

「教育の現象学」のデータサイエンス的転回

—全国学力・学習状況調査結果の分析から—

Data Scientific Turn of Educational Phenomenology

— Based on National Scholastic Tests —

田 端 健 人
Taketo TABATA

目 次

1. 研究背景
2. 現象学とは何か？
 2. 1. 現象学への最初の一步「志向性」
 2. 2. 「転回」を本質に含む現象学
 2. 3. ヨーロッパ諸学の危機と現象学
3. データサイエンスとは何か？
4. DS-EFAのリサーチクエスチョン
5. 教育データサイエンスの現象学的転回の第一歩
 5. 1. 信念
 5. 2. 直観
 5. 3. 意味
6. 生きられる数値
 6. 1. 生きられる（生きられた）数値の具体例
 6. 2. 体験における「存在秩序（調和）」ないしは「連関」
 6. 3. 「世界」という観点からの私たちの疑念の検討
7. 平均値差の現象学的データサイエンス
 7. 1. 平均正答率と平均正答数
 7. 2. 正答数の諸問題
 7. 3. 平均値差の帰無仮説検定
8. 平均値差の帰無仮説検定の試み
 8. 1. コルモゴロフ＝スミルノフ検定（KS検定）
 8. 2. t検定とz検定
9. 効果量
 9. 1. 帰無仮説検定から効果量へという世界的動向
 9. 2. 効果量（Effect Size）とは何か？
 9. 3. 効果量の基準値
10. 基準値という難問
 10. 1. 「レギュラトリーサイエンス」と「データサイエンス」
 10. 2. 効果量の基準値 $d=0.50$ の提案
 10. 3. 効果量の基準値からみた全国学力・学習状況調査のポイント差
11. 教育データサイエンスはどこまで科学的で現象学的か？
 11. 1. 教育データサイエンスの科学性
 11. 2. 教育データサイエンスの現象学性
12. 家庭の蔵書数の質問項目は児童生徒のSESの代替指標となるか？
 12. 1. 問題背景

12. 2. 検証のためのデータセットと手続きと検証結果
 12. 3. SESの合成変数の試行に関する現象学的省察
 12. 4. 教育データサイエンスにおける意味の空洞化の兆し
 13. 非認知能力と対話・探究学習と授業充実度の可視化
 13. 1. 「非認知『徳』(Virtus) 尺度」の開発
 13. 2. 「対話・探究学習 (Dia_Inq) 尺度」の開発
 13. 3. 「国語授業充実度 (Jpn_Ju) 尺度」と「算数・数学充実度 (Math_Ju) 尺度」の開発
 13. 4. 学力スコアと非認知「徳」スコア等との相関
 14. 自治体支援の事例
 15. むすびに代えて
- 資料：全国とトップ都道府県およびボトム都道府県の平均値差の効果量 (R4-H28, H21-H19)

1. 研究背景

2019年冬以来、筆者はテストスコアの分析や言葉の形態素解析に取り組んでいる¹。筆者はこれをチームで推進しており、私たちはこの取り組みを「DS-EFA (Data Science of Education for All: すべての子どものための教育データサイエンス)」と名づけている。

チーム全員ではないにしても筆者自身は、この教育データサイエンスを「教育の現象学」の一環と位置づけてきたが、初歩的なデータ分析を積み重ねるにつれその確信は強まっている。私たちの教育データサイエンスは現象学的だからこそ、テストスコアについての従来の専門的分析とは大なり小なり異なる着眼、異なる分析方法、異なるシステム開発、異なる分析結果になっているとの思いが強くなっている。

しかし私たちの教育データサイエンスは、本当に現象学的なのか、もし現象学的であるとすればどこまで現象学的なのか。現象学のデータサイエンス的転回なるものは可能なのか。かたやデータサイエンスとは何か。データサイエンスはどこまでサイエンス (科学) なのか。データサイエンスは現象学とどこまで地平融合できるのだろうか。こうした問題を本稿で考えてみたい。

2. 現象学とは何か？

そのためには、現象学とは何か、データサイエンスとは何かという根本問題を避けるわけにはゆかない。しかし、

現象学とは何か。フッサールの最初期の諸著作から半世紀も経ってなおこんな問いを発せねばならぬとは、いかにも奇妙なことに思えるかもしれない。それにもかかわらず、この問いはまだまだ解決からはほど遠いのだ。²

¹ 例えば、次の2冊はその成果の一部である。田端健人, 2022 『IRT分析ソフトEasyEstimationによる全国学力・学習状況調査の検証と経年比較』パイディア出版。田端健人, 2021 『子どもの言葉データサイエンス入門—形態素解析システムjReadabilityの活用と検証—』パイディア出版。

² メルロー=ポンティ, M., 1989 『知覚の現象学 1』竹内芳郎・小木貞孝訳, みすず書房, p.1.

これは1945年フランスで出版された『知覚の現象学』冒頭の言葉である。それからさらに75年以上経た現在、現象学とは何かの問いは解決に少しは近づいただろうか。おそらく現在でもこの問いは未解決のままであろう。「現象学とは何か」の問いに決まった答えがないならば、「私たちのデータサイエンスは現象学的か」の問いにどう答えればよいだろう。

2. 1. 現象学への最初の一步「志向性」

初心に帰り、初学者の道を歩んでみよう。それは私たちのデータサイエンスにふさわしいやり方である。現象学を修めた先人は初学者向けに「現象学とは何か」を解説してくれている。これは決定的な解決ではないにしても、現象学入門の第一歩になる。

「現象学」とは、哲学者フッサール(1859-1938)がその著書『論理学研究』の第2巻で始めた哲学上の意識分析の方法である。それは、認識対象と認識作用との相関関係を、「志向性」という概念を手がかりに分析する研究手法であった。³

引用文中の「認識対象と認識作用」は、「意識対象と意識作用」とか「ノエマとノエシス」とか「現出と現出作用」とも言い換えられる。両者を不可分とするのが「志向性」概念である。「志向性」とは、「なにかについての意識、なにかについての現出という、根本的な存在性格」のことである⁴。私たちは、何かを知覚するとか、何かを想起するとか、何かを表象するとか、何かを恐れるとか、何かを手に入れるとか、何かを決断するとかのように、何かへと向かう根本的な存在性格をもっている⁵。この「私たちが何かを・・・する」あるいは「何かの現われは、私たちにとっての現われである」という人間の根本的存在性格が志向性である。

この最初の一步を踏み出さない研究は、現象学と区別すべきであろう。例えば、自然現象を研究する場合、研究する私たちの意識作用を度外視するならば、自然科学とか物理学と呼ばれる研究になり、それは現象学的ではない。一方自然研究でも、私たちの意識作用を考慮に入れ始めると、その研究は現象学そのものではないにしても「現象学的」になる。現代物理学者ウェルナー・ハイゼンベルクの次のような自然像は現象学的である。光は粒としても波としても現れるが、それは観測する私たちの観測装置に依存することから、ハイゼンベルクは、「われわれが量子理論で数学的に公式化している自然法則はもはや粒子自体を取り扱っているのではなく、素粒子に関するわれわれの知識を取り扱っているのである」⁶とか、「この数学は、もはや素粒子の挙動を表わすのではなく、その挙動に関するわれわれの認識を表わしているのである」⁷とか、「われわれが劇場で観客であるばかりでなく、いつも共演者でもあることに気づかなければならない」⁸と述べ、素粒子が観測装置を利用した私たちの認識作用の志向的相関物であることを強調する。こうなると物理学も現象学的になる。

心とか意識とかを研究する心理学にしても然りである。従来の心理学が心とか意識とかを研究対象とし、研究している研究者の意識作用とか志向的相関関係を主題化しないならば、それは現象学的ではない。対照的に、研究対象の心理や意識を、対象化する意識作用との志向的相関において研究するならば、それは現象学的になる。

³ ハイデガー・フォーラム編, 2021『ハイデガー事典』昭和堂, p.32.

⁴ フッサール, E., 2004『ブリタニカ草稿—現象学の核心—』谷徹訳, ちくま学芸文庫, p.12.

⁵ フッサール, 2004, p.12.

⁶ ハイゼンベルク, W., 2006『現代物理学の自然像』尾崎辰之助訳, みすず書房, pp.8-9.

⁷ ハイゼンベルク, 2006, p.9.

⁸ ハイゼンベルク, 2006, p.9.

フッサール自身、「哲学的な現象学」に加え「現象学的心理学」を創設しようとした。

この哲学的現象学と同時に、しかしさしあたってはこれから切り離されることなく、方法的にも内容的にもこれと並行関係にある新たな心理学的な学科が生じてきた。その学科とは、アプリアリな純粋心理学あるいは「現象学的心理学」である。⁹

フッサールは、「哲学的な現象学」を「超越論的現象学」、「現象学的心理学」を「心理学的現象学」とも呼んでいる¹⁰。そうすると、現象学となる一定の条件を満たすならば、「教育の現象学」とか「○○の現象学」も、あるいは「現象学的教育学」とか「現象学的○○学」も、あながち牽強付会ではないだろう。

2. 2. 「転回」を本質に含む現象学

志向性を中核に据えることによって、現象学は現象学する研究者の多様性や各々の成長と不可分になる。つまり、志向性概念によって、現象学は「転回」という性格を本質に組み込むことになる。なぜなら、現象学の基本的態度は「事象そのものへ」であるが¹¹、「事象」は、志向性概念からすると、そこへと向かう人間の個性的な意識作用に応じて変容するからである。

このことは、先の現象学の定義における言葉の揺らぎにも表れている。

志向性を「意識対象と意識作用」の相関とみなすならば、現象学が中心に据える「事象」は「意識」や「意識的生」や「主観（性）」になる。これがフッサールの現象学であった。ところが、志向性を人間の「根本的な存在性格」とみなすなら、現象学の事象は、人間存在つまり「現存在」になる。これがハイデガーの現象学である。『存在と時間』執筆時のハイデガーは自身の探究を「現象学」と理解していたが¹²、本書によって現象学は「解釈学的転回」¹³ないし「存在論的転回」を迎える。ハイデガー自身の思索の歩みにおける「転回」は有名である¹⁴。フッサールにしても、1900/1年刊行『論理学研究』から1913年『イデー I』において「超越論的現象学への転回」¹⁵を見せている。転回させてこそ現象学だというのは言い過ぎかもしれないが、少なくとも現象学は、20世紀初めに台頭してきた「哲学運動」¹⁶であり、完結した哲学体系ではない。そうだとすれば、「現象学のデータサイエンスの転回」も、奇想天外な道ではないだろう。その最初の一步は、志向性がデータサイエンスの核心にどう関与するか、データサイエンスを志向性からどう展開できるか、にかかっている。

2. 3. ヨーロッパ諸学の危機と現象学

もう一つ、「現象学とは何か」を初歩的に理解するうえで欠かせないのは、現象学運動を突き動かす原動力が危機意識である、ということである。現象学は諸学問の危機から出発した。これはフッサール最晩年の著書名『ヨーロッパ諸学の危機と超越論的現象学』に凝縮されている。本書は1935年の講演をもとにしていて、その講演は次のような問いかけから始まる。

⁹ フッサール, 2004, p.7.

¹⁰ フッサール, 2004, p.57.

¹¹ ハイデガー・フォーラム編, 2021, p.344.

¹² Cf., ハイデガー・フォーラム編, 2021, p.33.

¹³ ハイデガー・フォーラム編, 2021, p.35.

¹⁴ Cf., ハイデガー・フォーラム編, 2021, p.7, p.399.

¹⁵ 谷徹, 2004「訳者解説」, フッサール, 2004, p.174.

¹⁶ フッサール, 2004, p.57.

学問のためにささげられたこの場所で行なうこの講演の「ヨーロッパ諸学の危機と心理学」という題目そのものからしてすでに異論を呼び起こすであろうことを、わたしは覚悟しておかねばなるまい。いったい、われわれの学問そのものが危機におちいつているなどとまじめに語られてよいものであろうか。¹⁷

この問いにフッサールは次のように応答する。

19世紀の後半には、近代人の全世界観は、もっぱら実証科学によって徹底的に規定され、また実証科学に負う「繁栄」によって徹底的に眩惑されていたが、その徹底性とは、真の人間性にとって決定的な意味をもつ問題から無関心に眼をそらす、ということの意味していた。単なる事実学は、単なる事実人をしかつぐらなない。このような傾向に対する一般的な評価の転換は、特に戦後（第1次大戦後）避けることのできないものとなったが、それは…（略）…若い世代のうちに、次第にこのような傾向に対する敵意に満ちた気分を惹き起こすまでになった。この事実学はわれわれの生存の危機にさいして、われわれに何も語ってくれない…（略）…この学問は、この不幸な時代にあって、運命的な転回にゆだねられている人間にとっての焦眉の問題を原理的に排除している。その問題というのは、この人間の生存全体に意味があるのか、それともないのかという問いである。¹⁸

ヨーロッパ諸学問は、「人間にとっての焦眉の問題」を排除していること、「真の人間性にとって決定的な意味をもつ問題から無関心に眼をそらす」していることが、フッサールを突き動かした学問の危機である。「真の人間性」を直視するのが現象学である。「真の人間性」とは、フッサールにとっては「志向性」に他ならない。

この危機意識は、なにも第一次世界大戦勃発ではじめて喚起されたわけではなく、いっそう先鋭化されただけである。フッサールが現象学をスタートさせた1900/1年『論理学研究』の序言に次の言葉がある。

〈数学およびあらゆる学問一般の客観性が論理的なものの心理学的基礎づけとどのように整合するか〉という原理的な疑惑さえもがますます私を不安にした。こうして、支配的な論理学の種々の確信に支えられていた私の方法…（略）…がすっかり動揺してしまったので、私は論理学の方法についての、特に認識作用の主観性と認識内容の客観性との相互関係についての一般的な批判的反省へとますます駆り立てられるのを覚えた。¹⁹

数学とあらゆる学問一般の原理的問題に対する「不安」から、フッサール現象学が始まったことがわかる。この不安は晩年の危機意識へとまっすぐにつながる。周知のように「不安」は、『存在と時間』によれば、人間を本来性へと転回させる根本気分である。なお、この引用文の「認識作用の主観性と認識内容の客観性との相互関係」が後に明らかにされる「志向性」である。

ヨーロッパ諸学への危機意識から、現象学にはあらゆる学問に対する強烈な批判意識が含まれることになる。「今日、実証科学はすべて成立基盤の危機に陥ってしまった」²⁰。しかしこの批判は、諸学問の単なる否定でも拒否でもない。ある時期、フッサールとハイデガーは、諸学問を基礎づける第1哲学として現象学を

¹⁷ フッサール, E., 1974 『ヨーロッパ諸学の危機と超越論的現象学』 細谷恒夫・木田元訳, 中央公論社, p.13.

¹⁸ フッサール, 1974, pp.16-17. 括弧内邦訳書。省略(略)は引用者(以下同様)。

¹⁹ フッサール, E., 1995 『論理学研究 I』 立松弘孝訳, みすず書房, pp.6-7.

²⁰ フッサール, 2004, p.145.

構想していた。

これ〔現象学〕がめざすのは、学問としての哲学をラディカルに新たに基礎づけること、そして、それによってすべての学問をラディカルに新たに基礎づけることである。²¹

筆者の個人的見解としては、現象学が哲学を含むすべての学問を基礎づけたとも、今後基礎づけるとも思わない。ここで押さえておきたいのは、諸学問に対する現象学の批判意識や危機意識は、諸学問を否定することでも、拒否することでもない、という点である。むしろ現象学は「基礎づけ」を標榜したほどに、諸学問を意味あるものに活かし返そうとする情熱をもっている。諸学問を「真の人間性」から考え直そうとする情熱である。

一方、確かに現象学には、諸学問と縁を切ろうとするかのような一面がある。「判断停止」「エポケー」「現象学的還元」と呼ばれる現象学者の態度変更である。

エポケーとは、次のような態度変更である。

意識が対象に関わる「自然的態度」を「遮断」し、それを根本的に変更する方法的懐疑の操作を意味する。その操作においては、自然的態度における一般定立が排去されて括弧に入れられ、いわばそのスイッチが切れ、定立がストップされる。この「括弧入れ」の操作を通じて、現象学に固有の新たな存在領域としての純粹意識が開示される。²²

フッサール自身の言葉によれば、

〔自然的態度のうちにある〕すべての実証科学は、超越論的に、エポケーに服さねばならないのであり、実証科学の対象領分すべてについても、それゆえまた心理学およびその意味における心的なものすべてについても、同様なのである。²³

このエポケー（判断中止）によって初めて、志向性とか志向的意識作用とかがあらわになる。とはいえ、実証科学や素朴な生活における自然的態度のスイッチを切るといっても、それは実証科学から縁を切るためではなく、実証科学を基礎づけようという情熱、諸学問を真の人間性に基礎づけようという情熱からであった。このことを心にとめておこう。

3. データサイエンスとは何か？

以上、「現象学とは何か」についての初歩的理解を確認した。

ではこの初歩的理解において、現象学はデータサイエンスとどのように切り結ぶことができるだろうか。そもそもデータサイエンスとは何だろうか。

「データサイエンス」という日本語は、一般に使われはじめてからまだ7年ほどしか経っていない。『広辞苑』第6版（2008年刊）には記載がなく、第7版（2018年刊）になって登場する。この言葉を、朝日新聞の記事データベース「聞蔵IIビジュアル」で検索すると、2021年末までの検索数は180件になる²⁴。これを年ごとに

²¹ フッサール, 2004, p.57.〔〕内引用者。

²² 木田元・野家啓一・村田純一・鷺田清一編集, 1994『現象学事典』弘文堂, p.42.

²³ フッサール, 2004, p.37.〔〕内は邦訳者の補足。

²⁴ 最終閲覧2022.01.19

グラフ化すると図1になる。2015年から記事に登場し、2021年に急上昇している²⁵。

もとの英語「data science」はどうだろう。3,000万冊をデータベースとするGoogle Ngram Viewer²⁶で、この語の使用頻度を検索してみた。比較対象として、この語のもととなった「computer science」と参考までに「phenomenologie」を合わせて検索した(図2)。data scienceは英米圏でも2010年を過ぎたところから使用頻度が上昇しはじめた若い言葉であることがわかる。またその使用頻度はじわじわ上昇しているものの、computer scienceと比べればまだ圧倒的に少ない。phenomenologieに比べれば、data scienceの方が頻度が高いこともわかる。

言葉の使用頻度からもわかるように、「データサイエンス」は「新しい分野」であり、「まだ完全な定義は

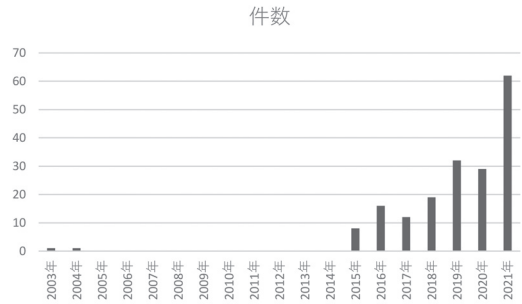


図1：聞蔵IIでのヒット件数

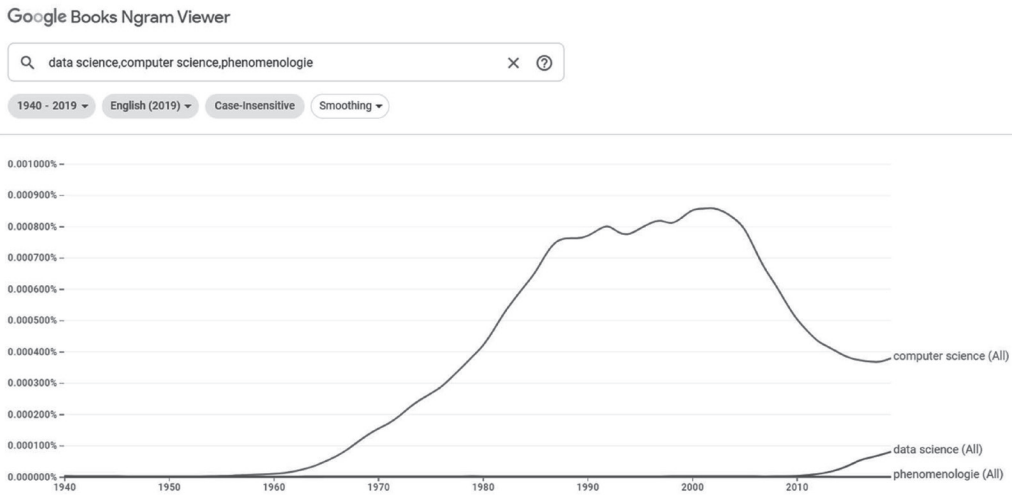


図2：Google NgramによるData ScienceとComputer Scienceの検索結果

ない²⁷。簡にして要を得ているのは、「データサイエンス」は「コンピューターサイエンスと統計学から産まれた子ども」であり、「統計的、コンピューターの、人間的」の3つを「決定的な構成要素」とするという解説である²⁸。

²⁵ ちなみにCiNiiで検索すると、2021年までで1,470件ヒットする（最終閲覧2022.01.19）。1999年に使用例があるが、年間の使用例が50件を超えるのは、2014年以降である。2019年から使用例が200件を超え、以後急上昇している。

²⁶ 2022.12.2アクセス。

²⁷ Skiena, S.S., 2020『データサイエンス設計マニュアル』小野陽子監訳, 長尾高弘訳, オライリー・ジャパン, p.1.

²⁸ Cf., Blei, M. D., & Smyth, P., 2017, Science and data science, *Proceeding of Natural Academy of Science*, 114(33), pp.8689-8690.

そもそもデータサイエンスで扱う「データ」とは何かも未解決の難問である²⁹。しかし、上記の便宜的定義からすると、データサイエンスの「データ」とは、コンピューターで扱える電子データに限定される。電子データをコンピューターで統計処理することがデータサイエンスの基本である。

しかも、このプロセスは完全には自動化できないため、「人間的」要素が不可欠になる。「コンピューターを利用した現代の統計ツールを、現代科学の問いに適用するには、人間による意味に満ちた判断と、訓練を積んだ深い知識とが必要である」³⁰。どのようなリサーチクエスチョンを立て、どのようにデータを分析・可視化するか、また計算された数値をどう評価するかは、分析主体の人間にかかっている。ここに、データサイエンスが現象学的になる余地がある。

先に確認したように、フッサールは人間性の核心を「志向性」にみた。データサイエンスを、「統計的、コンピューターの、志向性」の3要素で構想するとき、そのデータサイエンスは現象学への一歩を踏み出すことになる。

4. DS-EFAのリサーチクエスチョン

現象学的データサイエンスの道を切り拓くにあたり、私たちDS-EFAのリサーチクエスチョンを示そう。

私たちのプロジェクトの発端となったリサーチクエスチョンの一つは、全国学力・学習状況調査の学力スコアに関するものである。全国学力・学習状況調査は、小学6年生と中学3年生全員（各学年約100万名）を対象として、文部科学省が毎年1回実施する学力と学習状況に関する大規模調査である。基本的に、国語、算数・数学等の教科に関する調査と、児童生徒ならびに学校を対象とした質問紙調査で構成されている³¹。社会的に注目度が高く影響力も極めて大きな調査である。教科に関する調査問題が公表されると翌日の朝刊に紹介され、数ヶ月後に調査結果がリリースされると、翌日朝刊には、結果の都道府県ランキングとか、全国平均と自治体平均とのポイント差とかが大々的に報道される。

例えば、2021（令和3）年9月1日付『読売新聞』朝刊では、宮城県の結果として「3教科 平均下回る」の見出しで、次のような記事が掲載された。

国語は、小6の正答率が63%で、前回2019年度と同じく全国平均を2ポイント下回った。中3の正答率は65%で、全国平均と同じだったが、全国平均を1ポイント上回った前回よりも後退した。算数・数学は、小6の正答率が68%、中3が55%で、いずれも前回と同じく全国平均を2ポイント下回った。

また同朝刊には、「学テ 仙台、中学平均上回る 4年連続 政令市1位」とか「他地域は低迷 小学算数最下位」の見出しで、次の記事が踊った。

仙台市は中学生の国語、数学の平均正答率が全国20政令市でトップとなり、新型コロナウイルスの影響で中止となった昨年度を除き4年連続で1位となった。一方、仙台市以外の地域では小中ともに全国平均を2～5ポイント下回り、小学生の算数は全国最下位となった。

²⁹ 「データという語の使用は5世紀目を迎えているが、合意された定義はいまだに得られていない。」（ボーグマン, C.L., 2017『ビッグデータ・リトルデータ・ノーデータ—研究データと知識インフラ—』佐藤義則・小山憲司訳, 勁草書房, p.33.)

³⁰ Blei & Smyth, 2017, p.8690.

³¹ 詳細については田端, 2022『IRT分析ソフトEasyEstimationによる全国学力・学習状況調査の検証と経年比較』パイディア出版, p.1を参照されたい。

政令市1位の市民は鼻を高くするが、全国平均以下とか全国最下位の県民は惨めになる。全国平均以下とか全国最下位の結果は、県や市町村の議会でも槍玉に上がり、「学力向上」の圧力が各教育委員会、ひいては学校現場に重くのしかかる。「学力向上」の圧力は、「テストの点数を上げる」指導、「基礎基本の定着」、無味乾燥な「ドリル」学習となることも少なくない。

私たちのリサーチクエスションは、「全国平均を2~5ポイント下回することは、どのくらいの学力差なのか?」「それは大問題なのか小問題なのか?」「全国1位と全国最下位との学力差はどの程度なのか?」である。非常にシンプルで市民的な問いである。

この問いにまだ確かな解がないにもかかわらず、「全国平均以下」とか「全国最下位」とかのわかりやすい見出しが躍り、その評価に関係者はショックを受け続けてきた。私たちはそこに疑問を投じたのである。

問題をさらに複雑化し混乱させているのは、全国学力・学習状況調査の教科に関する調査は古典的テスト理論で設計されており、スコアを経年比較できないという問題である。新聞記者もそのことを知っているようで、例えば、2022(令和4)年7月29日『朝日新聞』朝刊1面では、次のような書きぶりになっている。見出しは「4年ぶり理科 正答率低下」である。

最も平均正答率が低かったのは、中3理科で49.7%。問題の難易度を調整していないため単純比較できないが、前回調査(66.5%)から16.8ポイント下がった。³²

下線部からすれば、記者は平均正答率が経年比較できないことを知っている。ところが「単純比較できない」と述べた直後に、「前回調査から16.8ポイント下がった」と単純比較するという矛盾に陥っている。

これらは氷山の一角である。このように、全国学力・学習状況調査は、社会的影響が極めて大きいにもかかわらず、全国平均とのポイント差の意味や重みが不明であり、経年比較できないにもかかわらず経年比較をしてしまうという混乱を招いている。この混迷を解決したいとの思いが私たちのリサーチクエスションには込められている。

5. 教育データサイエンスの現象学的転回の第一歩

私たちのリサーチクエスションを「志向性」と関連づけてみよう。門脇俊介氏の整理によれば、この志向性には「信念」「直観」「意味」の3種類がある³³。この3種類を整理しながら、これらと私たちのリサーチクエスションがどう関連するかを考えてみたい。

5. 1. 信念

信念とは、人間が「世界へコミットすることを可能にするさまざまな志向性のうち、…(略)…最も基本的なもの³⁴」である。「この世界は私が誕生する相当前から存在していたし、私の死後も長く存在するであろう」というのも信念である。「学力スコアは完全ではないにしても学力なる能力を数値化しているであろう」も一つの信念である。ある信念はその他の多種多様な信念と連動している。「学力スコア」についての信念は、「学力なる能力が存在する」という信念や「個人によって学力には差がある」といった信念と連動して

³² 下線は引用者。

³³ 門脇俊介, 2000「生活世界、志向性、人間科学」新田義弘編『フッサールを学ぶ人のために』世界思想社, p.280.

³⁴ 門脇, 2000, p.281.

いる。これらの信念は「全体論的な性格を持った一種の織り物として成立している」³⁵ことを発見したのがフッサールであり、彼はこの織物を「地平」という概念で明らかにした³⁶。そして「こうした地平的性格を持つ信念の織り物を通して現出してくるもの、言い換えれば、信念体系の相関者」が「世界」である³⁷。

私たちのリサーチクエスションは、「全国平均よりも2~5ポイント低い」という現出とか言説とかが、学力についての私たちの信念体系に織り込めないことに端を発している。私たちとは対照的に、ある県の平均が全国より2~5ポイント低いという現出を受けて、「これは報道しなければ」と思う記者たちや、この現出と言説を「問題」ととらえ教育政策を変更したり実行したりする人々にとっては、このポイント差は信念体系にうまく織り込まれていることになる。

私たちが、リサーチクエスションを起点として平均正答率2~5ポイント差を検証し、もしもこの差に実質的意味がないと結論を下したならば、それは「2~5ポイント差は大きい」という信念を否定することを意味する。そして、平均正答率とは異なるいっそう実質的意味のある評価システムを考案するとき、私たちは学力に関する信念体系を脱構築しようとしているのであり、学力の信念体系の相関者である学力の「世界」の一部を組み替えようとしている。

5. 2. 直観

私たちの信念は、単なる主観的な思い込みではなく、「世界の側から正当化という見返りを受け取る」³⁸。例えば、「世界はこの先も存在し続けるだろう」という信念は、私が朝目覚めるたびに正当化される。対して、世界によって正当化されない信念は、修正・変更を余儀なくされる。想像を絶する地震や津波は、私が住む地域が存続するという信念を震撼させる。読者には信じてもらえないかもしれないが、東日本大震災発生当日と数日続いた大きな余震、東京電力福島第一原発の爆発を身近に体験した私たちは、大学の会議室で「日本沈没」を否定する自信を失っていたことがある。

フッサールは、「信念の正当化は知覚をはじめとする直観によってなされる」とみなし、一方で「カテゴリーや数学的な対象などの抽象的存在者についての信念も、感覚的な知覚とは異なるタイプの直観によって正当化される」と考えた³⁹。しかし、管見ながら筆者が知る限り、フッサールは、この「感覚的な知覚とは異なるタイプの直観」の正体を明らかにしなかった。そこで以後、差し当たり、このタイプの直観を暫定的に「新しいタイプの直観」と呼ぼう。

私たちは学力なるものを知覚的に直観することはできない。児童生徒の学力であれ、自身の学力であれ、それを見ることも聞くことも、それに触れることもできない。国語の力、算数・数学の力なるものが児童生徒にあるという信念を私たちは素朴に抱いているが、その信念の正当化は間接的である。例えば、子どもが言葉を話したり書いたりするのを見聞きするとき、私たちは子どもの国語能力を感知する。しかしそれは、子どもの国語能力そのものを知覚的に直観しているわけではない。不正確な言い方だが、子どもの言動から間接的に推測しているに過ぎない。子どもの学力を感度良く現出させるツールの一つがテストである。テストスコアは、子どもの学力についての私たちの信念を正当化したりしなかったりする。この正当化は知覚による直観ではなく、新しいタイプの直観である。これがどのような直観であるかは、現象学的教育データサイエンスによって、おいおい明らかにしてみたい。

さらに、学力や非認知能力をテーマとするとき、知覚による直観が通用しなくなるもう一つの大きな要因

³⁵ 門脇, 2000, p.282.

³⁶ Cf., 門脇, 2000, p.283.

³⁷ 門脇, 2000, p.283.

³⁸ 門脇, 2000, p.284.

³⁹ 門脇, 2000, p.284.

がある。それは、私たちが関わるのは「個人の能力」だけでなく、全校の児童生徒とか地域全体の児童生徒とか日本全体の児童生徒といった「集団の能力」を相手にする点である。私たちが関わるのは、宮城県の子どもと仙台市の子どもの学力差の有無だったり、日本の子どもの学力水準だったりする。1学年約100万人は、私たちの知覚能力をはるかに凌駕している。100万人の学力を問題にし、日本のある学年の学力についての私たちの信念を正当化したり吟味したりするために、知覚的直観は全く無力であり、テストスコアをコンピューターで統計的に分析することに関わる新しいタイプの直観が必要になる。

二つの志向性、信念と直観について、ここでもう一つ重要な事実と言及しておきたい。近代の幕開け以降、科学技術を受け入れ進歩させてきた人類は、相容れない基本的信念に巻き込まれ、その矛盾は時代と共に大きくなっていくという事実である。それは、「大地は動かない」という信念と「大地は動いている」という信念、つまり天動説と地動説の信念体系である。地動説の信念体系には、「地球は丸い」とか「地球は自転と公転をしている」といった信念も織り込まれている。一方は知覚的直観が可能であり、他方は知覚的直観が不可能である。

大地は動かないという信念は、いつでも知覚を通して世界から正当化される。今も私はパソコンに向かいながら、大地が動いていないことを感覚的に直観できる。天体の運動についても、日の出や日の入りを見ることで直観できる。

ところが地球の運動を私たちは知覚的に直観できない。どうして、知覚的に直観できない信念、知覚的世界から正当化されない信念が揺るがないのだろうか。

知覚的直観が不可能な信念が一瞬わずかに揺らぐ思考実験を試みよう。地球の自転速度は、赤道直下で時速916km、新幹線の約5倍、ジェット機の約2倍、秒速約460mとされる⁴⁰。地球の公転速度はさらに速く、秒速約30km、音速の87倍（マッハ87）とされる⁴¹。『トップガン』のトム・クルーズでさえマッハ87の戦闘機は操縦できないだろう。これほどのスピードで地球が動いていることを想像するなら、「本当だろうか？」とわずかに信念が揺るがないだろうか。仮に揺らいだとして、次に座標系の知識を持ち出すならば、この揺らぎが収まるのがわかる。かなりの速度かもしれないが、同じ座標系の等速運動なのだから動きは感じないだろう、新幹線に乗っていても加速減速しない限り速度を感じることはないし、窓に置いた缶ビールも安定しているし、という具合に信念体系にうまく織り込むことができる。この思考実験から、知覚的に直観できない信念は、ある特殊な知的世界によって正当化されていると推測される。そして、この特殊な知的世界は、個人と集団において歴史的に形成されてきたと考えられる。しかしこのタイプの世界と直観の正体はまだ謎である。この正当化を「意味づけ」と呼びたいところだが、フッサールでは「意味」は第3の志向性の用語として直観とは区別される。

5. 3. 意味

私たちの「信念は、明示的な信念にせよ、より基底的な非明示的な信念にせよ、『意味』による分節化を受けて」おり、「意味による分節化とは、その意味を持つ信念の正当化がどのようなものであり他の信念とどのような関係をとるのが、規則として定められている」⁴²という。難解な解説だが、例えば「今日、A氏は研究室にいる」という命題は、「今日A氏は自宅にいるという信念と不整合をなすこと、A氏は今日、研究室のドアを開けるとい信念を含意すること、等々を規則として定める」⁴³という。初心者的に理解すれば、それぞれの信念は意味によって分節化されており、それらの意味はある種規則的で整合的であるという

⁴⁰ Webサイト「Gakken キッズネット」(2022.09.24閲覧) 参照。

⁴¹ Webサイト「Mayonez」(2022.09.24閲覧) 参照。

⁴² 門脇, 2000, p.285.

⁴³ 門脇, 2000, p.285.

ことだろうか。

「全国平均正答率を2~5ポイント下回る」という命題を私たちは間違いだとは思っていない。平均正答率で計算すれば、このポイント差になる。この計算は正しい。この命題は、「正答率とは、正答数を全問題数で割って100をかけた値」という命題や、「令和3年度の小6算数全国平均正答率は70%で、宮城県の正答率は68%だったため、宮城県は2ポイント低い」といった命題と整合的である。これらの命題は、言葉の意味としても計算規則としても間違っていない。

とするならば、私たちの疑問や疑念は、志向性の信念と直観に関わる問題であると同定できるだろうか。なぜかはわからないが、私たちにとってこのポイント差には、地球の自転速度がジェット機の2倍であるとか、公転速度がマッハ87であるという信じ難い命題よりも真実味がなく疑わしい。

よく考えてみると、意味の点でも、この命題に不整合なり疑念の余地がないわけではない。令和3年度の全国の小6は100万人弱、宮城県は2万人弱、これだけの集団の平均値で2~5ポイント差は、個人差と同じ意味をもつだろうか。これは意味の整合性の問題であろう。そうだとすれば、私たちの疑念は、3種の志向性全体に及んでいるかもしれない。

6. 生きられる数値

6. 1. 生きられる（生きられた）数値の具体例

英語圏の現象学には「生きられる経験 (lived experience : 生きられた経験)」という用語がある⁴⁴。フッサールのドイツ語Erlbnis (体験) の英語訳である。この英語訳をヒントに、「生きられる数値 (lived number)」あるいは「生きられた数値 (lived number)」という用語を造語してみたい。

私たちの日常生活には数値が溢れているが、それらの数値には、実感に裏打ちされた数値とそうでない数値がある。これを区別するための造語である。「生きられる」か「生きられた」かは、文脈に応じて使い分ける。

私たちの日常生活には、生活を左右する重要な意味をもつ数値とそうでない数値とがある。銀行員とか経理係とかエンジニアとか統計学者など数値を専門に扱う職業は当面除外しておこう。朝起きて顔を洗うという生活場面を考えよう。顔を洗うと私は体重計に乗る。私の体調管理として、61kg台なら安心するが、62kg台なら昨日の食生活や運動を反省し今日の食事や運動を微調整する。自分の腹回りや健康維持のためである。次に私はよく検温をする。新型コロナ感染症流行からおおよそ3年間頻繁に検温しているが、ワクチンの副反応を除き、37℃を上回ったことは不思議と一度もない。37℃を上回ればPCR検査を受け、それで陽性となれば、私は向こう5日間の生活様式を変更しなければならない。少し余裕があれば豆を挽いてコーヒーを入れる。豆20gに対して抽出量300ml、少し多めで25gに対して400mlが私の目安である。豆の特徴や焙煎の度合いによって微調整する。朝食を済ませば出勤のため車を走らせる。ある区間の制限速度は30km/hである。道路交通法の遵守ということもあるが、人通りの多い道路を私は30km/hかそれ以下で慎重に走行する。

これらは生活上現実味があり、行動を左右する数値である。こうした数値を「生きられた数値」と呼ぶことにしたい。

生きられる数値には他にも数多ある。スーパーで買い物をする時のものの値段、購入するリンゴの数、肉の重さ、預貯金の残高、海外旅行を計画している時には為替相場など、タイプは異なるがすべてが生活上有意義で、生活に影響を与える生きられる数値である。単位やカテゴリーの異なる数値が、日常生活では混在しているが、それは当たり前で、全体として調和を保っている。

⁴⁴ Moran, D. and Cohen, J., 2012, *The Husserl Dictionary*, Continuum, p.195.

一方、それまで生きられていた数値が、突如生きられなくなることもある。例えば、ニュースで報じられる「今日の全国の感染者数」である。それまではその数値を参考に、「感染が拡大している」とか「収まってきた」と目安になっていた数値、それを目安に翌日の外出を控えるとかしていた数値が、全数把握をやめた途端、実質的に機能しなくなった。報道される感染者数が感染者の多さを意味するのか少なさを意味するのか、まったくわからなくなった。

「全国平均正答率から2~5ポイント差」もこれに似ている。100点満点のテストで平均点が70点、ある子どもが68点だったとして、それを問題視するのはかなり教育熱心な保護者である。しかもこれは個人のスコアではなく、全国100万人とか宮城県2万人弱とかの集団スコアである。謎が謎を呼ぶ。個人スコアとしてはさほどの差ではないとも思われるが、都道府県単位の集団にすれば意味のある差になるのか。この疑問に誰か答えているのか。こうした疑念はこの数値の生きづらさの表現である。

6. 2. 体験における「存在秩序（調和）」ないしは「連関」

生きられる経験について、フッサールによる「体験 (Erlebnis)」の解明をもとに理解を進めてみよう。

どんな体験もみな、それ自身において、生成の流れである。…(略)…つまり体験は、過去把持と未来把持の絶えざる流れであって、過去把持と未来把持とは、原的状态というそれ自身が流れ行きつつある位相によって媒介されている。この原的状态の位相において、体験の生きいきとした今が、体験の「以前」と「以後」とに対比されて、意識されるのである。⁴⁵

意識の美しい記述である。「未来把持」は「未来予持」とも訳され、「生きいきした今」は「生き生きとした現在」とも訳される。今私が机を注視しているとしよう。私の知覚野の中心には机が現前している。その周囲には机ほど焦点が絞られていない床や壁や窓や窓際の植物などが広がっている。机が現前し続けているが、この間も私の体験は絶えず流れ続けている。私が注意を窓際の植物へと移すとき、それまで原的印象にとどまっていた机は過去把持へと転じ、先ほどまで私の注意を触発し未来予持されていた様々な周囲の物の中から窓際の植物が原的印象に流れ込む。これは時間的な「存在秩序」⁴⁶に注目した体験の記述であり、これを空間的な「存在秩序」の観点から記述すると、以下になる。

私は、たった今見て注意していた書きもの机から、私の注意を移動させて、部屋の中にある私の背後の今まで見ていなかった部分へと、注意を向け、さらにそれを通り過ぎて、ヴェランダへ、庭の中へ、あずまやの中にいる子供たち等々へと、注意を転じさせることができる。つまり、直接的に一緒に意識されている私の周囲の中のそこそこにそれらが存在するのを、私がまさに「知っている」ような、そうした客観のすべての方へと、私は私の注意を転じさせることができるのである。—今、「知っている」と言ったが、この知は、概念的思考を全く含んでいない。けれども他方そこへと注意が向けられると初めて、この知は、明瞭な直観へ、把握の意味における知覚へ、同じく確証的な経験へと、転化してゆくのである⁴⁷

⁴⁵ フッサール, E., 1996 『イデー I-II—純粹現象学と現象学的哲学のための諸構想 第1巻 純粹現象学への全般的序論—』 渡辺二郎訳, みすず書房, p.54.

⁴⁶ フッサール, E., 1992 『イデー I-I—純粹現象学と現象学的哲学のための諸構想 第1巻 純粹現象学への全般的序論—』 渡辺二郎訳, みすず書房, p.128.

⁴⁷ フッサール, 1992, p.126.

この記述では、過去把持や未来予持していることが、「概念的思考を全く含んでいない」で「知っている」状態と表現されている。そして、不明瞭な直観が明瞭な直観へと移行することが「確証」あるいは「確証的な経験」と表現されている。注意が向けられ原的印象において直観され確証されている対象の周囲を取り囲む不明瞭な直観、さらにそれを越えて時間的空間的に広がる開けが、先に触れた「世界」であり「地平」である。

世界というものは、むしろ、ある確固とした存在秩序をなしながら、限りないところにまで達しているのである。顕在的に知覚されたものや、また多少とも明瞭な仕方と一緒で一緒に現在しているかつ規定されているもの（もしくは少なくとも幾分かは規定されているもの）などは、未規定的な現実というあいまいに意識された地平によって、一部は浸透され、一部は取り囲まれるのである。私はその地平の中へと、注意の目差しという明るく照らし出す光線を、送ることができ、そのつどいろいろな成果を取めることができる。⁴⁸

知覚的世界は、「確固とした存在秩序」をなしており、この秩序に従って、注意を向けるごとに対象の原的印象が確証される仕組みになっている。これが、先に信念の正当化とか、直観の整合性と表現されたことである。信念の織物も世界の存在秩序に相関している。個々の直観や個々の信念を支え、最終的に正当化するの、全体として存在秩序を織りなす世界なのである。

この存在秩序は、別の個所では「調和」とも呼ばれる。

外的知覚の場合には、あらかじめ把握する予料が〔徐々に〕調和的に充実されていくという形式をもつことができ、こうした意識こそが、おのれを明示してくる物それ自体についての意識として、明証的な存在信憑に適うのである⁴⁹

「あらかじめ把握する予料」は「未来予持」、「存在信憑」は「信念」のことである。

フッサールは、この知覚的世界と「同じ直接性において、価値世界、財貨世界、実践的世界」が「現にそこに存在している」とみなしていた⁵⁰。価値や財貨や実践の世界も同様でありながら固有の存在秩序（調和）があると考えていたようである。しかし、邦訳者の渡辺二郎氏が指摘するように、この点の解明がフッサールでは不十分である。

「価値」や「実践」の諸性格が、「事象」としての「客観」に、どのような形で、「構成的」要素として含みこまれるようになるのか、という点は、問題として残る。一般にはフッサールはやはり、「物質的実在」の構成を一番基礎に置き、他のものはその上に「基づけられた統一」として重層化して成立すると見る…（略）…しかし『イデー II』では、それがそう単純な一方的基礎づけでなく、「循環」のあることが認められている…（略）…なお、ハイデggerは、「自然事物」に「価値述語」が「付加」されて「使用事物・実用品」が出てくるという「重層」…（略）…の考え方を、批判している…（略）…なぜならそれは、vorhandenの上にzuhandenを繋ぎ合わせる考え方で、本来的に、世界現象やzuhandenの現象を飛び越しているからである。⁵¹

⁴⁸ フッサール, 1992, p.127. 括弧内邦訳書。

⁴⁹ フッサール, 2004, p.74. [] は邦訳者の補足。

⁵⁰ フッサール, 1992, p.128.

⁵¹ 渡辺二郎, 1992「訳注」, フッサール, 1992, pp.352-353.

ドイツ語の「vorhanden」とは「事物的存在者」、「zuhanden」とは「道具的存在者」と訳される。目や手の前にあり知覚される事物と、手元にあつて利用される道具とは、存在性格が異なり、重層しているわけではないというのが『存在と時間』のハイデガーの捉え方で、ハイデガーは「・・・のため」という連関の全体を「世界」とみなした。

「ブリタニカ」論文でも、ハイデガーが編集すると、先の体験の存在秩序は、次のような書きぶりになる。

知覚—たとえばサイコロの知覚—は、原本的に把握するまなざしにおいて、この一つの物をもつ。けれども、この知覚は、〔志向的な〕体験として、物を単に内容空虚に〈そこにもつ〉ということなのではない。むしろ、知覚において、この〔ひとつの〕物が多様な「現出様式」とおしておのれを呈示しているのである。そして、こうした多様な「現出様式」の連関〔＝結びつき〕がまさにはじめて知覚を構成している⁵²

この現出様式の「連関」が世界であり、知覚される物の世界と、使用される道具の世界とでは現出様式が違うというわけである。

6. 3. 「世界」という観点からの私たちの疑念の検討

以上の考察のポイントを要約すると、次のようになるだろう。

- ① 個別の現出の背景には世界がある。
- ② 世界には、存在秩序とか調和とか現出様式の連関とか呼ばれる根本様式がある。
- ③ 個別の現出を取り囲む未来予持の地平に目を向けるとき、世界の存在秩序に調和する直観によってその未来予持が原的印象によって充実（つまり直観）されなければ、その個別の現出の信念は確証（正當化）されない。
- ④ 知覚的直観についてはフッサールが明らかにしたが、非知覚的直観の正体はいまだ謎である。

これらの観点から、私たちの疑念を改めて検討してみよう。

「平均正答率2~5ポイント差」は個人レベルならば、私たちの経験上、入学試験など人生を左右する試験でもない限り、大騒ぎするほどの大差ではない。これは私たちの経験からある種直観できる。個人レベルのこうした感覚からすれば、この差を大きく報道するのは不適切である。しかし、千や万という単位の集団に、この個人レベルの感覚、個人の生活世界の秩序を適用できるかは不明である。そもそも、この規模の集団的世界の存在秩序が依然不明である。この集団規模の世界に入ろうとした途端、上下左右の区別がつかない宇宙空間に投げ出されるような気分になる。この存在秩序を解明する必要がある。もしかすると、この存在秩序を創設するのが現象学的教育データサイエンスの使命なのかもしれない。「千や万という単位の集団において、2ポイント差は大きい」ことが何らかの仕方では確証されるなら、あるいは「大きくない」と確証されるなら、私たちは当初の疑念に終止符を打つことができ、このポイント差を生きていくことができるようになる。確証を得る際、意味の規則にも注意を払おう。

7. 平均値差の現象学的データサイエンス

そこで、意味連関に注意しつつ、このポイント差の大小について何らかの確証を得ることを試みたい。この確証は、知覚的直観とは異なる「新しいタイプの直観」であるが、その正体はまだ謎であるため、私たちのいわゆる勘所（常識的感覚ないし共通感覚か？）に訴えながら検討してみたい。

⁵² フッサール, 2004, p.101. [] 内はいずれも邦訳者の補足。

7. 1. 平均正答率と平均正答数

ポイント差は、平均正答率の平均値差のことである。「ポイント」という言葉も2つの正答率の差からきている。私たちが調査を進めて初めて知ったことだが、全国学力・学習状況調査の採点は、正答数でカウントされている。問題への配点はなく、正答1、誤答0で採点する。児童生徒ごとに正答数を全問題数で割り、100を掛けた値が正答率である⁵³。

志向性のこうした意味に関して、文部科学省がどうして平均正答率に換算して公表しているか疑問になる。

もちろん文部科学省は、平均正答数も公表している。公表される標準偏差は、平均正答数の標準偏差である。平均正答率の標準偏差は、筆者が知る限り公表されていない。それゆえ、のちに論及する「効果量」を計算する場合、平均正答率の公開データでは効果量を計算できない。この点に意味の**ばらつき**ないし**空所**がある。文部科学省が平均正答率の標準偏差を公表しないのは、憶測するに、2つの標準偏差を示すと教育関係者が混乱するからではないだろうか。加えて、教育関係者の多くは平均値は見るが、標準偏差は見ないという現状もある。その現状において、標準偏差を2つ掲載することは、さらなる混乱を招きかねない。図3に例示するように、「調査結果概況」では、「標準偏差」だけでなく、「中央値」「第3四分位」「第2四分位」「第1四分位」「集計値」「分布グラフ」すべてが正答数で計算表記されている。正答率は、「おまけ」的な後付けであることがわかる。

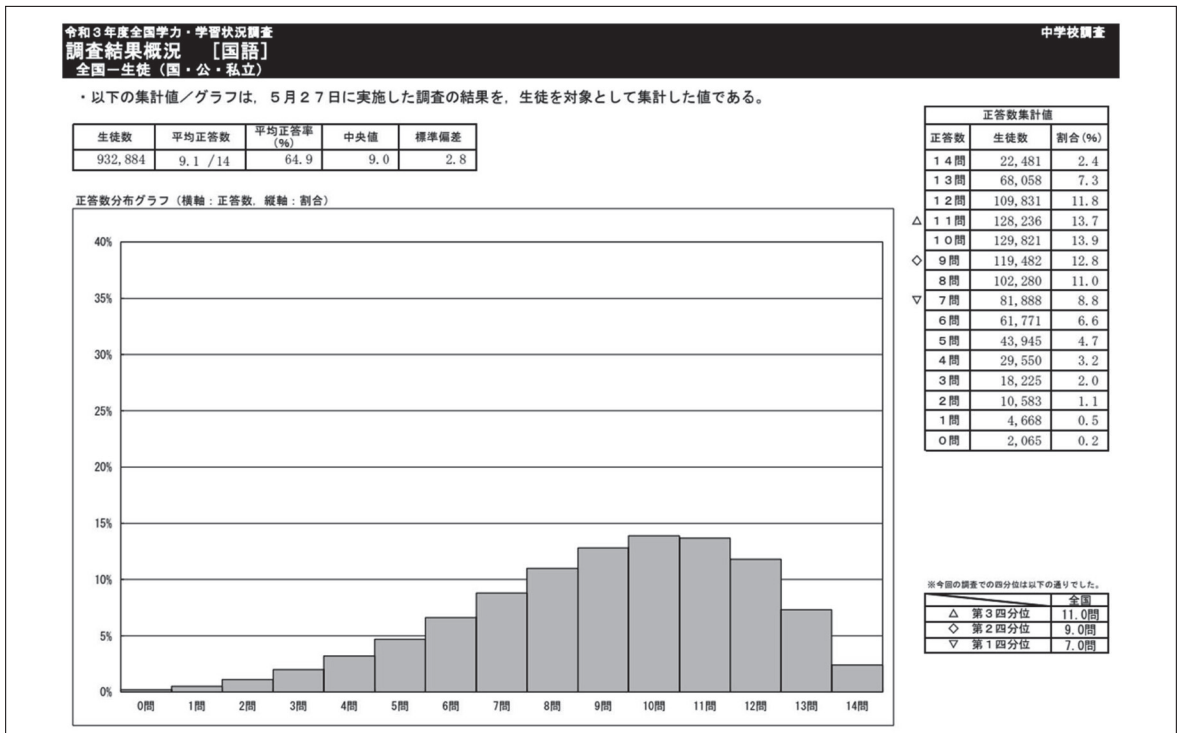


図3：「調査結果概況」令和3年度、中3国語の例

⁵³ Cf., 田端健人, 2021a 「全国学力・学習状況調査の平均正答率をどう受けとめるべきか? — 「生きられた数値」による〈子ども教育データサイエンス〉の構想—」『学ぶと教えるの現象学研究』19, p.2. 本稿第2版は田端2022, pp.93-113に所収されている。

なお、図3のグラフは、国立教育政策研究所のウェブサイト「全国学力・学習状況調査」(<https://www.nier.go.jp/kaihatsu/zenkokugakuryoku.html>)の「令和3年度 報告書・調査結果資料」から、「4. 令和3年度 全国学力・学習状況調査【中学校】調査結果資料」に進み、「2(1) 全国－生徒(国・公・私立)」をクリックするとウェブページとして表示され(https://www.nier.go.jp/21chousakekkahoukoku/factsheet/data/21m_201.pdf)、ダウンロードも可能である。

先に進む前にここで、この意味の綻びをわずか2個所だが令和3年度中3の国語と数学で繕っておこう。平均正答率の標準偏差を計算するために、文部科学省の匿名データ貸与制度で貸借した令和3年度中3の正答率換算データを利用する。

以下、データ操作に関わる表記では、元号の表記に頭文字のローマ字を、小6に「EL」、中3に「JH」、国語に「Jpn」、算数・数学に「Math」の略記号を用いる。

R3のJH匿名データからJpnとMathの正答数と正答率を取り出した中間生成物を作成する。貸与データ個票数は90,398名である。IBM社の統計ソフトSPSSで欠測値(SPSSでは「欠損値」)を削除し、記述統計を計算した結果が表1である。「NCA」は正答数、「CAR」は正答率を表わす。欠測値削除後の個票数は、表1の「度数」にあるように、90,237名である。

表1：R3_JHの平均正答数・平均正答率と標準偏差
記述統計量

	度数	最小値	最大値	平均値	標準偏差
NCA_Jpn	90237	.00	14.00	9.06	2.79
NCA_Math	90237	.00	16.00	9.15	3.69
CAR_Jpn	90237	.00	100.00	64.68	19.94
CAR_Math	90237	.00	100.00	57.20	23.05
有効なケースの数(リストごと)	90237				

文部科学省公表の国公私立データ(図3参照)と照合すると、平均正答数と標準偏差は小数第2位四捨五入で一致、平均正答率は小数第1位でわずかにズレるが、一方は全数データ、こちらは約10%の抽出データであること、またSPSSへのデータ取り込み時の小数の「まるめ」も関与することから、誤差の範囲に収まる。平均正答率で「効果量」を算出したいときは、表1の「率」の方の標準偏差を使えばよい。これで意味の空所2個所が埋め合わされた。こうした細かい修繕も現象学的データサイエンスの仕事である。

本題に戻ると、どうして意味連関として不完全な平均正答率を公表するのだろうか。文部科学省の真意を知ることはできないが、正答率と正答数の2値が公表され、しかもよく知られた「調査結果概況」(図3参照)では正答数が先に記載されているのに、報道機関はどうして正答率を選ぶのだろうか。この理由は明白で、市民に「わかりやすい」からである。この「わかりやすい」は、現象学のいう「直観」、前述の知覚的直観とは異なる「新しいタイプの直観」に訴えるということである。

仮に「全国平均正答数は全14問中9.7問だった」と朝刊で報じられたならどうだろう。読者はこれを百分率(正答率)に計算し直すであろう。あるいは、「どうして正答率で報道しないのか」という反応が予想される。これに対して、「全国平均正答率69%」という報道ならば、読者にはピンとくる。これは不思議な現象である。平均正答率が私たちにピンとくるのは、それが私たちの過去のテスト経験、つまり100点満点のテスト経験を隠に陽に覚起する(思い出させる)からだろう。現象学的に表現すれば、平均正答率は、私たちが慣れ親しんでいる学力テストの信念連関と意味連関に収まるからである。また平均正答率は、わたしたちによって習慣的に生きられた数値である。私たち一般市民には、正答数の報道を受け止める地平はないが、正答率の報道なら受け止める地平があるとも言える。

しかし、正答率の「わかりやすさ」から意味の混乱や綻びがはじまる。最大の綻びなり誤解は、「正答率にすることで、経年比較できる」という誤った印象を呼び起こすことである。先の新聞記事に表れていた通りである。しかし正答率に変換したところで、全国学力・学習状況調査の教科に関する調査のスコアは経年比較できない。先の新聞記事が言及しているように、年度ごとに問題の難易度が異なるからである。難易度の違いは、平均と標準偏差に影響を与える。意味連関をさらにたどるなら、難易度だけでなく、テストの識別力が揃えられていない（等化されていない）からである。常識的にわかるように、難易度は平均に表れる。難易度の高いテストの平均は低くなる。識別力はちらばり（分散や標準偏差）に表れる。同一集団に識別力の高いテストをすれば、ちらばりは大きくなり、識別力の低いテストをすれば、ちらばりは小さくなる。難易度と識別力を等化しなければ、異なるテスト結果を比較することはできない。ところが、平均正答率の「わかりやすさ」がこの難問を見えなくしてしまう。

平均正答率の問題については、このあたりにとどめておこう。平均正答率は、見た目の「わかりやすさ」のためであるが、この「わかりやすさ」の先には意味連関が欠如し、行き止まりになっている。「調査結果概況」（図3）を見ただけで、平均正答率がいわば「おまけ」であることは一目瞭然である。おまけを正面に据えるのは見込みがない。そこで、おもとの正答数に立ち戻ろう。たとえそこから疑問・疑念が出たとしても、教育データサイエンスで補修可能であり、新しい意味連関と直観連関を構築する見込みがあるからである。

7. 2. 正答数の諸問題

正答数に立ち返るや否や、種々の素朴な疑問や疑念が噴出する。1つは「毎年どうして問題数が変わるのか」であり、もう1つは「平成31年度以来問題数が激減しているが、テストの信頼性として大丈夫か」あるいは「全10問とか全14問で本当に学力が測れるのか」である。

平成24年度から令和4年度までの問題数を一覧にすると、表2になる。

表2：H24～R4_EL_JH_問題数一覧

	H24	H25	H26	H27	H28	H29	H30	H31	R 3	R 4
EL_Jpn	28	28	25	23	25	24	20	14	14	14
EL_Math	33	32	30	29	29	26	24	14	16	16
JH_Jpn	41	41	41	42	42	41	41	10	14	14
JH_Math	51	52	51	51	51	51	50	16	16	16

(単位：問)

H31以降のテスト問題数の減少によるテストの精度について、筆者は既に前著で多角的に検証し、結果として信頼性が低下していることを指摘した⁵⁴。そして、信頼性係数（クロンバックの α 係数）にして、国語は0.75、算数・数学0.80を「基準値」として提唱し、それを下回らないテストにするよう提言している⁵⁵。参照いただければ幸いである。

⁵⁴ Cf., 田端, 2022, 5・6・7章.

⁵⁵ 田端, 2022, p.36. 「基準値」については本稿で後に詳述する。令和4年度の信頼性係数については、今後計算する計画である。

7. 3. 平均値差の帰無仮説検定

本稿では、平均値差の大小の問題に焦点を絞る。

「平均値差の検定をして、有意差ありとか有意差なしと白黒つけられないものか」という素人考えは自然である。「有意な (significant)」という言葉は、素人にも専門家にも力強く重い言葉である。

有意差の検定は、帰無仮説検定といわれる。全国学力・学習状況調査の結果を帰無仮説検定できないものだろうか。

帰無仮説検定は、次の6つのステップを踏む⁵⁶。

第1ステップ：仮説 H_0 と H_1 を立てる

「 H_0 」という記号は「帰無仮説 (null hypothesis)」、それを否定する仮説「 H_1 」という記号は「対立仮説 (alternative hypothesis)」を意味する。たいていの場合、帰無仮説検定は H_0 を棄却し、 H_1 を立証するために用いられる。例えば、新薬開発などで、新薬を投与する実験群とそうでない統制群とを設定し、 H_0 「実験群と統制群に差はない」という帰無仮説を立て、検定でこれを棄却することで、 H_1 「実験群と統制群に差がある」を支持し、よって新薬には有意な効果があると立証するのが、わかりやすい使い方である。

第2ステップ：検定統計量 (T) と分布を決める

帰無仮説検定には複数の種類がある。「t検定」とか「 χ^2 (カイ2乗) 検定」とか「KS検定」とかである。帰無仮説からどの検定がふさわしいかを決めると、それぞれの検定統計量 (T) の計算式とそれが従う分布が決まり、データをその数式に当てはめる。どの検定を選ぶかは研究者の肉体的な判断だが、検定統計量の計算はコンピューターに任せることができる。

第3ステップ：有意水準 α を決定し、棄却域を決める

ここでもまた非常に肉体的な判断が介入する。

第3ステップでは、帰無仮説を棄却するために「棄却域」を設定する。棄却域となる基準値は「有意水準 (significance level)」と呼ばれ、慣習的に5%とされる。この「5%に論理的な根拠は何もなく、ただの慣習にすぎない」⁵⁷ことは、入門書や解説書で必ずと言ってよいほど力説されることである。

ただ「この5%という有意水準を分水嶺として『2値的』な判断を行うことが帰無仮説検定の本質」⁵⁸である。ここに仮説検定のわかりやすさと実用的な意義がある。白黒決着をつける魅力がある。「5%水準で有意差あり」と言えるか否かは、統計学という学術を背景としている点で、「その差は大きそうだ」という素人印象とは比べものにならないほどの重みがある。ただしこのわかりやすさがまた、5%水準があくまで慣習でしかないことを隠してしまう。

第4ステップ：データを取得し検定統計量 (T) を算出する

このステップについては、解説は必要ないだろう。データさえあれば、統計ソフトでP値 (P-value) が算出される。P値が有意水準以上か、それ未満かが分水嶺となる。

第5ステップ：仮説の棄却と採択

慣例的な有意水準5%の場合、計算結果のP値が $p < 0.05$ ならば帰無仮説を棄却し、対立仮説を採択 (支持) し、 $p \geq 0.05$ ならば帰無仮説を採択する。

以上が帰無仮説検定の考え方とプロセスである。

常に注意しなければならないのは、この検定は必ずしも正しいとは限らないことである。有意水準5%か

⁵⁶ Cf., 大久保街亜・岡田謙介, 2012『伝えるための心理統計—効果量・信頼区間・検定力—』勁草書房, pp.25-27.

⁵⁷ 大久保・岡田, 2012, p.27.

⁵⁸ 大久保・岡田, 2012, p.26.

らして既に、あくまで慣例であり、この水準が適切か否かには議論の余地がある。これは「基準値」の問題である。論理的な根拠がないとはいえ、都合の良いように変更できるわけではなく、それなりの根拠で設けられてきた値である。この問題はのちに改めて検討しよう。

加えて検定は、「確率に基づく決定である以上、判断に誤りはつきもの」⁵⁹である。その誤りには2つの過誤があるとされる。「第1種の過誤」は「帰無仮説が正しいにもかかわらず、帰無仮説を棄却する誤り」である⁶⁰。これは実際には無いものを棄却する誤りであり、諭えるなら、無罪の人間の主張を棄却し、有罪判決とする過誤である。のちに見るように、実質的にはわずかな差しかないのに、サンプル数が多いため $p < 0.05$ となり、帰無仮説を棄却し、差があると判定することも、第1種の過誤である。「第2種の過誤」は「帰無仮説が偽である…(略)…にもかかわらず、帰無仮説を採択してしまう誤り」である⁶¹。これは実際にはあるものを無いとする過誤であり、諭えるなら、有罪の人間を無罪にしてしまう間違いである。

このように検定と言っても絶対的に正しいわけではない。しかし、社会的な影響力を持つ平均差について、何の根拠もなく「その差は大きい」とされてきた状況を少しでも改善するためなら、統計学を背景とする帰無仮説検定は助けになるだろう。

8. 平均値差の帰無仮説検定の試み

そこで全国学力・学習状況調査について、帰無仮説検定を試みてみよう。

私たちの帰無仮説 H_0 の一つは「全国学力・学習状況調査における全国トップ都道府県平均と全国ボトム都道府県平均とに差はない」である。この帰無仮説が採択されるならば、連動して、この差の範囲内に収まる全国平均と宮城県（仙台市を除く）平均とに有意差はないことになる。また、全国都道府県ランキングがナンセンスであることになる。

たいがいの帰無仮説検定は、先の新薬開発の例のように、 H_0 を棄却することを狙っているが、私たちの検定にはそのような狙いはない。 H_0 が棄却されれば H_1 を採択する用意があり、その場合はメディアでの報道を受け止め、全国最下位の平均は有意に低いと判断し、その改善を自治体に勧めるし、一方 H_0 が棄却できなければ、その平均差は小さいとし、全国都道府県ランキング報道を控えるよう提案するだけである。

では、どの検定を選択するかである。私たちがたどった紆余曲折の道を紹介したい。

平均値差の検定でよく知られているのは、「t検定」である。ところが、統計学初心者も知る通り、t検定など帰無仮説検定は、サンプル数（「サンプルサイズ」ともいう）の影響を受け、サンプルサイズが大きくなるにつれて、有意になる傾向がある。全国100万人弱とか、宮城県2万人弱といったサンプルサイズなら、実質的な平均値差がなくても有意になってしまう恐れがある。

こうした不安がよぎるため、私たちはまず、専門家の指導のもと、サンプルサイズの影響が少ないと考えられるコルモゴロフ＝スミルノフ検定を試してみた。

8. 1. コルモゴロフ＝スミルノフ検定（KS検定）

（1）令和3年度、中3数学、トップ都道府県とボトム都道府県のKS検定

「コルモゴロフ＝スミルノフ検定（Kolmogorov-Smirnov test：KS検定）」とは、「2つの標本についてそれぞれの母集団の確率分布が一致しているかどうか、またはある標本の母集団の確率分布が帰無仮説で提示

⁵⁹ 大久保・岡田, 2012, p.28.

⁶⁰ 大久保・岡田, 2012, p.28.

⁶¹ 大久保・岡田, 2012, p.28.

した確率分布と一致しているかどうかを検定すること⁶²である。「標本」「母集団」「確率分布」「帰無仮説」といった統計学の専門用語に面食らってしまうが、データサイエンスを実践しながら、これらの専門用語を少しずつ体得していくことにしよう。初心者の一歩として、KS検定は、2つの集団(2群)の平均値差の有無を検証するテストと理解し、先に進もう。

KS検定のために、統計ソフトR⁶³のks.test関数を利用する。

データには令和3年度を選び、トップ都道府県とボトム都道府県の平均正答率の差の一覧(表3)から、差が最も大きい中3数学を選ぼう。表3右下のグレーで網掛けしたセルで、11ポイント差(0.9問差)である。トップは石川県、ボトムは沖縄県であった。

表3: R3_EL_JH_トップ都道府県とボトム都道府県の平均値

	トップ平均正答数	ボトム平均正答数	トップ平均正答率	ボトム平均正答率
EL_Jpn	10.0/14	8.6/14	71%	61%
EL_Math	11.9/16	10.8/16	74%	67%
JH_Jpn	9.6/14	8.4/14	69%	60%
JH_Math	10.1/16	8.2/16	63%	52%

データは、国立教育政策研究所のウェブサイトの「令和3年度 全国学力・学習状況調査【都道府県別】および【指定都市別】調査結果資料」に公開されており、このページ⁶⁴から、石川県と沖縄県の「中学校」「調査結果概要」をダウンロードする。石川県のJH_Mathの例を示すと図4になる。

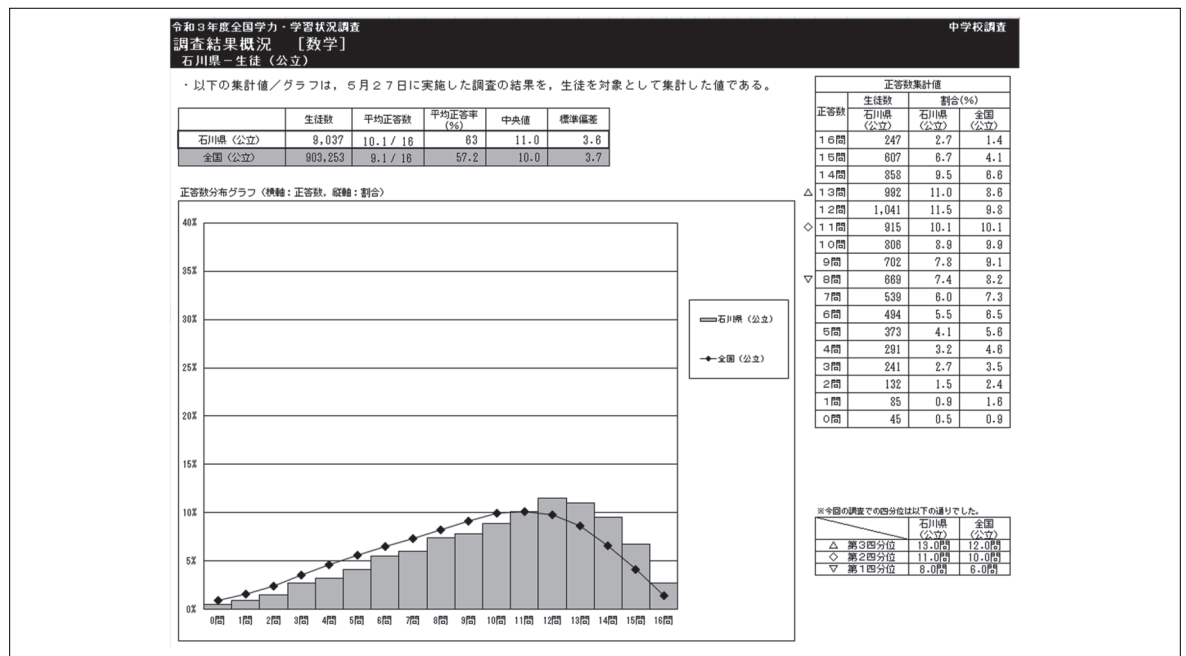


図4: R4_JH_Math_石川県の調査結果概要

⁶² 「BellCurve統計WEB」サイトの「統計用語集」より引用。

<https://bellcurve.jp/statistics/glossary/1360.html>

⁶³ <https://www.r-project.org/>

⁶⁴ <https://www.nier.go.jp/21chousakekkahoukoku/factsheet/prefecture-City.html>

図4の右肩の「正答数集計値」の「割合 (%)」の数値を使う。ボトム都道府県も同様である。Rには、0問正答の生徒割合0.5%から順に、各正答数の割合をベクトル化し、「x」という暫定的な変数に代入する。ボトム都道府県は、「y」の変数に代入する。

Rの入力画面「コンソール」のスク립トは、表4の上段のxとyの表記になる。

表4：R3_JH_Math_トップ都道府県とボトム都道府県のKS検定R、スク립ト

```
> x <- c(0.5, 0.9, 1.5, 2.7, 3.2, 4.1, 5.5, 6.0, 7.4, 7.8, 8.9, 10.1, 11.5, 11.0, 9.5, 6.7, 2.7)
> y <- c(1.5, 2.4, 3.1, 5.2, 6.7, 7.6, 7.9, 8.4, 8.2, 9.2, 8.6, 8.5, 8.1, 6.1, 4.9, 2.9, 1.0)
> ks.test(x, y)
```

Exact two-sample Kolmogorov-Smirnov test

```
data: x and y
D = 0.23529, p-value = 0.7307
alternative hypothesis: two-sided
```

あとは、ks.test(x, y)とコンソールに入力すれば、コンピューターがKS検定を実行してくれる。その結果は表4の下段「Exact two-sample …」以下になる。「p-value = 0.73」とある。私たちの棄却域は慣例にならって5%未満つまり $p < 0.05$ としているため、帰無仮説は棄却できず、私たちは帰無仮説を採択する。つまり、この2つの平均に有意な差があるとは言えない、という結果になる。

筆者は、同様のKS検定の結果を2021年2月発表の拙稿に記載した。データは、平成31年度小6国語、平成19年度中3数学Bのトップ都道府県とボトム都道府県と全国平均である。いずれの検定でも、 $p < 0.05$ で帰無仮説を棄却できるケースはなかった⁶⁵。

(2) 令和3年度、中3数学、トップ都道府県とボトム都道府県の正答数分布グラフ

私たちは「新しいタイプの直観」を模索している。

先のxとyのベクトルについて、さらに直観に訴えるため、xとyの正答数分布グラフを描いてみよう。Rのスク립トは、表5になる。

表5：ベクトルxとベクトルyの分布グラフ、スク립ト

```
# 正答数分布グラフ
plot(x, col="red", type="b",          #色指定、点と折れ線指定
     xlim=c(0, 17), ylim=c(0, 12),  #メモリが重ならないよう軸目盛の指定
     ann=F)                          #ラベル表示抑制
par(new=T)                            #グラフィックパラメタ「重ね描き」
plot(y, col="blue", type="b",
     xlim=c(0, 17), ylim=c(0, 12),
     xlab="正答数", ylab="生徒割合%") #ラベル
```

⁶⁵ Cf., 田端健人, 2021「全国学力・学習状況調査の平均正答率をどう受けとめるべきか? —『生きられた数値』による〈子ども教育データサイエンス〉の構想—」『学ぶと教えるの現象学研究』19, pp.4-5. 本稿の第2版は、田端, 2022, pp.93-114に所収している。

これをRで実行すると、図5のグラフが得られる。xが赤、yが青である。

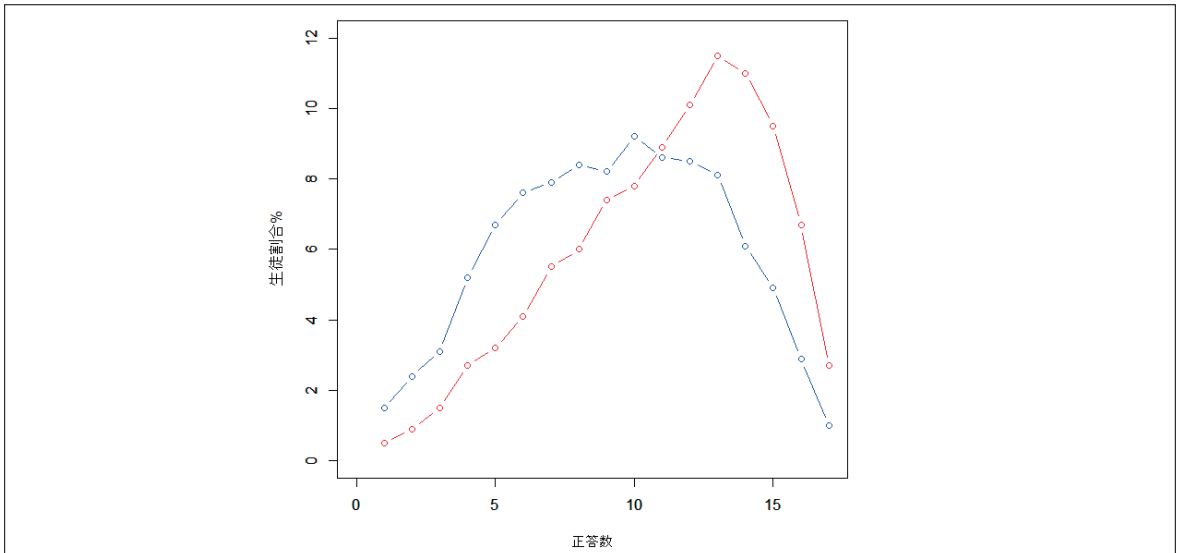


図5：表5のスク립トによる正答数分布グラフ

xとyの分布グラフは確かにズレているものの、KS検定によればこのズレは5%有意水準では有意ではない。

(3) KS検定での不整合と検証のためのシミュレーション

これまでのKS検定により、トップ都道府県とボトム都道府県の平均値には有意差がない、と決着がついたようにも見える。しかし、帰無仮説検定による判断は、必ずしも正しいとは限らないという先の警告を思い出そう。2種類の過誤がある。特に、帰無仮説検定はサンプル数の影響を受けるというよく知られた問題を思い出そう。

KS検定は、2群の分布の差を検定することができるが、ある群と正規分布との差を検定することもできる。これまでの調査研究で、筆者たちはいくつかの自治体から、全国学力・学習状況調査のデータを貸与いただき、いくつかの学校の正答数分布が正規分布と有意な差がないか調べてきた。そうしたところ、児童数14名の小学校でも $p = 0.06$ となり、帰無仮説を採択する結果となったことがあった⁶⁶。そのヒストグラムは、正規分布とはとても思えないような形状だった。現象学的データサイエンスでは、意味と直観の整合的な連関が重要であり、このケースはこの連関に不整合をもたらすことになった。

つぶさに見ると、この不整合は、KS検定の定義においても実は生じていたことが判明する。初心者の理解として単純化してとらえた部分である。再度引用しよう。KS検定とは、

2つの標本についてそれぞれの母集団の確率分布が一致しているかどうか、またはある標本の母集団の確率分布が帰無仮説で提示した確率分布と一致しているかどうかを検定すること

である。「2つの標本についてそれぞれの母集団の確率分布が」とある。xとyは「標本」であり、KS検定は

⁶⁶ Cf. 田端, 2022, p.124.

これら標本から、 x の「母集団」と y の「母集団」を推測する「推測統計」である。しかし、 x は石川県中3の全数であり、 y は沖縄県中3の全数である。これらを「標本」として推測しようとする2つの「母集団」とは何か、得体が知れない。

さらにKS検定を定義する数式を見ると、ここでも意味の綻びが現われる。

KS検定の検定統計量 D は、

$$D = \max_x |F(x) - F_n(x)| \quad (\text{式1})$$

で定義される。この D 値とサンプルサイズ (n_1, n_2) を用いて、

$$KS = |D| \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad (\text{式2})$$

の式から統計量 KS を求め検定を行う⁶⁷。やはり、サンプル数 (n_1, n_2) が使われている。しかし、先のベクトル x と y の計算で、石川県や沖縄県の受検者数は入力していない。では、KS検定の「サンプル数」とは、何のことなのか。Rはどのような計算をして、KS検定を行っているのだろうか。

この問いへの答えは、KS検定を熟知するよほどの専門家でないならば知られていないであろう。私たち初学者の一步もいったん踏み出してみると、このような専門的でレアな領域に踏み入ることになる。これは現象学的データサイエンスがいかにユニークかを示す証左かもしれない。

意味規則の連関を拡大するためにも、Rというツール（道具）の基本的な使い方に触れておきたい。

Rでは様々な関数を利用するが、その関数がどのようなものかを確認する方法がいくつかある。もっとも使い勝手が良いのは、関数のヘルプである。関数の前に「?」を入れると、ヘルプ画面が表示される。「?ks.test」で出力されるヘルプには、「Usage (使い方)」「Arguments (引数)」「Details (詳細)」「Value (関数が返す値)」などの最後に「Run Examples (実行例)」が示されている。ただこのヘルプには、どのような数式で計算しているかは明示されていない。

次に、やや高度になるが関数の「リソース・コード」を「R-Forge」のサイトで見る事ができる。チーム内外の専門家の助言を得て、KS検定のソース・コードを開いてみた⁶⁸。このソース・コードの読み取りも専門的であるが、163行目に、

```
n <- n.x * n.y / (n.x + n.y)
```

という式がある。これは上記の式2を、プログラミング言語で書いたものである。

$n.x$ や $n.y$ がどう定義されているかであるが、この式の前の155行目に、

```
n.x <- as.double(n)
```

とあり、この n がさらに前でどう定義されているかを見ると、

```
n <- length(x)
```

とある。 $length(x)$ は、ベクトル x の長さ（次元数）、つまりベクトル x に代入した値の数、私たちの場合で言えば、正答数割合の個数、つまり正答数0問からすべて正答16問までの17という数値である。 $n.y$ も156行目に、

```
n.y <- length(y)
```

とある。私たちの検定の場合、 $n.y$ の値も17となる。

この数式と、ヘルプの実行例とは整合的である。実行例の最初に表6のスク립トがある。

⁶⁷ <https://bellcurve.jp/statistics/glossary/1360.html>

⁶⁸ <https://r-forge.r-project.org/scm/viewvc.php/pkg/ks.test/R/ks.test.R?view=markup&root=ks-test&pathrev=54>

表6 : ks.testの実用例 (Rのヘルプより)

```
x <- rnorm(50)
y <- runif(30)
Do x and y come from the same distribution?
ks.test(x, y)
```

rnorm(50)というスクリプトは、正規分布に従う乱数を50個発生させるスクリプト、runif(30)というスクリプトは、0~1の間の数値で乱数を30個発生させるスクリプトである。こうして発生させた乱数(x, y)をks.test関数に代入する場合、n.xは50となり、n.yは30となる。このxとyのベクトルなら、一方は正負の値で他方は正の値であることから、当然差があるはずで、ないほうがおかしなことになる。このスクリプトを実施すると、どんな乱数でも決まって $p < 0.05$ になり、帰無仮説は棄却され、有意差ありとなる。

この実行例を利用して、私たちに必要なKS検定の意味連関を押し広げるシミュレーションを試行できる。n.xとn.yを17にするなら、令和3年度中3数学と同じ問題数の乱数が発生できる。また、正規分布の平均を50、標準偏差を10に設定するなら、私たちがよく知る「偏差値」の正規分布になる。乱数を17個発生させ、x1は偏差値50、標準偏差10の正規分布、y1は同じ標準偏差で、x1よりも偏差値にして5低い集団(偏差値45)を設定し、KS検定のシミュレーションをすることができる。スクリプトは表7になる。

表7 : 平均と標準偏差を指定した17個の乱数のスクリプト

```
x1 <- rnorm(17, 50, 10)
y1 <- rnorm(17, 45, 10)
ks.test(x, y)
```

このスクリプトの意味は、全16問の正答数で乱数を発生させた場合、偏差値にして5の差をKS検定にかけてみるということになる。有意水準はこれまでと同様5%とする。これを試みに10回やってみたところ、 $p < 0.05$ となったのが3回、 $p \geq 0.05$ が7回であった。このスクリプトを実行するごとに、コンピューターはx1とy1に対し、条件に近い新たな17個の乱数を発生させる。繰り返すごとに異なる乱数をコンピューターは選び出す。それが有意水準を下回るか否かは、ほとんど偶然になる(50%の確率)か、経験上やや有意水準を上回る(帰無仮説を支持する)傾向になる。

有意水準5%を下回って有意差ありになったケースのx1とy1のヒストグラムは、図6になる。

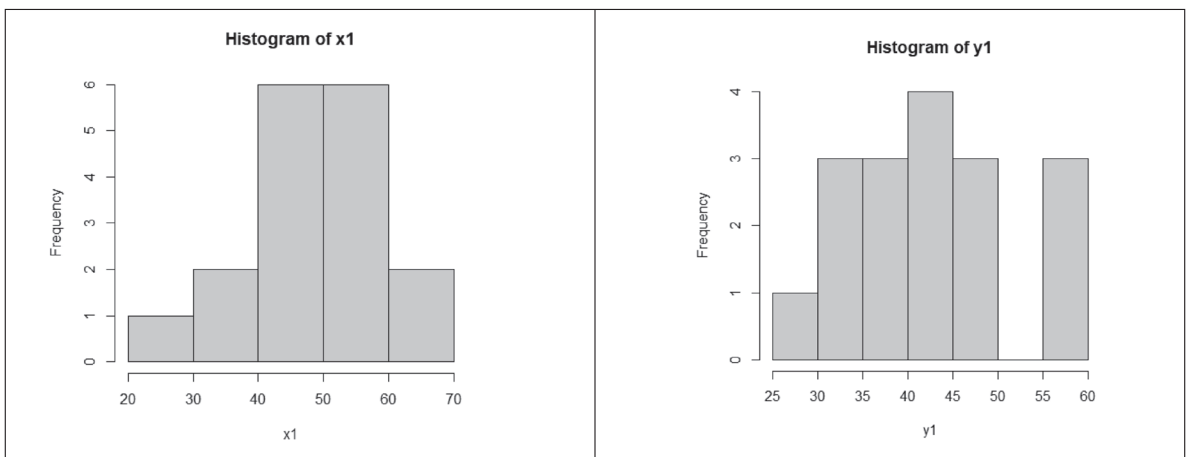


図6 : n=17の正規分布でKS検定で有意差ありのヒストグラム

対して、有意差無し histograms は、図7となる。

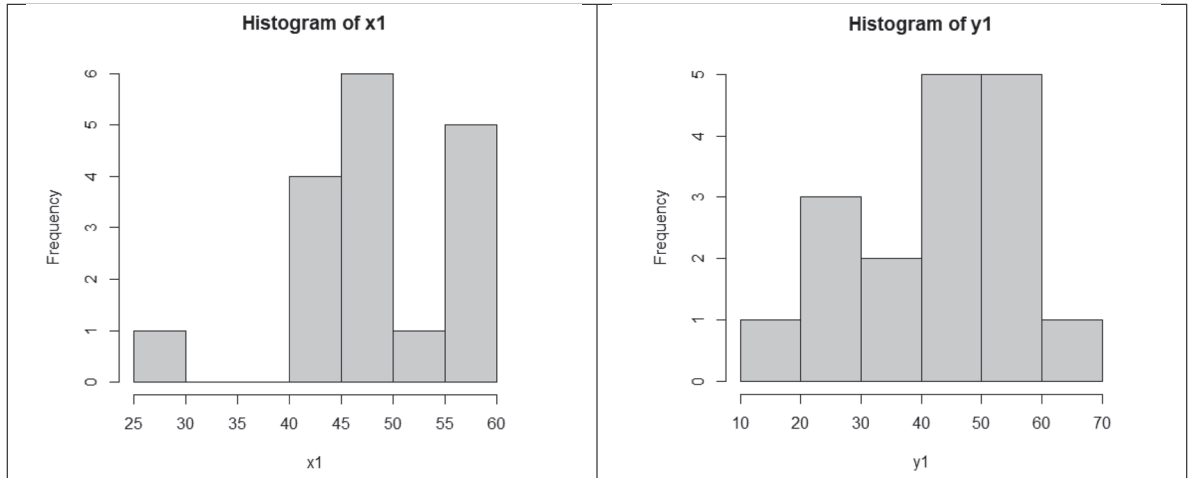


図7：n=17の正規分布でKS検定で有意差なしのヒストグラム

図7のヒストグラムを目視で比較すると、確かに図6よりもx1とy1が正規分布に近づいているように見えるが、図7のx1とy1のヒストグラムに「有意差はない」と言われても、感覚的には「差がある」ように見えてしまう。n=17ではKS検定の識別力は弱いと感じられる。

ではnの個数を増やすとどうだろう。乱数の個数を17から、石川県の中3受検者数とほぼ同じ9,000に変更し、KS検定すると、表8の結果となる。

表8：表7の乱数を1000倍したスクリプトとKS検定の結果

```
> x2<-rnorm(9000,50,10)
> y2<-rnorm(9000,45,10)
> ks.test(x2,y2)

Asymptotic two-sample Kolmogorov-Smirnov test

data: x2 and y2
D = 0.19911, p-value < 2.2e-16
alternative hypothesis: two-sided
```

「2.2e-16」という結果は、2.2のマイナス16乗を意味する。ゆえにP値は0.00となり、5%有意水準で帰無仮説は棄却され、2群の分布には有意差ありとなる。この乱数の発生を10回試したところ、10回ともP値は、0.00だった。

表8のx2とy2のヒストグラムは、図8になる。

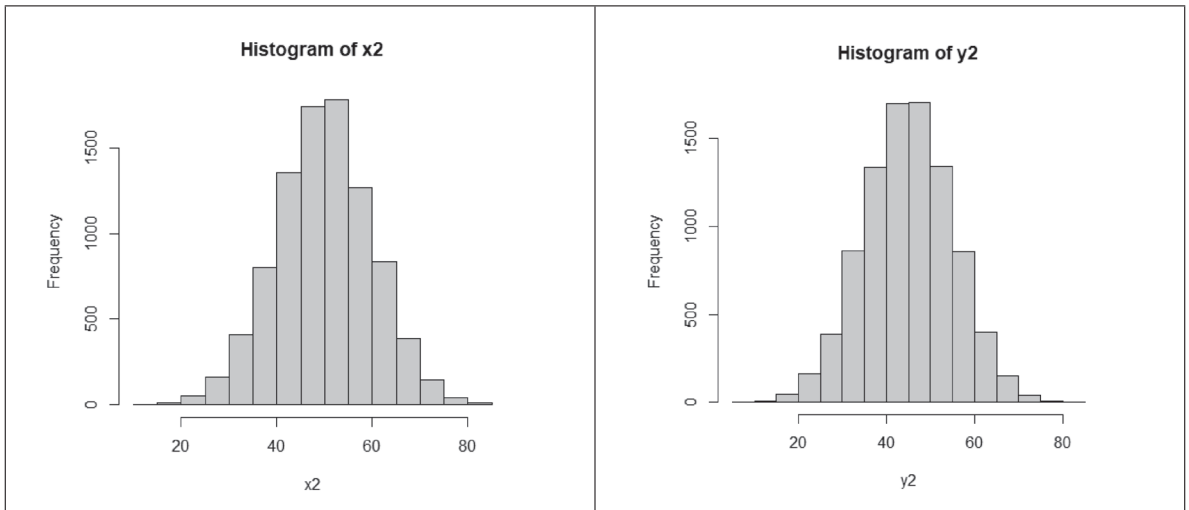


図8：n=9,000の乱数の正規分布（平均50と45、標準偏差10）

図8を見るといずれも平均50と平均45の正規分布になっていることがわかる。分布の形状は同じだが、平均が5だけズレており、このズレによりKS検定では5%水準で有意差ありになる。10回以上繰り返しても有意差ありである。

以上から、KS検定にも、本来のよく知られたサンプルサイズ問題が隠れていることが明らかになった。KS検定の「サンプル数」が17程度なら、そのサンプルが正規分布に従うという条件で偏差値5の差で乱数を発生させたとしても、乱数を発生させる偶然から必ずしも有意水準を下回るわけではなく、むしろ有意水準を上回る傾向にあり、「サンプル数」を9,000ほどに多くすると、同じ条件で偶然に左右される乱数でさえ決まって有意水準を下回るという結果になる。これは、サンプルサイズを大きくすれば、有意水準を下回る傾向にあるという帰無仮説の特性に他ならない。逆に以上の定義の再確認とシミュレーションにより、KS検定の意味と直観の綻びが部分的に修繕されたことになる。

8. 2. t検定とz検定

(1) 「t検定」とDS-EFA「平均値差検定システム」

t検定 (t-test) とは、「2つの異なるグループ (群) のテストの平均値に差があるか」⁶⁹を判定するテストである。あるいは「2標本t検定」とも呼ばれ、「2つの独立した母集団があり、それぞれの母集団から抽出した標本の平均に差があるかどうかを検定すること」⁷⁰である。「母集団から抽出した標本」とあるように、t検定は、標本から母集団を推測する「推測統計」⁷¹である。先の新薬開発の例でいえば、ある病気に対する新薬開発の場合、その病気に罹患した全員 (母集団) を調べるわけにはいかないの、サンプル (標本) としてその病気に罹患した何名かを抽出し、実験群と統制群に分け、その平均差から母集団の差を推測し有意か否かを検定するわけである。

t検定は、2標本の平均、標準偏差、人数がわかれば、計算できる。

⁶⁹ 小林雄一郎・濱田彰・水本篤, 2020『Rによる教育データ分析入門』オーム社, p.55.

⁷⁰ ウェブサイト「BellCurve統計の時間」の「2標本t検定とは」より引用。

<https://bellcurve.jp/statistics/course/9427.html>

⁷¹ 小林ほか, 2020, p.54.

そこで筆者たちのチームDS-EFAはこの検定が簡便にできるウェブ・システムを開発し、2022年5月よりホームページで無料公開している⁷²。開発者は、宮城教育大学の菅原敏教授と筆者である。このシステムは、t検定だけでなく、効果量、信頼区間、2群の分布グラフなども出力する。そこで次に、このシステムを活用し、サンプル数の問題について理解を進めよう。

(2) 帰無仮説検定におけるサンプル数問題

帰無仮説検定はサンプル数が多くなると有意差ありになりやすい（P値が0.05を下回りやすい）ことがよく知られている。100点満点の1点差は、経験的には僅差であるが、サンプル数がどのくらいになれば有意差ありと判定されるかをシミュレーションしてみよう。グループ1の平均を70、グループ2の平均を69とし、標準偏差はいずれも20としてみよう。両グループのサンプル数を3000にしても、 $p \geq 0.05$ となり、差はないと評価される。3,000からグループ2のサンプル数を徐々に上げていくと、 $n=3,153$ ではじめて $p < 0.05$ となり、有意差ありになる（図9）。

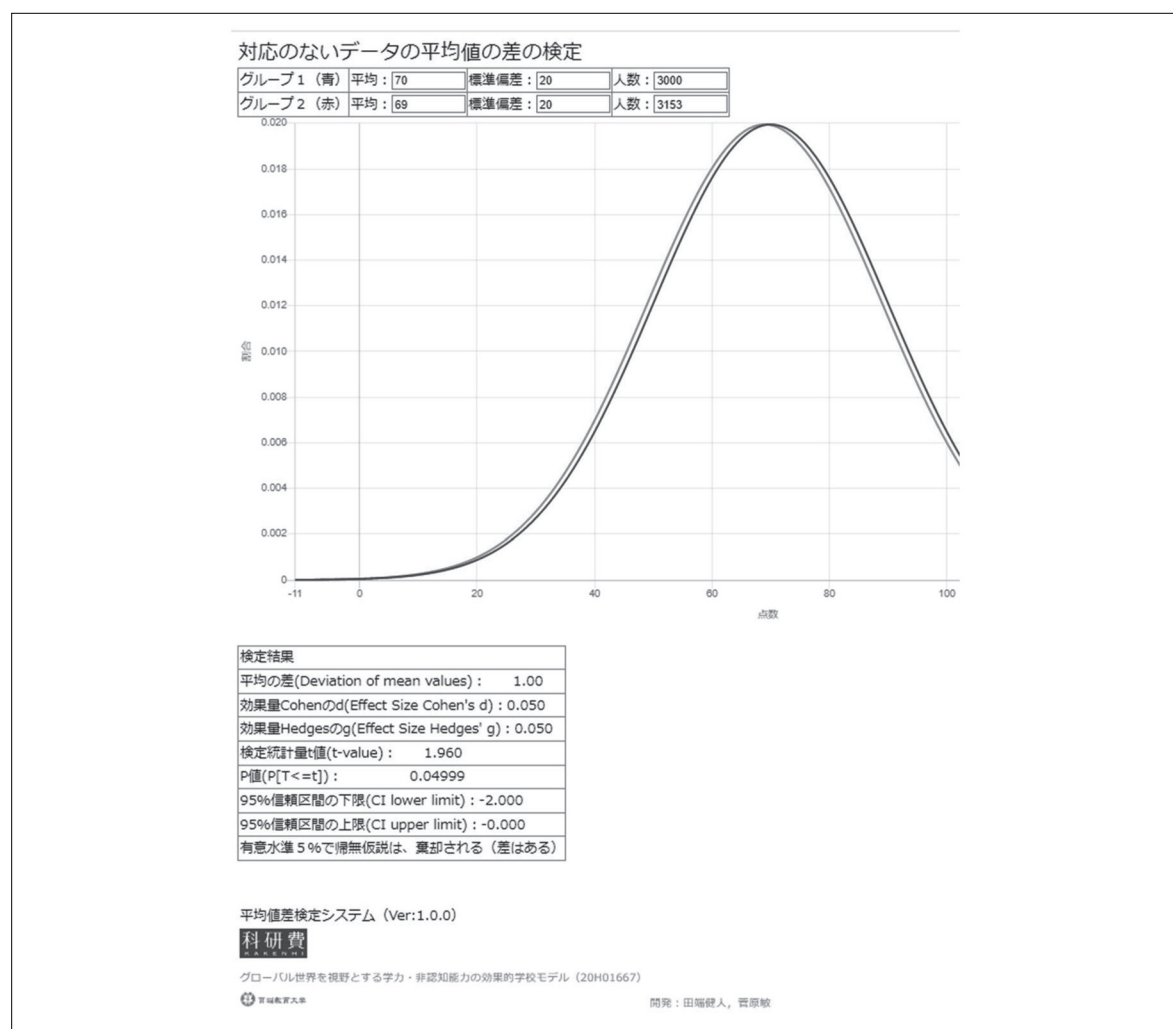


図9：平均値差検定システムによるシミュレーション

⁷² <https://ds-efa.info/cohensd/>

この僅差に有意差ありは直観的におかしい。図9のグラフの重なり具合からも、いくら統計学的に有意差ありと評価されても、素人感覚からは到底納得できない。しかも同じ平均差で、サンプル数が小さければ有意差なしで、大きければ有意差ありというのも、意味的直観的に不整合である。全国学力・学習状況調査の各都道府県の受検児童生徒数は、少なくとも5,000名以上はいるため、t検定なら平均正答率1ポイント差でも有意差ありとなり、このシミュレーションからもわかるように、t検定はまったく役に立たない。

(3) z検定

全国学力・学習状況調査は悉皆調査であるため、実は母集団を推測する必要はない。KS検定やt検定など推測統計の意味の不整合は実はここにもあった。母集団の平均と分散、つまり母平均と母分散が既知の場合には、z検定(z-test)が使える⁷³。筆者は既にz検定についても、全国学力・学習状況調査の文脈で検討し、t検定と同じくサンプル数問題が避けられないことを指摘した⁷⁴。

以上のように、ポイント差2~5ポイントは大きいのかに白黒決着をつけたいとの思いから、帰無仮説検定を試みてきたが、どの検定もサンプル数問題が障壁となり、その先に進むことができなくなった。そこで筆者たちが行きついた打開の道が「効果量」である。

9. 効果量

9. 1. 帰無仮説検定から効果量へという世界的動向

帰無仮説検定から効果量へという筆者たちがたどった試行錯誤の歩みは、実は統計学や心理学の世界的動向と軌を一にするものであった。

大久保街亜氏と岡田謙介氏によると、英語圏の心理学では、帰無仮説検定にかたよった旧来の統計解析に、1990年代から2000年にかけて「統計改革」が起こり、効果量や信頼区間(confidence interval, 以下CIと略)を併せて報告する研究が増えている⁷⁵。アメリカ心理学会(American Psychological Association)の出版マニュアルも第6版(2009年刊)からは、「帰無仮説検定は統計的分析のはじまりに過ぎず、効果量やCIなどを併せて記載することが結果を適切に報告するために必要である」⁷⁶(cf. 大久保・岡田, 2012, p.11)と記されるようになった。心理学だけでなく、統計学の領域でも、アメリカ統計学会(American Statistical Association)は2016年に声明を発表し、「P値や統計的有意性は、効果の大きさと結果の重要性を意味しない」とし、P値以外のアプローチとしてCIやベイズ統計などを推奨している⁷⁷。

対照的に日本では、少なくとも現在から10年前の時点では、「効果量を重視する流れはほとんどない」⁷⁸状況だった。その後、2019年7月から2020年6月までの国内学術誌『教育心理学研究』と海外学術誌 *Journal of Personality and Social Psychology* 掲載論文で、「効果量、信頼区間」の記載の有無を調べたところ、前者ではいずれも記載されていない論文が全体の44%を占め、後者では論文のすべてにいずれかの記載があったと

⁷³ ウェブサイト「BellCurve統計の時間」の「z検定」を参照。

<https://bellcurve.jp/statistics/glossary/1841.html>

⁷⁴ Cf., 田端, 2022, pp.124-127.

⁷⁵ Cf., 大久保・岡田, 2012, p.i, pp.5-6.

⁷⁶ Cf., 大久保・岡田, 2012, p.11.

⁷⁷ Wasserstein, R.L. and Lazar, N.A., 2016 「統計的有意性とP値に関するAPA声明」 藤俊哉訳, pp.2-3. <https://www.biometrics.gr.jp/news/all/ASA.pdf> [2022.09.28最終閲覧]

⁷⁸ 大久保・岡田, 2012, p.18.

いう⁷⁹。このように、日本では効果量は2020年現在でもまだ十分に浸透しているわけではない。教育学の領域ではなおのことである。試みに、筆者が属する学会の学会誌『教育学研究』と『教育方法学研究』とを過去20年遡ったところ、前者では⁸⁰、「効果量」やそれに関連する用語をタイトルとする論文はわずか2本しかなかった⁸¹。その2本も効果量d値には触れていない。後者では⁸²、「効果量」をテーマとする論文は皆無であった。こうした状況にあって、効果量を利活用した学習の可視化を提示したジョン・ハッティの研究、また効果量を日本の教育関係者に知らしめたハッティの邦訳書『学習に何が最も効果的か—メタ分析による学習の可視化—』（あいり出版、2017年刊）は先駆的な業績である。

9. 2. 効果量 (Effect Size) とは何か？

効果量とは、「効果の大きさをあらわす統計的な指標」⁸³のことである。例えば、同じ学力の2学級があり、同じ教師がある単元を指導したとして、一方の学級は対話活動を取り入れ、他方の学級は取り入れなかった場合、単元の終わりにテストをして前者が後者よりも得点が高かったとすると、その差は対話的学習の「効果」と評価できる。この「効果」を数量化したのが「効果量」である。

もう少し専門的に解説すると、効果量とは、「群間での平均値の差の程度、変数間の関連の強さなどを、データの単位に左右されないよう標準化したもの」⁸⁴である。

この説明にあるように、効果量には2つのタイプ、つまり群間の平均値差の程度を表わす指標と、変数間の関連の強さを表す指標とがある。前者はd族 (d family) の効果量、後者はr族 (r family) の効果量である。d族にはCohenのd値やHedgesのg値があり、r族には相関係数 (ピアソンの積率相関係数) r値などがある。2変数の関連の強さ (相関係数) を「効果量」と呼ぶのは日本語として違和感があるが、一方と他方の「影響 (effect)」関係の強さと考えれば、理解しやすい。

またこの説明にあるように、効果量は「データの単位に左右されないよう標準化」されていることも重要である。「標準化されている」とは、例えば100点満点で採点したテストで測定した効果量と、4件法で質問した項目スコアで測定した効果量とを、同一尺度上で比較できるということである。この特性を生かせば、経年比較できない設計の全国学力・学習状況調査のスコアを便宜的に経年比較することも不可能ではない。

効果量の最大の強みは、サンプルサイズの影響をほとんど受けにくいことである。

現象学的教育データサイエンスでは、d族の効果量として、今のところ、Cohenのdを活用する。私たちが開発したシステム「平均値差検定システム」にはCohenのdとHedgesのgの値を返すが、利用者はCohenのd値を見ればよい。前者と後者の違いは、記述統計か推測統計かの違いである。標本から母集団を推測する推測統計は今のところ必要ない。

Cohenのdの計算式は以下となる⁸⁵。グループ1の平均値を M_1 、人数 (サンプルサイズ) を n_1 、分散を S_1^2 (正

⁷⁹ Cf., 山内香奈, 2021 「『教育心理学研究』における統計改革の現状—サンプルサイズ設計を中心に—」『教育心理学年報』第60集, p.122.

⁸⁰ 『教育学研究』第89巻第1号 (2022年3月) から第69巻第2号 (2002年6月) までを調べた。

⁸¹ 川口俊明, 2006 「学力格差と『学校の効果』—小学校の学力テスト分析から—」『教育学研究』第73巻第4号 (pp.350-362) と志水宏吉, 2006 「学力格差を克服する学校—日本版エフェクティブ・スクールを求めて—」『教育学研究』第73巻第4号 (pp.336-349) である。

⁸² 『教育方法学研究』第47巻 (2021年) から第27巻 (2001年) までを調べた。

⁸³ 大久保・岡田, 2012, p.44.

⁸⁴ ウェブサイト「「BellCurve統計の時間」の「効果量」より引用。

<https://bellcurve.jp/statistics/glossary/1314.html>

⁸⁵ 大久保・岡田, 2012, p.55を参照し、平均差は絶対値とした。

確には「標本分散」、グループ2の平均値を M_2 、人数(サンプルサイズ)を n_2 、分散を S_2^2 とする。

$$d = \frac{|M_1 - M_2|}{S_p} \quad (\text{式3})$$

$$S_p = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2}} \quad (\text{式4})$$

Cohenのdの計算式(式3)は、2群の平均値差を「プールした標準偏差(S_p)」(式4)で割っている。

式4からわかるように、効果量もサンプルサイズの影響を全く受けないわけではない。しかし、 n_1 と n_2 の数は、値の大きさではなく、両者の比の問題であることが式4からわかる。つまり、2群のサンプル数は、その比が各群の分散を平均する際の重みづけになっている。このことは、 $n_1 = x$ 、 $n_2 = ax$ という代入、つまり、グループ2の人数がグループ1の人数のa倍というケースを想定してみるとわかりやすい。このケースを式4に代入すると、

$$S_p = \sqrt{\frac{x S_1^2 + ax S_2^2}{x + ax}} \quad (\text{式5})$$

となり、この式5は、

$$S_p = \sqrt{\frac{(S_1^2 + a S_2^2)x}{(1+a)x}} \quad (\text{式6})$$

と書き換えることができ、結果的に、

$$S_p = \sqrt{\frac{S_1^2 + a S_2^2}{1+a}} \quad (\text{式7})$$

と x (サンプルサイズ)を消すことができ、 n_1 と n_2 の比が式に関与していることがわかる。

ちなみに、この式3と式4を一つに統合した式を、コンピューターで計算しやすいよう、標準的なプログラム言語で記述すると、

$$d = \text{abs}(m1 - m2) / \text{sqrt}((n1 * s1^2 + n2 * s2^2) / (n1 + n2)) \quad (\text{式8})$$

となる。式8の分母 $\text{sqrt}((n1 * s1^2 + n2 * s2^2) / (n1 + n2))$ を関数としてエクセルのセルに入力すれば、簡単なシミュレーターが作れるので、 $n1$ や $n2$ のセルに様々な数値を入れ込んで、プールした標準偏差がどのように変化するかをシミュレーションしてみたい。サンプル数のボリュームではなく、サンプル数の比のことが腑に落ちるはずである。

現象学的視点から非常に興味深いのは、Cohenのdは相関係数 r に次の式により変換できることである⁸⁶。

$$r = \frac{d}{\sqrt{d^2 + 4}} \quad (\text{式9})$$

変換できることがなぜ現象学的に興味深いかというと、 d 族の統計的世界と r 族の統計的世界とが存在秩序を保ち調和するからである。両者の地平が融合するとも言えるだろう。実際にデータを扱っていると、この変換によって、意味連関が格段に拡張し、新しいタイプの直観が得られる経験が圧倒的に増加する。

9. 3. 効果量の基準値

上記のような計算で算出される効果量 d 値や r 値を、どのような目安で評価すればよいだろうか。これは「基準値」という重大な問題であるため、章を改めて検討したい。ここでは、効果量の一般的な基準値を紹

⁸⁶ Cohen, J., 1988, *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition*, Routledge, p.23.

介しておく。

Cohenのdの開発者ヤコブ・コーエン（Jacob Cohen）は表9の基準値を提唱している⁸⁷。

表9：Cohenのdの基準値

d = 0.2 (効果小) / d = 0.5 (効果中) / d = 0.8 (効果大)

この基準値には、「一定の有効性」があるとされ、「研究遂行前に、これは差がありそうだ、とか、厳しそうだ、という直感が働くことがあるが、その直感は概ねCohenの基準の大小に沿うのではないか」という経験知も紹介されている⁸⁸。この「直感」は不思議である。本稿でいう新しいタイプの直観とオーバーラップする。一方、コーエンの基準に否定的な見解⁸⁹もある。先のハッティは、独自の観点から、d = 0.40を基準値としている。彼によれば、これは学校で1年間に達成できる平均的な値である⁹⁰。さらに「分野ごとの基準が必要である」とする見解もあり、「外国語教育研究の分野において、Plonsky and Oswald(2014)は、… (略) …対応なしの場合、d=0.40 (効果量小)、d=0.70 (効果量中)、d=1.00 (効果量大)、対応ありの場合、d=0.60 (効果量小)、d=1.00 (効果量中)、d=1.40 (効果量大) という基準を提唱して」いる⁹¹。

このように、効果量の基準値については、種々の議論がある。ただ、上記の議論から共通して見えてくるのは、ハッティのd=0.40は効果量としては必ずしも大きくはないということである。というのも、この基準値はCohenが中程度とする0.50に届かず、Plonsky and Oswaldでは対応のない2群でも効果小とされるからである。コーエンの0.50にしても、Plonsky and Oswaldの「中」に届かない。

次に、相関係数rの一般的な基準値は、表10になる。

表10：相関係数rの基準値

$0.7 \leq r \leq 1.0$	： 強い相関
$0.4 \leq r < 0.7$	： 中程度の相関
$0.2 \leq r < 0.4$	： 弱い相関
$0.0 \leq r < 0.2$	： 相関なし

相関係数rは、-1から1までの値をとり、絶対値にして表10が目安とされている。

先述のようにdはrに変換できるので、コーエンの基準値と相関係数の一般的な基準値とを対応させると、表11になる。

表11：dとrの対応表

効果量dの基準	d	r	相関係数rの基準
	1.95	0.70	大
	0.90	0.41	中
大	0.80	0.37	
中	0.50	0.24	
	0.40	0.20	小
小	0.20	0.10	

⁸⁷ Cohen, 1988, p.82.

⁸⁸ 村井潤一郎・橋本貴充, 2018「統計的仮説検定を用いる心理学研究におけるサンプルサイズ設計」『心理学評論』Vol.61, No.1, p.123.

⁸⁹ Cf., 村井・橋本, 2018, p.124.

⁹⁰ Cf., ハッティ, J., 2018『学習に何が最も効果的か—メタ分析による学習の可視化：教師編—』原田信之訳者代表, あいり出版, p.2.

⁹¹ 小林ほか, 2020, p.93.

dとrの大中小はズレている。コーエンが「大」とする $d=0.80$ も一般的なrの基準値では「中」に届かない。dの意味連関とrの意味連関は単純に接続できるわけではない。しかし意味の裂け目にかける橋がないわけではないだろう。

表11からわかるのは、コーエンの「大中小」は、相関係数にすれば、「中」に届かない小さな効果の区分である。rで0.40未満の小さな相関の効果を微細に区分するのがdの基準値とみなすことができる。相関係数0.20に届かないd値(0.40未満)は、実質的な効果はないとみなして良いだろう。とすれば、実質的な平均差dの基準を、0.40とするか0.50とするかになる。ハッティのメタ分析との接続を考えれば $d=0.40$ も実用的であり、先に紹介した経験知からすればコーエンの $d=0.50$ も汎用性が高い。

この難問は、理屈から演繹的に解を導ける問題ではない。先の有意水準5%にしても、その開拓者フィッシャーが「偏差が有意と考えられるかどうかを判定する限界として、便宜的にこの値〔 $P=0.05$ 〕をとるものとすれば」⁹²と記しているように、論理的に導かれた解ではなく、便宜的に設定した仮定(仮説)であることがわかる。この仮説が以後多くの研究者の共感を得て、あるいは惰性で受け継がれ自明となった値である。「基準値」というこの不思議な数値について次に考察しよう。

10. 基準値という難問

10. 1. 「レギュラトリーサイエンス」と「データサイエンス」

基準値の問題がいかに重大かを理解するには、この問題解決のために一つの新しい科学、「理系的知識のみならず文系的素養も必要とする新しいタイプの科学」⁹³が存在することを指摘すれば十分であろう。

基準値は世の中のいたる所に溢れている。食品の消費期限や賞味期限、水道水の基準値、大気汚染の基準値、放射性物質の基準値、生態系保全の基準値、危険物からの距離の基準値、交通安全の基準値など、挙げればきりが無い。「お酒は20歳から」も「18歳成人」も基準値である。基準値の多くは生きられる数値である。

基準値は数値でありながら非常に人間的であり、主観の印象や感覚あるいはコミュニティの寛容度などと不可分である。このことは、食品の消費期限と賞味期限の基準値がどのように決められるかを見るとよくわかる。

日本のガイドラインでは、「期限設定のための保存検査」が実施されるようだが、それは①理化学的試験、②微生物学的試験、③官能検査で構成されている⁹⁴。①と②は科学的に測定できるが、③は「人間の視覚、味覚、嗅覚などの感覚を通して、色や香り、食感などを一定の条件で評価する」⁹⁵。それゆえ「食品の『おいしさ』を決めるのは、機械ではなく結局は人である」⁹⁶。これはデータサイエンスが「人間的」であらざるをえないことと同じである。

官能検査は、知覚的直観の領域である。しかし、フッサールやメルロ＝ポンティの知覚の現象学は、消費や賞味の期限を判定する人間の知覚を明らかにしはしない。明らかにするどころか問題にさえしなかった。官能検査の基準値は「食べて安全な限界」とか「食材の味が損なわれない限界」である。この官能検査をさら

⁹² フィッシャー, R. A. 『研究者のための統計的方法』遠藤健児・鍋谷清治訳, 森北出版, p.34.〔〕内は引用者。

⁹³ 村上道夫・永井孝志・小野恭子・岸本充生, 2019 『基準値のからくり—安全はこうして数字になった—』講談社, pp.9-10.

⁹⁴ 村上ほか, 2014, p.27.

⁹⁵ 村上ほか, 2014, p.27.

⁹⁶ 村上ほか, 2014, p.27.

に鋭くしたのが、私たちの日常的な知覚的直観である。私たちの日常的な知覚的直観はしばしば、「食材が食べて安全かどうか」よりもさらに厳しく、「食材がおいしいか否かの限界」に照準を絞っている。消費や賞味期限内の肉や魚や野菜や果物を私たちは店頭で見るが、私たちがえてして目を光らせるのは、そうした食材の鮮度である。魚の鮮度の知覚は、机やサイコロを見たり、そこから視線を移す知覚作用とは、質的に全く異なる。机やサイコロを見るという知覚作用をいくら深く解明しようと、食材の鮮度を見抜く知覚の解明にはいたらない。ハイデガーの道具連関の分析も、食材の鮮度の知覚解明にはいたらない。そもそも食材は道具ではない。この一件だけとっても、知覚の現象学や道具的世界の現象学は、生活世界を全面的に解明しているわけではないことがわかる。いずれも生活世界の究極的な記述というわけではない。それだけ現象学の伸び代があるとも言える。食材の鮮度は、見れば「わかる」。この「わかる」は推測に似ているが、認知的推測ではなく、信念に近い。食材を見れば、私たちは「この食材は新鮮だ」という信念なり感覚を抱き、購買意欲をそそられる。この信念なり感覚が確証されるのは、それを食べてみた時である。味覚が「やっぱり新鮮だった」という確証や、「新鮮そうに見えたのに新鮮ではなかった」という反証を与える。

以上の検討から明らかになるのは、基準値は安全と危険の限界の一線にあり、生活世界の豊かさに関わるのは、食材の鮮度のように、基準値よりいっそう厳しい値であるということである。教育データサイエンスに当てはめると、安全と危険を区分けし、アラートを鳴らす「基準値」は、実践の豊かさを区分けする基準よりも高めに設定するのが良いということである。「効果量の基準値は0.40か0.50か」という問いでは、基準値としては0.50の方が相応しいし、実践の豊かさを微細に評価するには0.40とか0.30とかでも良いかもしれない。そもそも効果量 d の一般的な基準値は、相関係数 r からすればかなり微細な効果を評価する。このことを考慮すれば、0.50の基準値も相当にミニマムな基準であり、もう少し高くてもよいくらいである。アラートという社会的な意味を持つ線引きとしては、0.40より0.50の方が妥当である。

ちなみに基準値の専門家である村上道夫氏たちは、基準値を相手どる新しい科学を「想定・評価・判断をとまなう科学」とし、「レギュラトリーサイエンス」と呼んでいる⁹⁷。「レギュラトリー」には、「規制」や「調整」の意味が込められている⁹⁸。レギュラトリーサイエンスの特徴としては、先に引用したように理系的知識と文系的知識の融合に加え、測定・分析された数値と主観的な感覚や推定との融合があり、さらに「従来型の科学で得られたデータや知見と、政治や行政による規制・調整・政策判断などとの間にある大きなギャップを埋める『橋渡し』の役割」⁹⁹とか、「専門的な知見と市民の感情」¹⁰⁰との組み合わせとかがある。レギュラトリーサイエンスは、多種多様な知や情や意が入り混じり、それらのバランスをとるハイブリッドなアリーナである。教育データサイエンスは、基準値の問題においてレギュラトリーサイエンスと接合する。

10. 2. 効果量の基準値 $d=0.50$ の提案

(1) 社会経済的状況 (SES) の効果量を参照値として

効果量の基準値 $d=0.50$ の妥当性を検証する参照値として、「社会経済的状況Socio Economic-Status (以下「SES」)」をとりあげてみよう。

学力研究の領域では、SESと学力との相関関係の研究が蓄積されてきた。SESと学力とは相関し、SESが高いと学力を押し上げる効果があること、またその逆の効果が実証されてきた¹⁰¹。

⁹⁷ 村上ほか, 2014, p.17.

⁹⁸ 村上ほか, 2014, p.17.

⁹⁹ 村上ほか, 2014, p.17.

¹⁰⁰ 村上ほか, 2014, p.21.

¹⁰¹ 例えば、お茶の水大学, 2018「保護者に対する調査の結果と学力等との関係の専門的な分析に関する調査研究」(https://www.nier.go.jp/17chousa/pdf/17hogosha_summary.pdf) や松岡亮二, 2020『教育格差一階層・地域・学歴—』ちくま新書などを参照。

文部科学省も令和3年度の全国学力・学習状況調査の質問紙調査から、SESの代替指標となる家庭の蔵書数の質問項目を加え、学力との相関係数を公表している¹⁰²。その一覧を示すと表12になる。上段が公表されたr値で、下段はそれを筆者がd値に換算した。

表12：令和3年度文部科学省公開、学力とSESとの相関係数

	小6国語	小6算数	中3国語	中3数学
学力とSESの相関係数r	0.233	0.251	0.220	0.203
学力に対するSESの効果量d	0.478	0.518	0.452	0.414

もちろんSESをどう定義し、どのスコアで測定するか、また個人スコアかグループスコアかで学力との相関係数も変動する。しかし、文部科学省公開の表12のデータは、他の質問項目との相関係数との比較が容易であり、市民の意味連関や直観連関に組み込みやすいため、目安になる値である。

表12を一瞥すれば、SESが学力に与える影響は、効果量dにして0.40から0.50あたりである。「dの基準値は0.40か0.50か」という先の議論と重なり合う。平均を取れば、d=0.47である。この結果も、0.50に軍配があがる。SESの影響力と同等の効果を「実質的な効果」の基準値とするなら、d=0.40では弱く、少なくともd=0.50は必要であろう。

(2) PISA2018の長期トレンド評価を参照値として

さらに、国際学力テストPISAの長期トレンドについてのOECDの評価を参照してみよう。

図10は国立教育政策研究所発表の資料からの抜粋である¹⁰³。この図からわかるように、PISAの読解力(Reading)で日本は、2000年から2003年にかけてスコアを下げ、2006年から2009年、2012年とスコアを上げている。その度にマスコミは「読解力低下」や「読解力向上」を声高に報じてきた。

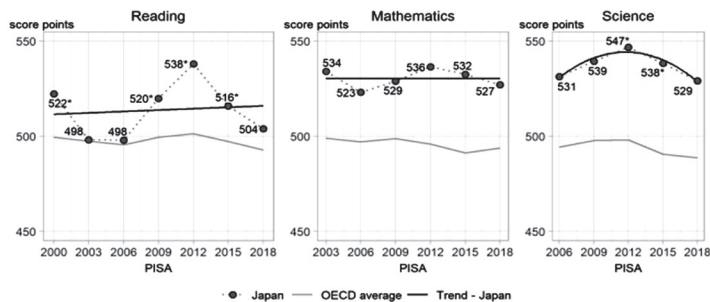


図10：PISA 2000-2018各リテラシーのトレンド

しかし、OECDの長期トレンド分析を国立教育政策研究所は次のように要約し、日本の学力に「有意な」

¹⁰² 国立教育政策研究所の以下のURLから、R3「報告書・調査結果資料」→「3. 令和3年度 全国学力・学習状況調査【小学校】調査資料結果」あるいは「4. 令和3年度 全国学力・学習状況調査【中学校】調査資料結果」→「(3) 相関係数、クロス集計表」の「相関係数(・・・)全国【表】」の順でクリックすると、エクセルファイルが自動ダウンロードされる。その一覧の右2列に教科と項目との相関係数が記載されている。

<https://www.nier.go.jp/kaihatsu/zenkokugakuryoku.html> [2022.09.30最終閲覧]

¹⁰³ https://www.oecd.org/pisa/publications/PISA2018_CN_JPN_Japanese.pdf

変動はないと評価した。

平均得点の2000年～2008年の長期トレンドに関するOECDの分析によると、日本の読解力は、平均得点のトレンドに統計的に有意な変化がない国・地域に分類され、そのうち「平坦」タイプに該当。¹⁰⁴

図10はOECD発表のCountry Noteからの転載である。図10の直後（PDFのp.3）には読解力の経年の変動の不安定さに関する言及があり、OECDは微妙な注釈をつけている。この注釈について筆者は2021年発表の拙稿にて検討したので、ここでは繰り返さない¹⁰⁵。重要なのは、微妙な注釈をつけながらもOECDが図10のグラフに日本の読解力のトレンドとして平坦な棒グラフを引いている点である。平均スコアの上下があるのは当然だが、長期トレンドとしては平坦な棒グラフで表すことができるという評価である。仮に国立教育政策研究所が、OECDの発表を国民向けに「わかりやすく」まとめたとしても、その発表に統計学の専門家も特段の批判をしてはいない。2000年と2003年との平均差、2003年（同点の2006年）と2012年との平均差は、統計的に有意な差ではないというのがOECDと国立教育政策所の長期トレンド評価である。

そうだとすれば、これらの平均値差の効果量を算出すれば、その効果量は大きくないと評価する一つの参照値になるだろう。

日本の2000年の読解力の平均値は522、標準偏差は86¹⁰⁶、2003年は平均498、標準偏差は106¹⁰⁷、2012年は平均538、標準偏差は99である¹⁰⁸。この値を「平均値差検定システム」に入力したところ、2000年と2003年との差の効果量 $d=0.25$ 、2003年と2012年との差で $d=0.39$ であった。 d 値で0.40にも届かないという結果である。

2003年平均と2012年平均との差は図10では大きく見えるが、かつて拙稿で指摘したように、これは「グラフのマジック」であり、450点から550点をクローズアップしているからである。この差は効果量にして0.39である。0.39をOECDや国立教育政策研究所は有意な差ではないと評価している。このことから、アラートとしては、やはり0.40よりも0.50の方が効果量の基準値には相応しい。

以上の検討から、私たちDS-EFAは、実質的な平均値差の基準として、 $d \geq 0.50$ を提案する。

10. 3. 効果量の基準値からみた全国学力・学習状況調査のポイント差

紆余曲折の議論となったが、これだけの検討を経て初めて、全国学力・学習状況調査での「全国平均との2～5ポイント差」が大きいのか小さいのかを判定できる。

先に引用した新聞記事には、「仙台市以外の地域では小中ともに全国平均を2～5ポイント下回り、小学生の算数は全国最下位となった」とあった。そこで、令和3年度の小6算数で、仙台市を除く宮城県の平均

¹⁰⁴ https://www.nier.go.jp/kokusai/pisa/pdf/2018/01_point.pdf 当該PDFのp.3参照。下線は引用者。

¹⁰⁵ その第2版である田端, 2022, p.99を参照いただければ幸いである。

¹⁰⁶ OECD, 2000, *Literacy Skills for the World of Tomorrow: Further Results from PISA 2000*, p.281.
https://read.oecd-ilibrary.org/education/literacy-skills-for-the-world-of-tomorrow_9789264102873-en#page3

¹⁰⁷ OECD, 2003, *Learning for Tomorrow's World: First Results from PISA 2003*, p.444.
https://read.oecd-ilibrary.org/education/learning-for-tomorrow-s-world_9789264006416-en#page1

¹⁰⁸ OECD, 2012, *PISA 2012 Results: What Students Know and Can Do: Student Performance in Mathematics, Reading and Science, Volume I*, p.382.
https://read.oecd-ilibrary.org/education/pisa-2012-results-what-students-know-and-can-do-volume-i-revised-edition-february-2014_9789264208780-en#page1

と、全国平均との差の効果量を出してみよう。仙台市を除く宮城県の平均は10.5 (66%)、標準偏差3.6であり、全国(公立)の平均は11.2 (70.2%)、標準偏差3.5である。正答率で4.2ポイント差である。平均値差検定システムで計算したところ、効果量 d は0.20となる。基準値0.50には及ばず、有意差なしという結果である。「全国最下位」は全国平均と実は僅差なのである。

では、5ポイント差はどうだろう。令和3年度、仙台市を除く宮城県の平均が、全国平均より5ポイント(正確には5.2ポイント)低かったのは、中3数学である。宮城県平均は8.3 (52%)で標準偏差3.7、全国(公立)平均は9.1 (57.2%)で標準偏差3.7である。この平均差の効果量 d は0.22である。やはり僅差である。とりたてて報じるほどの差ではない。

以上の検証から、「全国最下位」とか「全国平均を2~5ポイント下回る」という報道は、取るに足りない差をことさら重大視する点で不適切である。こうした報道は、小さな差を大きな差と思わせ、過度の序列化や競争を煽ることになるため、今後は控えてほしい。今後もこうした報道を継続するのであれば、この差が大きいことを証明してからにしてほしい。

では、先に引用した「仙台市は中学生の国語、数学の平均正答率が全国20政令市でトップとなり、…(略)…4年連続で1位となった」という記事はどうだろうか。令和3年度中3国語で、仙台市平均は9.5、標準偏差2.6であり、全国(公立)の平均は9.0、標準偏差2.8である。平均差の効果量は0.18となり、これまた僅差という結果である。令和3年度中3数学では、仙台市の平均9.6、標準偏差3.7、全国(公立)の平均9.1、標準偏差3.7であり、平均差の効果量は0.14である。さほど大きくない効果量0.50(基準値)にさえとても届かない。47とか67個のどんぐりを並べて背を測り、「あなたは1位、あなたは最下位」と言っているような報道が適切だとはとても思えない。

最後に全国都道府県ランキングの妥当性についても検討しておく。トップ都道府県平均と全国平均との差、またボトム都道府県平均と全国平均との差には、効果量にしてどれほどの大きさがあるのだろうか。令和4年度、令和3年度、平成31年度、平成30年度、平成29年度、平成28年度、さらに全国学力・学習状況調査がスタートした平成19年度、その後の平成20年度、平成21年度について、小6の国語と算数、中3の国語と数学のそれぞれの平均差の効果量を計算してみた。平成31年度以前は、A問題とB問題に分かれていたのので、A問題とB問題を別々に計算した。その結果の一覧が、本稿末の資料である。

全120回の計算結果のうち、 $d \geq 0.50$ だったのはわずか5回、全体の4%しかなかった。全国平均と各都道府県平均との差もまた、全国学力・学習状況調査スタート時から僅差であったと結論づけることができる。

以上の検討から、都道府県や政令市について、「1位」とか「最下位」とかとラベルを貼ったり、順位づけたりすることは、小さな差を序列化し競争を煽る不適切な報道であると結論づけたい。

11. 教育データサイエンスはどこまで科学的で現象学的か？

ここで立ち止まって、私たちの教育データサイエンスは根本的に何をしているのか、それはどこまで科学的であり、どこまで現象学的であるかを自己理解してみたい。

11. 1. 教育データサイエンスの科学性

私たちの教育データサイエンスがやっている根本は、フッサールの言葉を借りるなら、「自然の数学化」ならぬ人間集団の数学化であり、人間集団ないしは児童生徒集団を「理念化」し、「数学的多様」にしている¹⁰⁹。ガリレイが相手取ったのは、私たちの感性的直観に与えられる自然ではなく、「それ自体において数

¹⁰⁹ フッサール, 1974, p.38.

学的な自然、式として与えられる自然、式からはじめて解釈される自然¹¹⁰であった。同様に、教育データサイエンスが相手取るのは、眼前の生身の児童生徒ではなく、大人が作成した質問項目によって測定された児童生徒の学力や生活習慣やパーソナリティや学習状況の数量であり、その数量を待ち受けている統計的数式であり、計算結果を可視化する幾何学である。つまり、教育データサイエンスが相手取っているのは「数学的な人間集団、式として与えられる人間集団、式からはじめて解釈される人間集団」である。ガリレイとは自然か人間かの違いしかない。

フッサールが鋭く洞察したように、ガリレイの自然科学の根幹には「測定術」がある。自由落下運動の実験でガリレイが経過時間と落下距離を測定したように、教育データサイエンスもまた、児童生徒の学力や生活習慣やパーソナリティや学習状況などの測定から出発する。この測定は広い意味で「テスト」と呼ばれる。自然科学や測定術に関するフッサールの洞察は、私たちが試みている科学がどのような性格を持つかの自己理解を促してくれる。

フッサールによると、測定術は「測定の精密さを高め、それを完全化を高める方向へますます押し進める技術」¹¹¹である。この指摘の通り、テストも、古典的テスト理論（CTT）から項目反応理論（IRT）へと精度を高め、近年ではペーパーベース（PBT）からコンピューターベース（CBT）のテストに進化しようとしている。完全化を高める技術のこうした運動には終わりが無い。さらに測定術は、「できあがった方法としての技術なのではなく、それは同時に、つねに新しい技術的手段（たとえば道具）を発明することによって、その方法をたえず新たに改良する方法でもある」¹¹²。統計ソフトエクセルやRは比較的新しく開発された道具であるが、そのおかげで平均値や標準偏差や信頼性係数などが素人にも簡単に一瞬で計算できるようになり、多種多様なグラフも一瞬で作図できるようになった。既存の道具を利用して、私たちのチームDS-EFAでも「平均ゾーンシステム」とか「平均値差検定システム」などの新しい道具を発明・開発できた。この技術改良の運動もまた無限の歩みであり、ますます多くの人を巻き込んでいく拡張の運動でもある。完全化や改良や巻き込みといった科学技術の無限の運動を、誰も止めることはできないし、誰もコントロールできない。科学技術の無限の運動は、人間の手中にあるのではなく、逆に人間の方が、科学技術の無限の運動の手中にある。科学技術の科学技術による科学技術のための無限の運動と言いたいくらいである。この運動は、ハイデガーのいう「総かり立て体制（Gestell：ゲシュテル）」かもしれない。データサイエンスがまだ一部の研究者にしか浸透していない教育の領域に、私たちがそれをもちこむことは、まさにこの科学技術の無限運動を教育の領域でさらに拡張・加速させることになる。

私たちの教育データサイエンスも統計学の指導のもとにあったため、様々な仮説を扱い、その仮説を検証してきた。仮説と検証は、まさに自然科学の理念に他ならない。この理念について、フッサールは瞠目すべきことを述べている。「ガリレイの理念は一つの仮説であり、しかもそれはきわめて注目すべき性質の仮説なのである」¹¹³。「注目すべき性質」とは、「仮説はそれが検証されるにもかかわらずいぜんとして永遠に仮説でありつづける」¹¹⁴という性質である。

無限に仮説でありつづけ、無限に検証でありつづけるということが、自然科学に固有な本質であり、アприオリに自然科学のあり方なのである。¹¹⁵

¹¹⁰ フッサール, 1974, p.73.

¹¹¹ フッサール, 1974, p.60.

¹¹² フッサール, 1974, p.60. 括弧内邦訳書。

¹¹³ フッサール, 1974, p.61.

¹¹⁴ フッサール, 1974, p.61.

¹¹⁵ フッサール, 1974, p.61.

ご多分に洩れず、私たちの教育データサイエンスも仮説を検証してきたが、絶対的な解にいたることはなかった。これは私たちのデータサイエンスが未熟だったからではなく、それが自然科学と同じ本質を共有していたからである。例えば、「正答率で5ポイント差は僅差でしかない」というのはあくまで仮説であり、それを支持する証拠をいくつか挙げたにせよ、仮説であり続け、検証され続ける性質のものである。逆にこれを反証する証拠が出たとしても、その反証が科学である限り、対立仮説もまた仮説であり続け、反証もまた反証であり続ける。「白黒決着をつけたい」というのは見込みのない夢であったことが、フッサールの洞察から判明する。

加えて、自然にせよ人間にせよ、それを数学化し、式によって解釈可能にすることによって、「粗雑な予見」を超えた「学的予見」が可能になる¹¹⁶。

一度式を手に入れば、それによって、具体的、現実的な生活の直観の世界…(略)…において、経験的確かさをもって期待されうるものを、実践的に望ましい仕方、あらかじめ予見できるようになる。¹¹⁷

放物運動の公式がわかれば、一定の初速と角度で発射した大砲の球がどれだけ先に落下するかが予測できるようになる。実際には空気抵抗や風などの影響を受けるため、現実には計算通りでないとしても、この予見が現実と与えるインパクトは大きい。同様に、効果量の計算式がわかれば、教育的働きかけの効果を計算でき、それぞれの働きかけの効果を予見できるようになる。確かに経験知からも、働きかけの効果を素朴に予見することはできるが、それは効果量という学術的予見のような精度も客観性もない。学術的予見は経験的予見を裏付けたり修正したりすることができる。

ただし、テストによって数学化された人間の世界と現実世界とのギャップは、物理学的自然と現実の自然とのギャップに比べて格段に大きい。完成度を増しつづける統計学によって、個人や集団の将来がこの先いっそう高い精度で予測できるとしても、それはあくまで数学化された世界での予測であり、生身の個人や集団の実存的な将来がその通りになるとは限らない。この点にも十分注意しよう。

11. 2. 教育データサイエンスの現象学性

フッサール現象学のおかげで、自然科学やデータサイエンスの内部からでは曖昧だった次元が区別できるようになる。それは、感性的に直観される経験的世界と、数学化された理念的世界との区別である。フッサールはこの違いに様々な表現を与えている。ガリレイは「経験的なものを数学的な極限理念へ関係させようとする」¹¹⁸というのもその一つの表現である。あるいは、前者は「内容的充実の世界」、後者は「形態の世界」とも呼ばれる¹¹⁹。また、「われわれが学問以前の生活において経験する色や音、熱、物体そのものの重さ、因果的に考えてまわりの物体を熱くする熱線などは、物理学的には、音波や熱波の振動などといった、純粋に形態の世界の出来事を告知している」¹²⁰とも言われる。この区別が現象学によってはじめて明らかになるとすれば、私たちの教育データサイエンスは、この区別を知る点で現象学的である。

また、フッサールが示唆するようにガリレイや自然科学者は理念の世界こそを「真に存在するもの」とみなし、主観にとって相対的な経験的世界(生活世界)をそうでないとみなしたとすれば、私たちのデータサ

¹¹⁶ フッサール, 1974, p.73.

¹¹⁷ フッサール, 1974, p.63.

¹¹⁸ フッサール, 1974, p.44.

¹¹⁹ フッサール, 1974, p.55.

¹²⁰ フッサール, 1974, pp.54-55.

イエンスはそうした自然科学者の立場には立たない点でも現象学的である。私たちは、理念的世界も生活世界も、いずれも真に存在するとみなす。

私たちはさらに、ガリレイと同時に始まったとされる「理念化された自然を学以前の直観的自然にすりかえること」¹²¹に注意深くあることで、フッサールが批判する自然科学とは一線を画す点でも、現象学的である。フッサールの次の洞察もまた重要である。

すでにガリレイのもとで、数学的な基底を与えられた理念性の世界が、われわれの日常的な生活世界に、すなわちそれだけがただ一つ現実的な世界であり、現実の知覚によって与えられ、そのつど経験され、また経験されうる世界であるところの生活世界に、すりかえられているということは、きわめて重要なこととして注意されねばならない。このすりかえは、ただちにその後継者たち、つまり引きつづく数世紀間の物理学者たちによって相続されることになった。¹²²

「すりかえる」にあたる原語のドイツ語は、unterschiebenであり¹²³、「あるものを別のものの下にすべり込ませる（押し入れる）」「あるものに別のものをこっそり押しつける」といった意味がある。このドイツ語は、「下」を意味する前綴りunter-と、「押す」「押しつける」を意味する語幹schiebenから出来ている。現代人の発達プロセスからすれば、私たちが幼少期に最初に出会うのは五感を通して経験される自然であり、自然科学の理念的世界は主に学校教育を通して後から学ばれる自然である。しかし、いつの間にか理念的な自然が経験的自然の下にすべり込まれ、理念的自然の方が本当の自然であるかのような錯覚さえ生み出してしまふことがある。「私が立っているこの大地は平坦であるかに見えるが、実は丸い」という理解も同じである。人間集団の理解でも同じで、「この学校の小学6年生の学力は、全国平均より5ポイント低い」と評価するときも、この「すりかえ」ないし「すべり込ませ」が起きている。現象学に学ぶことで、私たちはこの「すべり込ませ」に対する感受性をもつことができるようになる。

さらにこの洞察から、私たちの生活世界の理解には、理念的世界の理解がすべり込み、両者が混交・癒合していることが判明する。生活世界は、知覚によってのみ経験される純粋経験の世界ではなく、科学技術的な理念的世界が部分的にその下にすべり込んでいるハイブリッドな世界である。

もう一つ、フッサールの重要な発見として「意味の空洞化」という危機的現象がある。

この幾何学の算術化ということは、当然であるかのように、ある仕方では幾何学の意味の空洞化へゆきつく。「純粋直観」という一般的な名称のもとに幾何学的思考に本原的に呈示されていたような、実際に空間時間的な理念性が、いわば純粋数形態、代数的形象に代わってしまう。代数的計算においては、幾何学的意味はおのずから斥けられる。いな、まったく脱落させられてしまう。人は計算するだけで、せいぜい最後になって、数は大きさを意味するはずであったということを思い出すくらいのものである。¹²⁴

難解な文章だが、要するに、幾何学が算術化され、作図などによって幾何学的に直観されていたことが単なる計算になってしまうことで、そもそもそれが何を意味するのかが見失われ、意味が空洞化される現象と

¹²¹ フッサール, 1974, p.71.

¹²² フッサール, 1974, p.69.

¹²³ Vgl., Husserl, E., 1967, *Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie*, Martinus Nijhoff., S.48.

¹²⁴ フッサール, 1974, p.64.

理解しておきたい。次の文章も同じく意味の空洞化の解説である。

人は文字や結合記号や関係記号（+、×、=などのような）を、その結合のゲーム規則に従って、すなわち事実上本質的にはカードや将棋といったゲームと同じような仕方でも操作するのである。その技術的操作に真の意味を与え、かつ規則に合った結果に真理性……を与える根源的思考は、ここでは排除されている。¹²⁵

計算規則に従って計算するだけの「単なる技術」¹²⁶に教育データサイエンスが陥ってしまう危険に注意しよう。その計算や計算結果が何を意味するのかを考える「根源的思考」に努めよう。思えば、「平均正答率2～5ポイント差は大きいのか」という私たちのあまりにも初歩的な疑問も、その差がどの程度であるかをグラフによって直観化したのも、根源的思考へと立ち返ろうとする試みだったかもしれない。初心には意味の空洞化への対抗も含まれているかもしれない。意味の空洞化への抵抗が現象学の一要素であるならば、初心から始める教育データサイエンスはこの点でも現象学的である。

以上のように、私たちの現象学的教育データサイエンスは、現象学と教育データサイエンスとを往還しつつ、両者の地平を融合し、新しい地平を拓いてゆく営みのようである。現象学は、私たちが何をしているのか、つまり私たちがしている教育データサイエンスとは何かの自己理解を豊富にしてくれる。逆に教育データサイエンスは、知覚とかコミュニケーション的行為によって直接的に交流する子どもを超えた、大規模な子ども集団の世界を開いてくれる。たとえそれが血肉を備えた人間集団ではなく、数学化された人間集団であるとしても。自然の場合と同様に、血肉を備えた個人や集団と、数学化された個人や集団との関係や交渉もまた解明すべき課題である。少なくとも現象学は、これら二つの世界を区別することを教えてくれるし、それぞれの世界の仕組みを教えてくれる。さらに現象学は、数学化された個人と集団にも異なるタイプがあり、生きられる数値による数学化と、意味が空洞化した数値による数学化とがあることを教えてくれる。

本稿の最後に、私たちの教育データサイエンスの意味連関をさらに拡張する3つのデータ分析（12、13、14章）を紹介したい。

12. 家庭の蔵書数の質問項目は児童生徒のSESの代替指標となるか？

12. 1. 問題背景

令和3、4年度には、児童生徒の家庭のSES（社会経済文化的状況）を把握するために、家庭の蔵書数の質問項目が児童生徒質問紙に盛り込まれた。すでに言及したように、SESは学力と弱く相関する。相関係数としては大きくないが、決して無視できない影響である。教師たちも、SESが学力に与える影響を、経験上大なり小なり感知している（ただし多くの教師の経験知は、測定された相関係数からすれば、その影響を若干過剰評価している傾向にあると筆者は感じている）。それだけに、この質問項目がどれほどの精度で家庭のSESを反映しているかは、検証しておきたい問題である。私たちのチームDS-EFA内でも、「この質問だけで、本当にSESがわかるのか？」とか、「中3ならともかく小6の子どもの自宅の蔵書数をどれだけ正確に回答できるか？」とか、「児童生徒の質問項目には『雑誌、新聞、教科書は除きます』としかなく、子どもはおそらく漫画の冊数をカウントしているはずだから、SESの指標としては信ぴょう性がない」といった疑念や異論が絶えなかった。SESの代替指標としてこの質問項目の精度は信頼に足るだろうか。

そこでこの質問項目の精度を検証するために、次のような研究デザインを考案してみた。令和3年度は保

¹²⁵ フッサール, 1974, p.66. 括弧内邦訳書。

¹²⁶ フッサール, 1974, p.66.

護者に対する調査も実施しており、保護者にも家庭の蔵書数を質問している。児童生徒の回答状況とその保護者の回答状況とがどの程度整合的かを検証し、相関関係が強ければ、児童生徒への質問項目の精度は一定程度信頼できると評価できるだろう。また、保護者アンケート調査では、世帯収入（税込年収）も質問している。蔵書数は世帯収入ともある程度の相関があるはずである。以上の仮説からすれば、①児童生徒による蔵書数の回答、②保護者による蔵書数の回答、③保護者による世帯収入の回答、これら3者にはそれなりの相関関係が想定される。それゆえ、これら3者を合成変数（合成尺度）にすることもでき、その合成変数の信頼性係数 α は、調査研究に必要なとされる0.70前後はあると仮定できる。まずはこの仮説を検証してみよう。

12. 2. 検証のためのデータセットと手続きと検証結果

文部科学省の貸与制度を利用し、私たちは令和3年度の保護者に対する調査の匿名データを借り受けた。そのデータセットでは、保護者のIDごとに、その児童生徒の学力値、児童生徒質問紙回答スコア、当該保護者の質問紙回答スコアが接続されている。

① 令和3年度の児童生徒質問紙調査では¹²⁷、小学校版も中学校版も、項目（22）が蔵書数の質問になっている。データ処理では「STQ_022」と略記する。質問文は、小学校版は次のように記されている。

あなたの家には、およそどれくらいの本がありますか。（雑誌、新聞、教科書は除きます。）

中学校版もほぼ同じであるが、「雑誌」の箇所が「一般の雑誌」という記載になっている。

児童生徒にイメージしやすいよう、図11のようなイラストが挿入されている。

回答選択肢は、「1 0～10冊」「2 11冊～25冊」「3 26冊～100冊」「4 101冊～200冊」「5 201冊～500冊」「6 501冊以上」の6件法である。

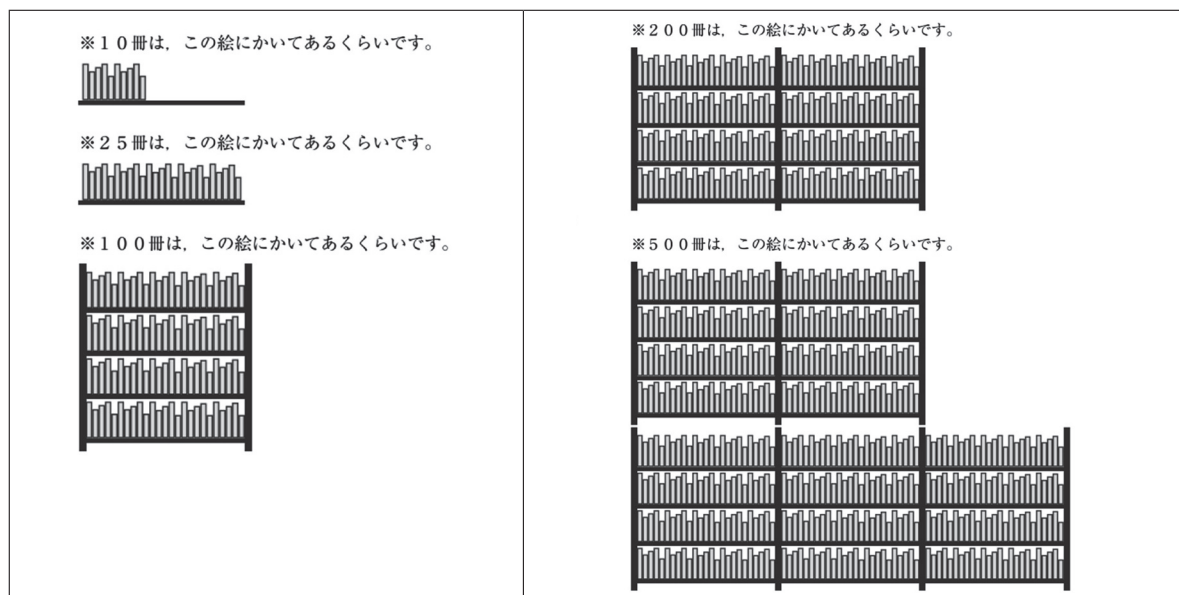


図11：質問項目（22）に添えられた蔵書のイラスト

¹²⁷ 令和3年度の児童生徒質問紙は、国立教育政策所の「教育課程研究センター『全国学力・学習状況調査』」のウェブページから、令和3年度の「調査問題・正答例・解説資料」をクリックすると、「○質問紙調査」のコーナーに、小学校と中学校の児童生徒質問紙のPDFが用意されている。

② 令和3年度の保護者アンケート調査では¹²⁸、小学校版も中学校版も、家庭の蔵書数の質問項目は「6-(3)」である。データ処理ではこの項目を「PTQ_06_03」と略記する。質問文は次のとおりである。

あなたの家には、およそどのくらいの本がありますか。本棚1メートルにつき、約40冊の本が入るとします。(電子書籍や雑誌、新聞、子供向けの本は含みません。)

児童生徒に対する質問文と異なるのは、含まない本として「電子書籍」と「子供向けの本」が追加されている点である。除外対象が異なるため、児童生徒と保護者の回答は完全には一致しないと予想される。

選択肢は、児童生徒調査と同じく、「1. 0~10冊」「2. 11冊~25冊」「3. 26冊~100冊」「4. 101冊~200冊」「5. 201冊~500冊」「6. 501冊より多い」の6件法である。

③ 家族の世帯収入は、保護者アンケート調査の質問項目「7-(9)」である。データ処理では、「PTQ_07_09」と略記する。質問文は、次のとおりである。

あなたのご家族全体の世帯収入(税込み年収)は次のどれにあてはまりますか。

回答選択肢は13件法で「1. 100万円未満」「2. 100万円以上200万円未満」「3. 200万円以上300万円未満」「4. 300万円以上400万円未満」「5. 400万円以上500万円未満」「6. 500万円以上600万円未満」「7. 600万円以上700万円未満」「8. 700万円以上800万円未満」「9. 800万円以上900万円未満」「10. 900万円以上1,000万円未満」「11. 1,000万円以上1,200万円未満」「12. 1,200万円以上1,500万円未満」「13. 1,500万円以上」である。

学力値とSESの相関も見たいため、データセットには、国語(Jpn)と算数・数学(Math)の正答数も加えた。データ処理として、まず児童生徒回答から無回答「0」とその他「99」をエクセルで空欄に置換し、同様に保護者回答から非該当「-99」無回答「-9」誤記入「-8」を空欄化、もとの欠測値である空欄と合わせ、SPSSで空欄のあるIDをデータセットから削除した。

その結果、分析対象の総人数は、小6で28,924名、中3で63,920名となった。

STQ_022とPTQ_06_03とPTQ_07_09を合成した場合の信頼性係数(クロンバックの α 係数)をSPSSで計算したところ、小6で0.49、中3で0.44となった。この結果は、正直予想外であった。

なお「信頼性係数」とは、心理測定尺度の「一貫性(内的整合性)」を検証する値であり、「ある心理測定尺度を構成する複数の項目に共通した反応(回答)が得られるか」を評価するための指標である¹²⁹。複数の項目が同じ方向を向いているかを判定するための指標とも言える。一般に、調査研究では0.70くらいは必要とされる。繰り返すが、上記の3項目の信頼性係数が0.70にとっても及ばないことは、筆者には大きな驚きであった。

世帯収入の質問項目PTQ_07_09を除き、蔵書数の質問項目STQ_022とPTQ_06_03に限定した場合、この2項目の合成変数の信頼性係数は小6で0.60、中3で0.59に上昇した。0.70には届かないが、これら2項目はそれなりの内的整合性があると評価できる。自宅の蔵書数については、児童生徒の回答と、保護者の回答は、

¹²⁸ 令和3年度の保護者アンケート調査は、国立教育政策所の「教育課程研究センター『全国学力・学習状況調査』」のウェブページから、令和3年度にある「保護者に対する調査」をクリックし、「小学校調査票」「中学校調査票」をクリックすることでPDFが閲覧・ダウンロードできる。

¹²⁹ 宮本聡介・宇井美代子編, 2015『質問紙調査と心理測定尺度—計画から実施・解析まで—』サイエンス社, p.67.

それなりに整合的と評価できそうである。

対して、世帯収入と蔵書数とは相互に比較的独立性の高い変数であることも判明する。この結果は、収入が高いからといって本を買うとは限らず、収入が低くても本を買う家庭も少なくないことを意味している。図12、13にあるように、保護者回答の世帯収入と蔵相数の相関係数は小6で0.31、中3で0.26と驚くほど弱い相関しかない。蔵書数は、保護者の「経済的状况」よりも「文化的状况」を色濃く映し出していると解釈できる。そこで、以後は、この指標を、SESにCulturalを加味したSECS (Socio-Economic Cultural Status : 社会経済文化的状況) の代替とみなすことにしたい。

Rのpairs.panels関数を使って、各項目の相関係数を計算してみると、図12、13になる。

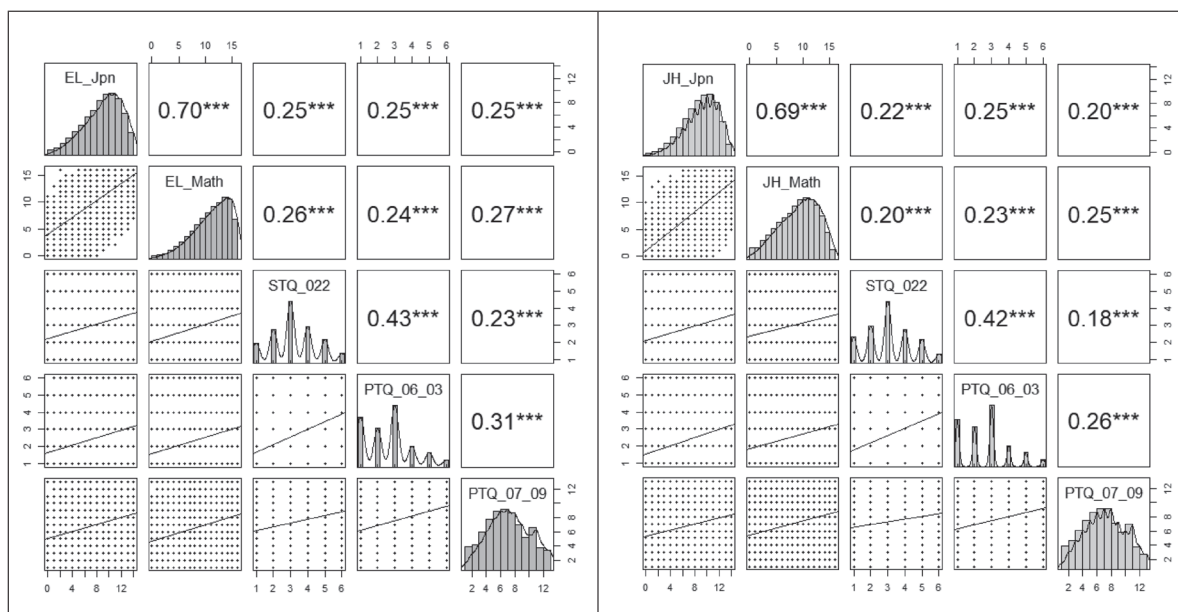


図12：R3小6の相関係数一覧

図13：R3中3の相関係数一覧

蔵書数の児童生徒回答 (STQ_022) と保護者回答 (PTQ_06_03) の相関係数は、小6で0.43、中3で0.42とほぼ同じであり、先に紹介した一般的な基準値で、中程度の相関の強さである。蔵書数についての児童生徒回答と保護者回答は、ある程度整合的であると言える。それゆえ、蔵書数に関する児童生徒への質問項目の精度は、悪くはないと評価できる。

さらに、学力と児童生徒回答の蔵書数スコアとの相関係数は、学力と保護者回答の蔵書数スコアとの相関係数とほぼ同じになっている。SECSの代替指標として蔵書数を知りたい場合、学力値への影響を知るには、児童生徒の回答でも、保護者の回答でもさほど違いはない、という結果である。

これらから、蔵書数に関する児童生徒質問項目は、SESなりSECSの代替指標としてそれなりの信ぴょう性があると評価しておく。注意深く言えば、この質問項目がSESなりSECSの代替指標であるという見方はあくまで仮説であるが、今回の検証でこの仮説を棄却する証拠は出なかった。

今回の検証の副産物として、次の3つの仮説も浮上する。

1つ目は、国語と算数・数学との相関は、SECSに比べ、圧倒的に強いという仮説である。この相関の強さから、これら以外の教科との相関の強さも予見される。

2つ目は、蔵書数にせよ、世帯収入にせよ、SECSが学力に与える影響は、相関係数にして0.20から0.30程度にとどまるという仮説である。先に紹介した文部科学省公表の学力値と質問項目との相関係数を見ると、この程度の相関を示す質問項目は他にいくつもある。例えば、令和3年度小6調査の相関係数では、「普段

(月曜日から金曜日)、1日当たりどれくらいの時間、テレビゲーム(コンピュータゲーム、携帯式のゲーム、携帯電話やスマートフォンを使ったゲームも含む)をしますか」の項目と国語スコアとの相関係数は0.27、算数スコアとは-0.26である。ゲームの時間が短いほど学力が高いという弱い相関がある。あるいは「5年生までに受けた授業では、課題の解決に向けて、自分で考え、自分から取り組んでいましたか」の項目は、国語とは0.25、算数とは0.25の相関がある。あるいはまた「国語の授業の内容はよくわかりますか」の項目は、国語と0.31、算数とも0.26の相関がある。このように、学校には、家庭のSECSを挽回する要素がいくつもあることが見えてくる。これは、教育関係者や児童生徒や保護者に希望を与える知見ではないだろうか。同時に教育関係者の責任の重さを示唆する結果でもある。

3つ目は、小6よりも中3の方がSECSの影響が若干弱まっているという仮説である。図12、13で、同じ項目同士の相関係数を比較すると、中3の方がわずかに相関係数が小さいことがわかる。学年が上がるにつれ、家庭の影響よりも、本人や仲間や学校の影響が強まっているという仮説が浮上する。

12. 3. SESの合成変数の試行に関する現象学的省察

以上から、科学の仮説がどのようなものであるかの理解も進む。フッサールが指摘したように確かに仮説は仮説であり続ける。しかし、仮説は修正されるときに否定され、別の仮説にとってかわられることもある。本章では、3つの項目の合成変数の信頼性係数を、筆者ははじめ0.70程度と予測していたが、検証の結果0.50にも届かず、3項目は合成変数には相応しくないと修正を迫られた。この一件からもわかるように、仮説が絶対的に証明されることはなく、仮説のままであり続ける一方で、仮説が検証によって修正あるいは否定されることは確かである。まさにカール・ポパーが指摘するように、「科学者は、理論が偽であることの確立には合理的な確信がもてるばあいがあるのに反して、自分の見出したことが真であることについては、決して確実に知ることができない」¹³⁰。真であることは証明できないが、偽であることは確実に知ることができる。これが真正な科学の仮説や推測である。フッサールと同じくポパーも、「科学理論は正真正銘の推測〔つまりは仮説〕である」¹³¹とか「法則や理論はすべて推測、あるいは暫定的な仮説」¹³²であるという。反証や反駁が可能なのは科学の弱みではなく、強みなのだ。ある理論が科学か非科学かを定める「判定基準は、その反証可能性、反駁可能性、ないしテスト可能性である」¹³³。逆に、あらゆる事例がある理論によって解釈できる場合、その理論は「非科学」である¹³⁴。対照的に、科学の予測には「リスク(反駁の危険性)」¹³⁵がつきものである。「万一、観測の結果、予測された効果がまったく発見できなくなれば、その理論は直ちに反駁されてしまう」¹³⁶というリスクである。私たちの教育データサイエンスも、予測は立つが、実際にデータを計算してみないとわからないというリスクの連続である。これは、私たちの教育データサイエンスが、科学の判定基準をクリアしていることを示しているだろう。私たちの教育データサイエンスは科学の領域に踏み込んだと言えそうである。

ひるがえって、私たちがもう一本の導きの糸としている現象学は、科学なのだろうか、それともあらゆる事例を解釈できる非科学なのか、それともこの二分法を超える何かなのだろうか。

¹³⁰ ポパー, K. R., 1980『推測と反駁』藤本隆志・石垣壽郎・森博訳, 法政大学出版会, p.183.

¹³¹ ポパー, 1980, p.183.〔〕内は引用者。

¹³² ポパー, 1980, p.93.

¹³³ ポパー, 1980, p.64.

¹³⁴ こうした非科学的理論の例として、ポパーは、マルクスの歴史理論、フロイトの精神分析理論、アドラーの個人心理学をあげ、科学であるアイシュタインの相対性理論と対比している (cf., ポパー, 1980, pp.58-62)。

¹³⁵ ポパー, 1980, p.62. 括弧内邦訳者。

¹³⁶ ポパー, 1980, p.62.

ともあれ、反証可能性という観点から私たちの教育データサイエンスを見直してみると、このデータサイエンスは2つのタイプの予測と反証可能性に曝されている。一つは、先の合成変数の信頼性係数のように数値データの世界での予測と反駁である。もう一つは、現実の生身の世界での予測と反駁である。例えば、次の13章で検証するように、全国データから対話・探究学習が学力と非認知能力を押し上げると予測される場合、ある学校がその学習に力を入れた結果、その効果が現れるか否かが試される。その効果は、教師たちの実践感覚とテストとの2つのタイプの観察によって把握される。私たちの教育データサイエンスに特徴的なのは、予測と反駁が、数学化された人間の世界だけで完結せず、生身の人間の世界にも開かれていることである。それゆえ、私たちの教育データサイエンスは実践的性格をもつし、教育実践現場に対し重い責任を負っている。

さらに加えて、先の予測と検証から、フッサーが指摘しなかった現象、すなわち仮説と検証は新たな仮説を次々と生み出し、仮説のネットワークを次々と広げていくことも判明する。このネットワークがデータサイエンスの地平なのかもしれない。実は仮説は問いでもある。仮説が検証を求めるのは、それが問うからである。データサイエンスの地平は、問いと検証の無限の連関かもしれない。また私たちの出発点にある問いはささやかながらユニークだったかもしれない。「SESの格差はどれほど教育格差をもたらすか？」は有名な問いだが、私たちの問いは「全国平均正答率との2~5ポイント差は大問題か小問題か？」である。この問いを共有する現場関係者は少なくない。マスコミ関係者は答えを決め、「大問題」という前提で報道している。結局この問いは仮説であった。「差はない」という帰無仮説の検証の問題に等しい。この問いを出発点として検証をすれば、連鎖反応としてまた別の問いが現れ、独自の無限の問いと検証の連鎖となっている。例えば、この問いをきっかけに、効果量に行き着いたが、教育の効果量の基準値はどの程度かの問いが浮上し、それを検証する一環としてSESと学力との効果量が参照され、蔵書数の質問はどれほどSESを反映しているかの問いになった。こうした連鎖反応は、現象学的に言えば意味連関とか存在秩序ということになるが、これは実はトーマス・クーンのいう「パラダイム」ではないだろうか。ポパーの論敵だったクーンによれば、「パラダイム」とは、「一般に認められた科学的業績で、一時期の間、専門家に対して問い方や答え方のモデルを与えるもの」¹³⁷である。私たちの教育データサイエンスは、今のところ「一般に認められた科学的業績」にはほど遠いが、このデータサイエンスがそれ自身に固有の問いを次々と投げかけてくることは確かである。このデータサイエンスは、パラダイムなるものを形成しはじめているのかもしれない。

12. 4. 教育データサイエンスにおける意味の空洞化の兆し

データ分析の話に戻れば、先の分析(12. 2)は個人スコアをもとにしているが、グループスコアをもとにすることもできる。SECSの階層ごとにグループ化する方法である。小6の児童を、蔵書数の回答ごとにグループ化し、そのグループごとに保護者の蔵書数回答の平均 μ_1 と世帯収入の平均 μ_2 を計算し、それを一覧化すると表13になる。この一覧から、児童グループの階層が上がるにつれ、2つの平均値が徐々に上がっていることがわかる。この上昇を可視化するために、さらに μ_1 と μ_2 の散布図を描き、散布図に回帰直線を引き、その回帰直線の決定係数 r^2 値を計算してみると、図14になる。

¹³⁷ クーン, T., 1995 『科学革命の構造』 中山茂訳, みすず書房, p.v.

表13：児童回答蔵書数1～6ごとの保護者回答蔵書数平均と保護者回答世帯収入平均

児童回答蔵書数	保護者回答蔵書数	保護者回答世帯収
1	1.75	6.1
2	2.13	6.59
3	2.51	7.3
4	2.97	7.8
5	3.45	8.27
6	4.03	8.47

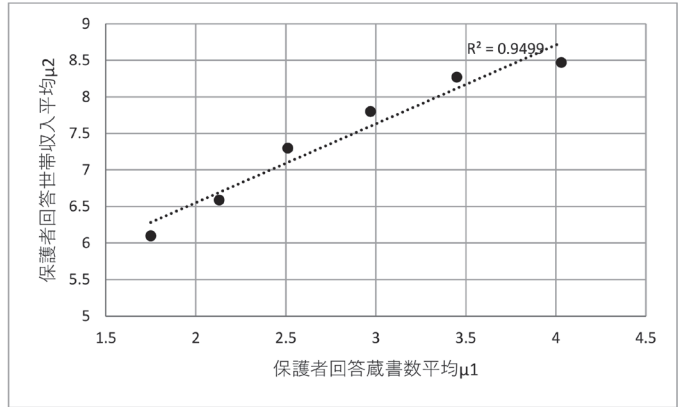


図14：児童回答蔵書数ごとの $\mu 1$ と $\mu 2$ の散布図と回帰直線および決定係数 (R^2)

決定係数で0.95となり、階層ごとの格差が歴然とする。

ただし、このグループスコアによる分析には、意味の空洞化の兆しがある。確かに、手続きや計算の一つひとつの意味は明快で理解できる。しかし、理解に苦しむのは、図14の6つの点が何を意味するかである。確かにその意味は、「1とか2とかと回答した児童グループの保護者の蔵書数平均と世帯収入平均を2次元の座標上に表現した点」である。しかし「2値の平均を座標上に表現する」というのは抽象的な操作であり、意味充実が難しい。

例えば、1と回答した児童 ($n=3,126$) を取り出し、個人スコアでその保護者が回答した蔵書数と世帯収入のスコアの散布図を描くと図15になる。また6と回答した児童 ($n=1,491$) の同じ散布図は図16である。赤のドットが各グループスコアで2値の平均を座標に表現したものである。

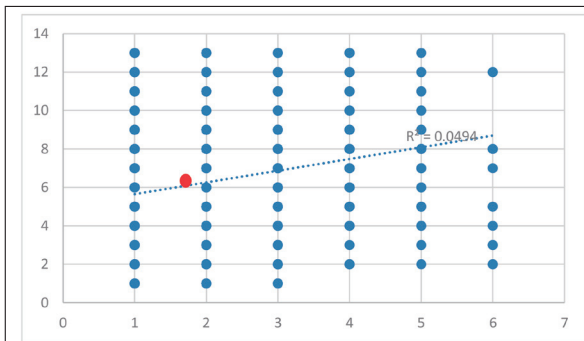


図15：1と回答した児童の保護者の蔵書数と世帯収入の散布図および決定係数 (R^2)

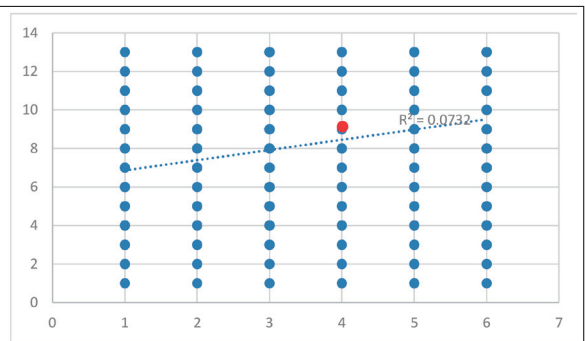


図16：6と回答した児童の保護者の蔵書数と世帯収入の散布図および決定係数 (R^2)

1と回答した児童生徒の保護者3,126名は、蔵書数について1と回答した者もいれば、2や3や4や5や6と回答した者もあるし、世帯収入についても同様に散らばっている。決定係数 r^2 はそれぞれ0.049と0.073で、その平方根の相関係数はそれぞれ0.22と0.27となり、弱い相関しかない。それゆえ、確かに平均すればその代表値が得られるが、これだけまんべんなく散らばっている青のドットを、赤のドットが本当に代表しているの

か理解に苦しむのである。ここに意味の空洞化の兆しがある。

確かに、この2つのグループの保護者の世帯収入回答のヒストグラムを描くと図17、18となり、1と回答した児童の保護者と6と回答した児童の保護者の、世帯収入に関する回答傾向の差を理解できる。

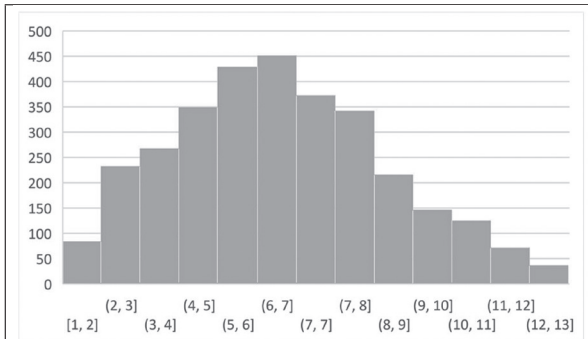


図17：1と回答した児童の保護者の世帯収入のヒストグラム

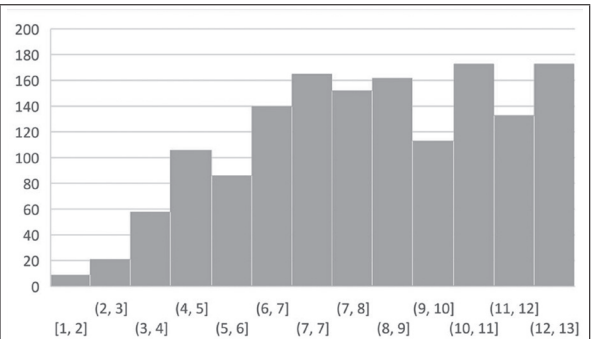


図18：6と回答した児童の保護者の世帯収入のヒストグラム

確かに、1と回答した児童の保護者に比べ、6と回答した児童の保護者の世帯収入の回答は、高い傾向にある。グループごとの階層性がないわけではない。しかし、上記の赤のドットは抽象的で意味充実が難しい。それゆえ、直観や意味充実を重視する現象学的な教育データサイエンスでは、抽象度が上がるグループスコアより、意味充実しやすい個人スコアを優先したい。

ちなみにグループスコアを利用したデータ分析でよく知られているのは、いわゆる「クロス分析」である。児童生徒質問紙調査でも、文部科学省はクロス分析を公表している。令和4年度児童質問紙調査のクロス分析の次の例（図19）¹³⁸は、表13や図14と類似のデータ処理である。

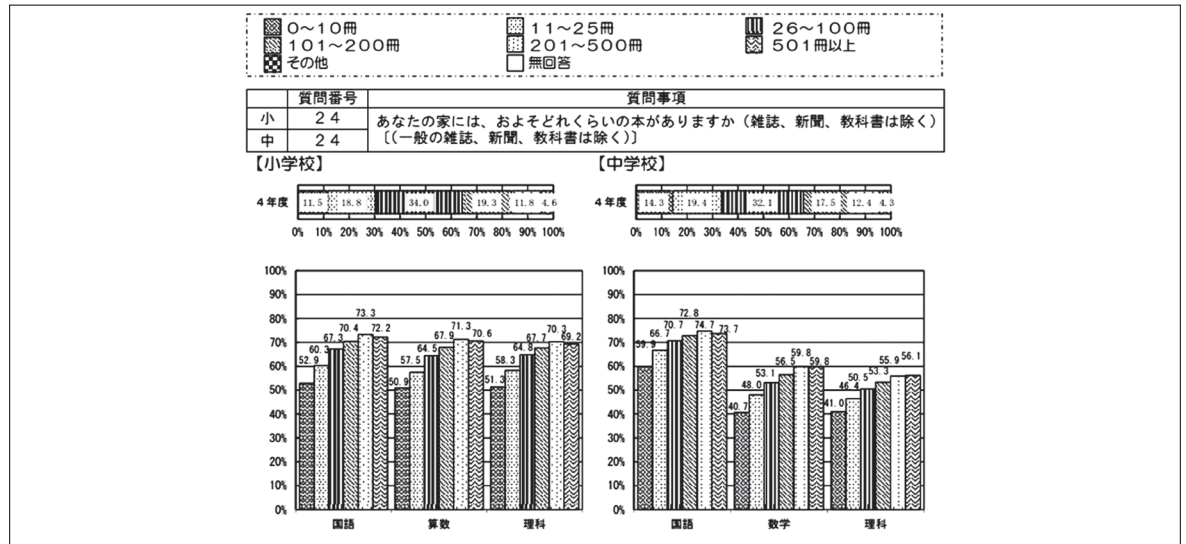


図19：令和4年度児童生徒質問紙、蔵書数のクロス分析（文部科学省公表）

¹³⁸ 文部科学省「令和4年度 学力・学習状況調査 報告書【質問紙調査】」の小学校版121ページより転載。
https://www.nier.go.jp/22chousakekkahoukoku/report/data/22qn_03.pdf

図19は蔵書数の回答ごとに児童生徒をグループ化し、6つのグループの教科の平均正答率を棒グラフで比較している。蔵書数が増加するにつれ、教科の平均正答率が上がることが印象的である。

ただし、こうしたグループスコアによる可視化では、個人スコアではさほどでない効果が大きく見えてしまう問題がある。例えば、学校質問紙「調査対象学年の児童生徒は、授業中の私語が少なく、落ち着いていると思いますか」の項目を、グループスコアで可視化すれば、図20のようになる。

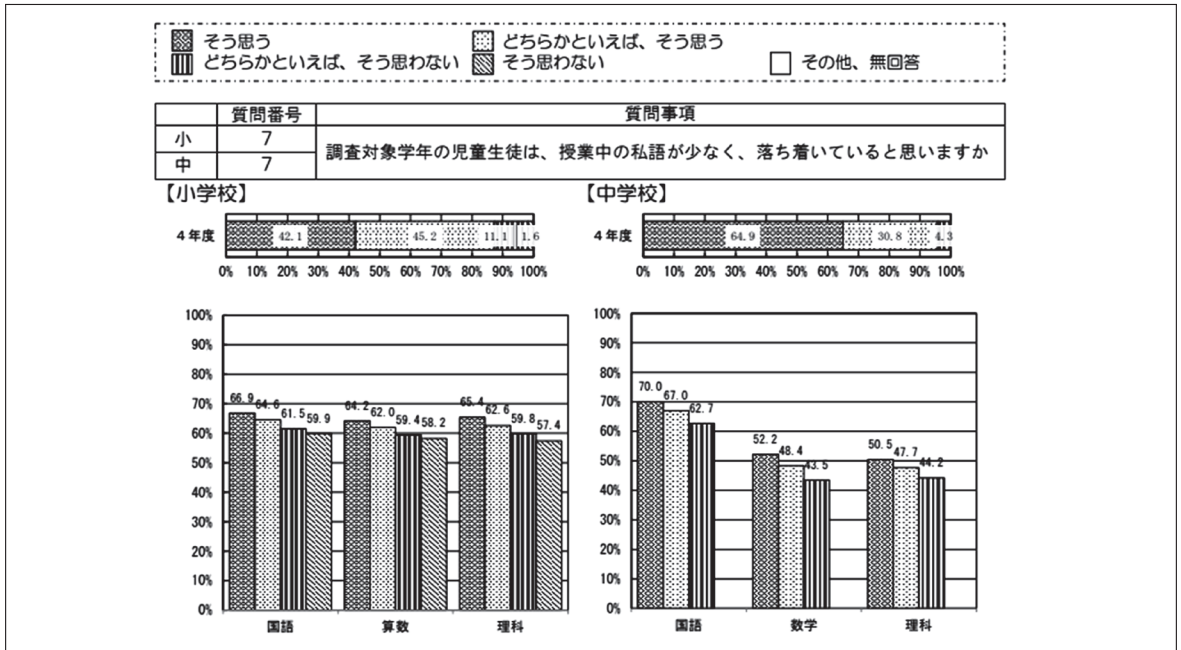


図20：令和4年度学校質問紙、読書に関するクロス分析（文部科学省公表）¹³⁹

「当てはまる」と回答した学校ほど学力スコアが高いこと、逆に「当てはまらない」と回答した学校ほど学力スコアが低いことがはっきり出ている。ところが、学校ごとの個別スコアの相関係数を見ると、この項目と学力との相関は、小6で国語0.063、算数0.059、理科0.067であり¹⁴⁰、中3で国語0.073、数学0.079、理科0.073と¹⁴¹、相関関係を認めることはできない。抽象度を高めることによる意味の空洞化の弊害である。学校質問紙に回答するのは多くの場合教頭であるが、教頭が当該学年を観察して「落ち着いている」と判断しても、その学年には学力の高い子から低い子までちらばっているため、学校単位の個別スコアで相関関係が認められないことは、当然の理である。この当然の理を意味充実することは容易である。ところが、グループスコアにすると図20のように、あたかも相関関係があるかのような結果になる。グループスコアの利活用には注

¹³⁹ 文部科学省「令和4年度 学力・学習状況調査 報告書【質問紙調査】」の小学校版139ページより転載。

¹⁴⁰ 国立教育政策研究所「令和4年度 全国学力・学習状況調査 調査結果資料【全国版／小学校】」の(3)相関係数、クロス集計表」中「相関係数(学校質問紙—教科)全国【表】」の質問項目(7)より。
<https://www.nier.go.jp/22chousakekkahoukoku/factsheet/primary.html>

¹⁴¹ 国立教育政策研究所「令和4年度 全国学力・学習状況調査 調査結果資料【全国版／中学校】」の(3)相関係数、クロス集計表」中「相関係数(学校質問紙—教科)全国【表】」の質問項目(7)より。
<https://www.nier.go.jp/22chousakekkahoukoku/factsheet/middle.html>

意が必要である。

こうしてみると、「個人スコア」と「グループスコア」との区別は非常に重要であることがわかる。「個人スコア」と「グループスコア」という概念は、実は東北大学大学院柴山直教授のご教示によるものである¹⁴²。柴山教授によると、これらの概念は統計学の入門書や専門書から得たものではなく、長年にわたり統計分析に携わるなかで柴山教授のなかで形成されてきたとのことである。柴山教授は、これらの概念の根拠資料として、Robert Brennan ed., 2006, *Educational Measurement*, Praeger Pubの第19章タイトル「Educational Measurement Chapter 19 Monitoring Educational Progress with Group-Score Assessments」で使われている「Group-Score」をあげている。

個人スコアでは現れない相関がグループスコアでは現れるという現象は、ウィリアム・ロビンソンが指摘した「生態学的誤謬 (ecological fallacy)」と同じである。ロビンソンは「個人相関」と「生態学的相関」という用語を用い、生態学的相関では0.946でも、個人相関では0.203しかない例を示している¹⁴³。

筆者としては、グループスコアの相関を「誤謬」と断じるつもりはないし、グループスコアの分析にはそれ相応の意義があると考えている。しかし、これらをしっかり区別し、個人スコアにはない相関関係がグループスコアでは出やすいことや、グループスコアでは意味の空洞化が起きやすいことに注意したい。

13. 非認知能力と対話・探究学習と授業充実度の可視化

全国学力・学習状況調査の匿名データを分析してみると、この調査が実に良くできていることがわかってくる。調査結果のデータは、教育関係者が知りたいことに貴重な示唆を与えてくれるからである。例えば、昨今注目される非認知能力 (Non-Cognitive Skills) の全国平均がどのくらいなのか、また非認知能力に影響を与えるのはどのような学習か、非認知能力は学力と相関するのかなどの問いを検証し仮説を得ることができる。

13. 1. 「非認知『徳』 (Virtus) 尺度」の開発

ノーベル経済学賞を受賞したジェームズ・ヘックマンの一連の調査研究で注目を浴び、2015年ごろからOECDも重要性を強調している「非認知能力」は、心理学者が開発した「the Pupil Behavior Inventory (PBI)」や「the Ypsilanti Rating Scale (YRS)」といった心理テストで測定される「パーソナリティ能力」である¹⁴⁴。近年では、「ビッグ・ファイブ (Big Five) 尺度」と呼ばれるパーソナリティ・テストが、非認知能力の測定では普及し、OECDも参照している¹⁴⁵。ビッグ・ファイブの構成要素は、「開放性 (Openness)」「誠実性 (Conscientiousness)」「外交性 (Extraversion)」「協調性 (Agreeableness)」「情緒安定性 (Neuroticism)」である。

昨今話題になる非認知能力は、ビッグ・ファイブのように複数の要素を合わせたパーソナリティ・スコアである。「自己有用感」とか「やり抜く力 (グリット)」とかの各要素が単体で注目されることもあるが、非

¹⁴² この場をお借りして、柴山直教授にお礼を申し上げたい。なお、あくまで筆者の見解であり、誤解などの責任は全て筆者にある。

¹⁴³ 森幸雄, 1987「生態学的データ利用における誤謬の問題—ロビンソンの生態学的誤謬問題を中心として—」『Sociologica』, 12巻1号, pp.24-26.

¹⁴⁴ Heckman, James J., Pinto, Rodrigo, and Savelyev, Peter., 2013, Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes, *American Economic Review*, 103(6), p.2067.

¹⁴⁵ Cf., 経済協力開発機構 (OECD) 編著, 2019『社会情動的スキル—学びに向かう力—』無藤隆・秋田喜代美監訳, 明石書店, p.53, p.57.

認知能力を測定する場合は、複数の要素をバランスよく合成しなければならない。その上で、各々の構成要素を見るのが、非認知能力の評価としては適切である。

また「パーソナリティ(人格)が測れるのか?」という疑問は、「学力は測れるのか?」と同じく、また「自然(物体)は測れるのか?」と同じく、おそらく解くことのできない謎である。学力よりも物体の方が測れると思うのは、測定を机や土地の大きさを測るような単純な技術的測定としか考えないからである。自然を数学化したガリレイはこんな測定には無縁であった。彼が測ろうとしたのは、それまでの生活世界とは無縁の自由落下運動などである。また、OECDが「社会情動的スキルの測定は困難ではあるが、信頼性をもって実施できる」¹⁴⁶と答えたとしても、この答えはあくまで仮説に過ぎない。自然科学が解明するのが数学化された自然であるように、教育データサイエンスが解明するのは数学化されたパーソナリティである。

それよりいっそう深淵な謎は、学力にせよパーソナリティにせよ自然にせよ、よくできた質問を投げかけると、個人や集団や自然は意味に満ちた数値を返してくれる、という事実である。「現実の経験〔ガリレイが行った実験〕は…(略)…彼〔ガリレイ〕の仮説的な予想がそのつど要求したとおりのことを示してくれた」¹⁴⁷ように、全国学力・学習状況調査の測定値は、教育データサイエンスの仮説的な予想の要求に、少なくとも今のところ、納得できる結果を返してくれるのである。

非認知能力を測定する質問項目群は、専門家が利用するもの¹⁴⁸から、ネット上で素人が簡易判定できるものまで多種多様である。全国学力・学習状況調査の児童生徒質問紙にも、ビッグ・ファイブとある程度対応する質問項目が盛り込まれている。令和4年度調査でみると、小6中3調査の質問番号7~18は、文部科学省の分類では「挑戦心、達成感、規範意識、自己有用感等」とまとめられている¹⁴⁹。私たちDS-EFAは、これらは非認知能力の代替指標として利用できると考えた。今回は平成31、令和3、令和4年度に共通する表14の8項目を合成することで、非認知能力の代替指標としてみる。

表14：非認知能力代替指標の質問項目¹⁵⁰

児童生徒質問紙項目一覧 (R4~H31)					斜線部は、該当質問項目なし								
番号 小 中	質問項目【R4】				番号 小 中	質問項目【R3】		番号 小 中	質問項目【H31(R1)】				
7	7	自分には、よいところがあると思いますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない				6	6	R4に同じ		5	5	R4に同じ	
9	9	将来の夢や目標を持っていますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない				7	7	R4に同じ		8	8	R4に同じ	
11	11	難しいことでも、失敗を恐れずに挑戦していますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない				9	9	R4に同じ		10	10	R4に同じ	
12	12	人が困っているときは、進んで助けていますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない				10	10	R4に同じ		14	14	R4に同じ	
13	13	いじめは、どんな理由があってもいけないことだと思いますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない				11	11	R4に同じ		15	15	R4に同じ	
15	15	人の役に立つ人間になりたいと思いますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない				12	12	R4に同じ		16	16	R4に同じ	
16	16	学校に行くのは楽しいと思いますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない				13	13	R4に同じ		12	12	R4に同じ	
30	30	地域や社会をよくするために何をすべきかを考えることがありますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない				25	25	R4に同じ		24	24	R4に同じ	

¹⁴⁶ 経済協力開発機構, 2019, p.56.

¹⁴⁷ フッサール, 1974, p.59. [] 内引用者。

¹⁴⁸ 専門的な日本語版としては、和田(1996)「Big Five尺度」がある (cf., 堀洋道監修・山本真理子編, 2014『心理測定尺度集 I—人間の内面を探る〈自己・個人内過程〉—』サイエンス社, pp.123-128)。

¹⁴⁹ 文部科学省「令和4年度 全国学力・学習状況調査 報告書—質問紙調査—」の「目次」の次のページ(ページ数なし)に分類表が記載されている。

<https://www.nier.go.jp/22chousakekkahoukoku/report/data/22qn.pdf>

¹⁵⁰ 「令和4年度 全国学力・学習状況調査 報告書」, pp.182-183より引用。

<https://www.nier.go.jp/22chousakekkahoukoku/report/question.html>

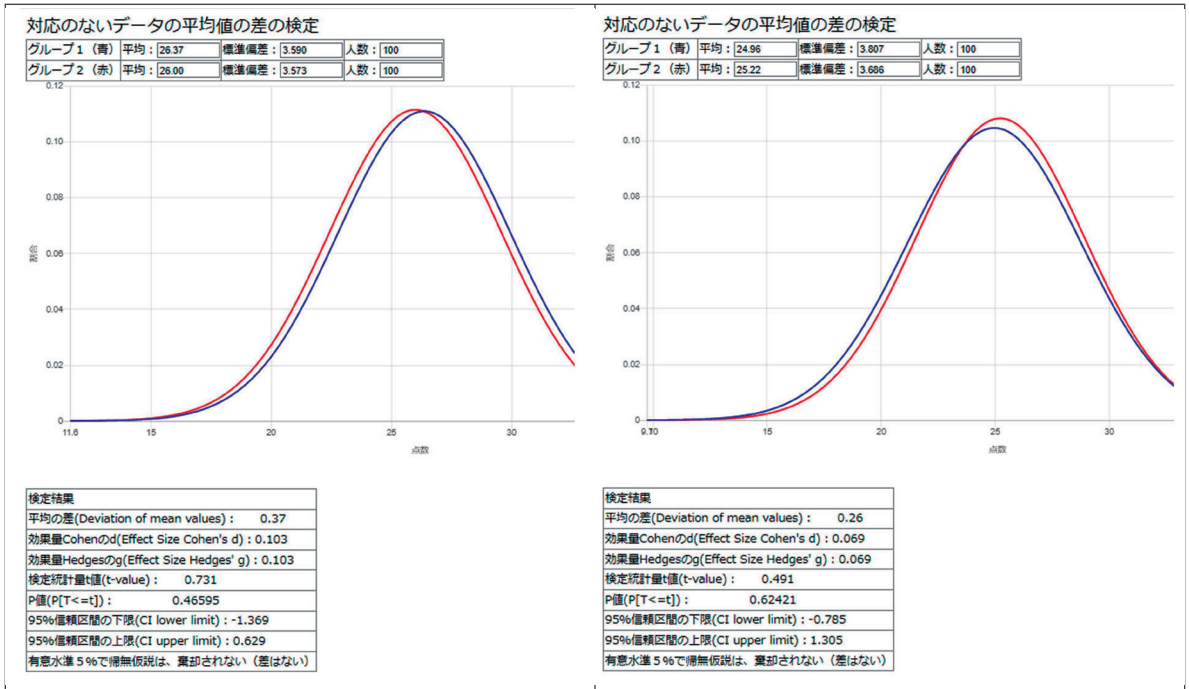
これら8項目の回答は「1 当てはまる」「2 どちらかといえば、当てはまる」「3 どちらかといえば、当てはまらない」「4 当てはまらない」の4件法である。そこで、データセットでは、これらの数値以外の誤入力等を空欄に置換し、空欄のある全てのIDをSPSSで削除した。また当てはまる方の値が小さいため、SPSSで「1→4」「2→3」「3→2」「4→1」に逆転した。

小6貸与匿名データでは、平成31年度はn=10,1698で、これら8項目の信頼性係数を計算したところ、 $\alpha = 0.72$ であった。どの項目を削除しても信頼性係数は低下することも確認した。スケールの統計量は、平均値26.37、標準偏差3.59となった。令和3年度はn=96,492、 $\alpha = 0.69$ 、平均値26.00、標準偏差3.57であった。令和4年度はn=929,990、 $\alpha = 0.70$ 、平均値26.12、標準偏差3.64であった。

中3データでは、平成31年度はn=91,600、 $\alpha = 0.72$ 、平均値24.96、標準偏差3.81であった。令和3年度はn=84,475、 $\alpha = 0.70$ 、平均値25.22、標準偏差3.69、令和4年度はn=84,674、 $\alpha = 0.70$ 、平均値25.21、標準偏差3.66である。

小6でも中3でも信頼性係数が0.69以上となったため、これらを合成変数としても良いと判断した。新しく開発した合成変数であるため、これを「非認知『徳』尺度（スコア）」と名づけた。名づけ親は宮城教育大学教職大学院の丸山千佳子特任教授である。学校は子どもの「知徳体」の成長に責任をもつ場所という認識が学校現場にあり、学力テストで測定される「知」と、体力テストで測定される「体」に対し、この合成尺度はそれらのテストでは直接測定されない「徳」の要素を測定すると考えたからである。データ分析では「Virtus」と略記する。

非認知「徳」スコアを算出してわかるのは、この全国値が年度でほとんど変動しないことである。平均値差検定システムで、小6と中3それぞれ、グループ1（青）に平成31年度、グループ2（赤）に令和3年度の平均、標準偏差を入れグラフ化すると図21、22になる。T検定はサンプルサイズの影響から無意味であるため、便宜的にn=100を代入した。



年度間に差がないことは、図21、22のグラフから直観的にうかがわれる。効果量もそれぞれ $d=0.10$ と $d=0.07$ であり、ほとんど差がないことを示している。全国規模で見れば、同一学年の学力スコアが年度間でほとんど差がないのと同じく、非認知「徳」スコアも「中心極限定理」により年度間でほとんど変動しないという仮説を置くことができる。

そうすると、これら3年度の匿名データを合算すれば、全国各学年の徳スコアの平均値が得られることになる。これら3年度分を合算すると、小6で $n=291,180$ 、 $\alpha=0.70$ 、平均値26.17、標準偏差3.60となり、中3で $n=261,749$ 、 $\alpha=0.71$ 、平均値25.13、標準偏差3.72となった。

以上の信頼性係数、平均値、標準偏差を一覧化すると、表15になる。

表15：非認知「徳」尺度の信頼性係数、平均値、標準偏差

	年度	信頼性係数 α	平均値	標準偏差
小6	R4	0.70	26.12	3.64
	R3	0.69	26.00	3.57
	H31	0.72	26.37	3.59
	3年度合成	0.70	26.17	3.60
中3	R4	0.70	25.21	3.66
	R3	0.70	25.22	3.69
	H31	0.72	24.96	3.81
	3年度合成	0.71	25.13	3.72

3年度合成スコアを基準値とすれば、各自治体や学校の非認知「徳」スコアの高低を評価できる。目的に応じて、比較対象となる年度の全国値を基準とすることもできる。学力スコアと同じく非認知「徳」スコアも、全国平均との差が効果量 d 値にして -0.50 以下となる場合は、改善が必要と評価できる。また全国平均と同程度であれば、個別の自治体や学校での非認知「徳」能力の育成が順調であると評価できる。

13. 2. 「対話・探究学習 (Dia_Inq) 尺度」の開発

全国学力・学習状況調査には、児童生徒が学校で対話や探究学習にどの程度取り組んでいるかを測定する質問項目も盛り込まれている。調査時期までの学校で、児童生徒が対話や探究学習にどのくらい取り組んできたか、児童生徒の自己評価を、これらの項目から私たちは知ることができる。そこで、本研究チームは、直近3年間に共通する6項目を抽出し、「対話・探究学習尺度 (スコア)」と名づけた合成尺度を開発した。データ処理上の略記号は「Dia_Inq」である。

抽出した項目群は、表16の通りである。表16は小6児童への質問紙項目であるが、中3生徒にも同じ番号で同じ質問がなされており、違いとしては、「5年生までに受けた授業」の部分が、「1、2年生の時に受けた授業」という表記になっているだけである。

表16：「対話・探究学習（Dia_Inq）」尺度の質問項目¹⁵¹

児童生徒質問紙項目一覧（R4～H31）					斜線部は、該当質問項目なし				
番号		質問項目【R4】	番号		質問項目【R3】	番号		質問項目【H31(R1)】	
小	中		小	中		小	中		
38	38	5年生まで(1,2年生のとき)に受けた授業で、自分の考えを発表する機会では、自分の考えがうまく伝わるよう、資料や文章、話の構成などで工夫して発表していたか ①当てはまる/②どちらかといえば、当てはまる/③どちらかといえば、当てはまらない/④当てはまらない	32	32	R4に同じ	36	38	R4に同じ	
39	39	5年生まで(1,2年生のとき)に受けた授業では、課題の解決に向けて、自分で考え、自分から取り組んでいましたか ①当てはまる/②どちらかといえば、当てはまる/③どちらかといえば、当てはまらない/④当てはまらない	33	33	R4に同じ	35	37	R4に同じ	
43	43	学級の友達(生徒)との間で話し合った活動を通じて、自分の考えを深めたり、広げたりすることができますか ①当てはまる/②どちらかといえば、当てはまる/③どちらかといえば、当てはまらない/④当てはまらない	37	37	R4に同じ	29	32	学級の友達と(生徒)の間で話し合う活動を通じて、自分の考えを深めたり、広げたりすることができますか ①当てはまる/②どちらかといえば、当てはまる/③どちらかといえば、当てはまらない/④当てはまらない	
45	45	総合的な学習の時間では、自分で課題を立てて情報を整理して、調べたことを発表するなどの学習活動に取り組んでいますか ①当てはまる/②どちらかといえば、当てはまる/③どちらかといえば、当てはまらない/④当てはまらない	39	39	R4に同じ	31	34	総合的な学習の時間では、自分で課題を立てて情報を整理して、調べたことを発表するなどの学習活動に取り組んでいますか ①当てはまる/②どちらかといえば、当てはまる/③どちらかといえば、当てはまらない/④当てはまらない	
46	46	あなたの学級では、学級生活をよりよくするために学級会(学級活動)で話し合い、互いの意見をよき生かして解決方法を決めていますか ①当てはまる/②どちらかといえば、当てはまる/③どちらかといえば、当てはまらない/④当てはまらない	40	40	R4に同じ	32	35	あなたの学級では、学級生活をよりよくするために学級会(学級活動)で話し合い、互いの意見をよき生かして解決方法を決めていますか ①当てはまる/②どちらかといえば、当てはまる/③どちらかといえば、当てはまらない/④当てはまらない	
47	47	学級活動における学級での話し合いを生かして、今、自分が努力すべきことを決めて取り組んでいますか ①当てはまる/②どちらかといえば、当てはまる/③どちらかといえば、当てはまらない/④当てはまらない	41	41	R4に同じ	33	36	学級活動における学級での話し合いを生かして、今、自分が努力すべきことを決めて取り組んでいますか ①当てはまる/②どちらかといえば、当てはまる/③どちらかといえば、当てはまらない/④当てはまらない	

文部科学省の分類では、令和4年度の質問番号38～44は「主体的・対話的で深い学びの視点からの授業改善に関する取組状況」に、質問番号45～48は「総合的な学習の時間、学級活動、特別の教科 道徳」にカテゴリ分けされている¹⁵²。本研究チームは、質問内容からして、これらは対話と探究学習に関わるまとまりのある項目群と解釈した。

回答法は非認知「徳」項目と同じ4件法である。

データ処理も、非認知「徳」項目と同じく、1～4以外の数値を空欄にし、空欄のあるIDを削除し、各項目を逆転した。処理後のデータ数は非認知「徳」項目と同じである。

それぞれの信頼性係数（ α ）、平均値、標準偏差は、表17のとおりである。

表17：対話・探究学習尺度の信頼性係数、平均値、標準偏差

	年度	信頼性係数 α	平均値	標準偏差
小6	R4	0.81	18.11	3.64
	R3	0.79	18.01	3.53
	H31	0.82	17.63	3.74
	3年度合成	0.81	17.91	3.65
中3	R4	0.82	17.96	3.59
	R3	0.80	17.85	3.55
	H31	0.83	16.99	3.85
	3年度合成	0.82	17.58	3.70

信頼性係数としてはどの年度、学年でも0.70以上あるため、合成尺度として妥当と評価できる。一覧からわかるように、この尺度の全国スコアも、3年程度ではさほど変動しない。

ただ、平成31年度から令和4年度にかけて、小6中3ともに、平均値が微増している。対話・探究的学習

¹⁵¹ 「令和4年度 全国学力・学習状況調査 報告書」, p.160より引用。

<https://www.nier.go.jp/22chousakekkahoukoku/report/question.html>

¹⁵² 注135を参照。

が全国的に徐々に盛んになっている兆候とも思われるが、3年度分だけの分析結果であるため即断は避けたい。

13. 3. 「国語授業充実度 (Jpn_Ju) 尺度」と「算数・数学授業充実度 (Math_Ju) 尺度」の開発

同じ要領で、国語の授業充実度や算数・数学の授業充実度に関する質問項目を合成し、新たな合成尺度を開発した。略記号はそれぞれ「Jpn_Ju」と「Math_Ju」である。

合成した項目は表18、19のとおり、それぞれ4項目である。

表18：「国語授業充実度 (Jpn_Ju)」尺度の質問項目 ¹⁵³								
児童生徒質問紙項目一覧 (R4～H31)								
番号 小 中		質問項目【R4】	番号 小 中	質問項目【R3】	番号 小 中	質問項目【H31(R1)】		
49	49	国語の勉強は好きですか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない	43	43	R4に同じ	37	40	R4に同じ
50	50	国語の勉強は大切だと思いますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない	44	44	R4に同じ	38	41	R4に同じ
51	51	国語の授業の内容はよく分かりますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない	45	45	R4に同じ	39	42	R4に同じ
52	52	国語の授業で学習したことは、将来、社会に出たときに役に立つと思いますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない	46	46	R4に同じ	40	43	R4に同じ

表19：「算数・数学授業充実度 (Math_Ju)」尺度の質問項目 ¹⁵⁴								
児童生徒質問紙項目一覧 (R4～H31)								
番号 小 中		質問項目【R4】	番号 小 中	質問項目【R3】	番号 小 中	質問項目【H31(R1)】		
53	53	算数(数学)の勉強は好きですか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない	52	52	R4に同じ	46	49	R4に同じ
54	54	算数(数学)の勉強は大切だと思いますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない	53	53	R4に同じ	47	50	R4に同じ
55	55	算数(数学)の授業の内容はよく分かりますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない	54	54	R4に同じ	48	51	R4に同じ
56	56	算数(数学)の授業で学習したことは、将来、社会に出たときに役に立つと思いますか ①当てはまる／②どちらかといえば、当てはまる／③どちらかといえば、当てはまらない／④当てはまらない	55	55	R4に同じ	49	52	R4に同じ

これらの質問項目群は、文部科学省の分類では「学習に対する興味・関心や授業の理解度等」となっている。私たちは、これらの項目を、授業の充実度と解釈し、「授業充実度尺度 (スコア)」と命名した。

データ処理としては、非認知「徳」スコアや対話・探究学習スコアと同じ置換・削除・逆転を実施した。先の2つのスコアと一つにしたデータセットで処理したので、データ数は、先の2つのスコアとすべて同じである。

それぞれの信頼性係数、平均値、標準偏差は、表20、21のとおりである。

¹⁵³ 「令和4年度 全国学力・学習状況調査 報告書」, pp.160-161より引用。
<https://www.nier.go.jp/22chousakekkahoukoku/report/question.html>

¹⁵⁴ 「令和4年度 全国学力・学習状況調査 報告書」, p.161より引用。
<https://www.nier.go.jp/22chousakekkahoukoku/report/question.html>

表20：国語授業充実度尺度の信頼性係数、平均値、標準偏差

	年度	信頼性係数 α	平均値	標準偏差
小6	R 4	0.72	13.11	2.30
	R 3	0.71	13.08	2.29
	H31	0.72	13.17	2.30
	3年度合成	0.75	13.12	2.30
中3	R 4	0.72	12.89	2.30
	R 3	0.74	12.75	2.38
	H31	0.74	12.64	2.46
	3年度合成	0.74	12.76	2.39

表21：算数・数学授業充実度尺度の信頼性係数、平均値、標準偏差

	年度	信頼性係数 α	平均値	標準偏差
小6	R 4	0.72	13.44	2.44
	R 3	0.73	13.68	2.41
	H31	0.75	13.57	2.47
	3年度合成	0.73	13.55	2.44
中3	R 4	0.75	12.30	2.76
	R 3	0.78	12.17	2.88
	H31	0.78	12.12	2.93
	3年度合成	0.77	12.19	2.86

いずれも $\alpha \geq 0.70$ なので、合成尺度として妥当と判断した。

この一覧からもわかるとおり、これらの合成スコアの全国平均や標準偏差も、3年を通してほとんど変動しない。

また国語と算数・数学の授業充実度は、それぞれわずく4項目であるが、それでも信頼性係数は0.70を上回った。それほど4項目の間には内的整合性が強く、項目間の相関係数も高いことになる。

例として令和3年度の国語充実度の項目間の相関係数を計算すると、表22の結果になった。

表22：R3_EL_Jpn_Juの項目間の相関

項目間の相関行列				
	STQ_043	STQ_044	STQ_045	STQ_046
STQ_043	1.000	.373	.494	.288
STQ_044	.373	1.000	.384	.543
STQ_045	.494	.384	1.000	.311
STQ_046	.288	.543	.311	1.000

特に高いのが項目44 (STQ_044) と46 (STQ_046) の相関で、 $r=0.54$ もある。項目44「国語の勉強は大切だと思いますか」と項目46「国語の授業で学習したことは、将来、社会に出たときに役に立つと思いますか」との質問である。勉強の大切さに、将来社会に出たときに役立つことが含まれるような関係と解釈できる。

両者に中程度の相関があるという結果も納得がゆく。

次に高いのが、項目43と45の相関係数0.49である。質問としては、項目43「国語の勉強は好きですか」と項目45「国語の授業の内容はよく分かりますか」とである。国語の授業がよくわかるから好きだ、また国語の授業が好きだからわかりにくい内容でも努力してわかるようになる、という心理構造が想定できる。相関係数で比較すると、図12に示したように、小6保護者の蔵書数と世帯収入の項目間相関は、0.31であった。これに比べても、上記2つの相関の強さがうかがわれる。

同様に、令和3年度小6算数の授業充実度項目間相関は、表23になる。

表23：R3_EL_Math_Juの項目間の相関

項目間の相関行列				
	STQ_052	STQ_053	STQ_054	STQ_055
STQ_052	1.000	.392	.620	.277
STQ_053	.392	1.000	.391	.591
STQ_054	.620	.391	1.000	.285
STQ_055	.277	.591	.285	1.000

$r=0.62$ と相関が最も高いのは、項目52「算数の勉強は好きですか」と項目54「算数の勉強はよく分かりますか」である。また項目53「算数の勉強は大切だと思いますか」と項目55「算数の授業で学習したことは、将来、社会に出たときに役に立つと思いますか」も、 $r=0.59$ と相対的に高い相関になっている。国語と同じ傾向が読み取れる。

令和3年度中3の授業充実度の相関行列も同様の結果であった。結果のみ記せば、令和3年度中3国語の項目44と46の相関係数は0.62、項目43と45の相関係数は0.52であった。数学の項目52と54の相関係数は0.62、項目53と55の相関係数は0.70であった。

13. 4. 学力スコアと非認知「徳」スコア等との相関

では、これらの合成変数は、相互にどれほど相関し、学力とはどれほど相関するだろうか。

令和4年度の小6と中3のデータをもとに、Rのpsychパッケージのpairs.panels関数で各スコアの相関係数 r 、ヒストグラム、散布図を一覧にしたのが、図23、24である。

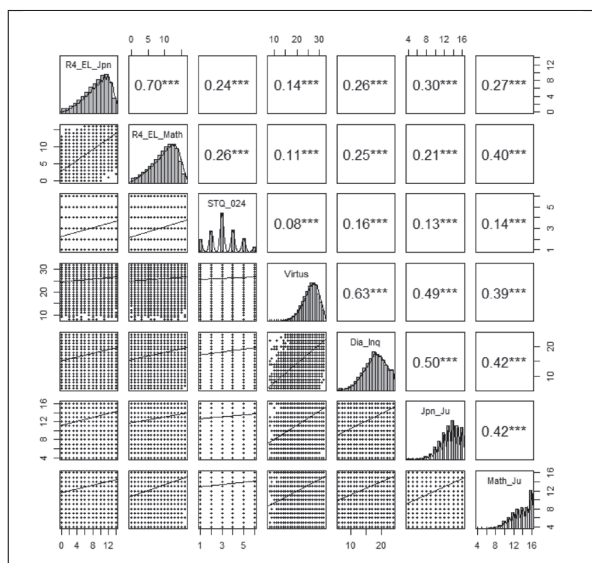


図23：令和4年度小6、学力非認知等相関一覧

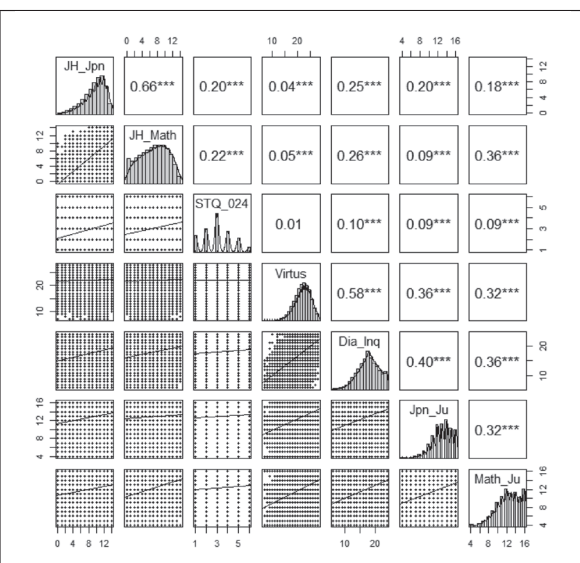


図24：令和4年度中3、学力非認知等相関一覧

令和3年度も同じく計算したが、結果は令和4年度とほぼ同じであった。

この一覧から、以下のことがわかる。

- 1) 国語と算数・数学の教科間の相関は0.70前後でやや強い。
- 2) SECSと学力との相関は、やはり0.25前後である。
- 3) SECSと非認知「徳」スコアとの相関はない。SECSが高くても非認知徳スコアが低い場合も多ければ、その逆もまた然りである。
- 4) 非認知「徳」スコアは学力とは相関しない。非認知「徳」スコアが高くても学力が高いとは限らないし、その逆もまた然りである。この知見は、「非認知能力が上がれば、学力も上がる」という素朴な見方が今回のデータでは裏づけられないことを示している。学力と非認知能力は相互に独立性の高い能力であるという仮説が成り立つことになる。教科間の学力の相関の強さとは対照的である。
- 5) 対話・探究学習は、SECSと同程度かやや強い程度に国語や算数・数学の学力と相関する。小学5年まで、あるいは中1中2と対話・探究学習に積極的に取り組んだと感じる児童生徒ほど、学力が高いという結果であり、その逆でもある。
- 6) 対話・探究学習は非認知「徳」能力と中程度の相関（小6で $r=0.63$ 、中3で $r=0.58$ ）がある。対話・探究学習に積極的に取り組んだと思う児童生徒ほど非認知「徳」スコアが高い。合成した各質問項目からうかがわれるように、非認知「徳」能力は、「知徳体」の育成に責任をもつ学校でぜひとも育成したい力である。この力を育成するには、対話・探究学習が効果的であるという仮説を支持する結果である。
- 7) 授業充実度は教科の学力と弱い相関にある。図23のように、小6では、国語の授業充実度が国語スコアだけでなく、算数スコアとも相関しているし、算数の授業充実度も算数スコアだけでなく、国語スコアとも相関している。図24の中3では、国語の充実度は国語のスコア、数学の充実度は数学のスコアのみとの相関である。この点が、令和3年度の分析と異なる点で、令和3年度は小6も中3も、国語充実度が国語と算数・数学と、算数・数学充実度が国語と算数・数学と相関していた。ある教科の授業充実度は当該教科ばかりでなく、他教科の学力スコアにも影響を与えるという仮説が成り立つ。
- 8) 国語と算数・数学の授業充実度は、非認知「徳」能力とも0.30以上、ときに0.50に迫るほど相関する。

これは令和3年度の分析でも同様だった。授業が充実することで、教科の学力ばかりでなく、児童生徒の非認知能力も育成されるという結果である。

14. 自治体支援の事例

教育データサイエンスのこうした知見や手法を、私たちは個別の自治体の支援に活かしている。2022年度現在で、私たちは都道府県単位の1自治体、市町村単位の3自治体と協働し、データの分析と支援を実施している。

本章では、先の13章で示した合成尺度を活用した自治体支援事例を紹介したい。自治体を匿名化するために、若干のフィクションを織り交ぜている。どのような分析・支援かの概略をお伝えしたい。

X自治体は、長年にわたり探究的な学習に取り組んできた先進地域である。幼稚園保育所にはじまり、小中高校を通して、地域と交流し、地域を調査し、地域に発信する学習に取り組んできた。一方で、こうした対話的で探究的な学習が、学力向上につながっているかは不明であった。またX自治体の分析によれば、ここ数年、全国学力・学習状況調査の全国平均正答率とのポイント差で、全国を下回る状況が続いていることが問題になっていた。そのため、探究学習は「活動あって学びなし」ではないか、「探究学習はもうやめて、学力向上にシフトすべきではないか」との批判も出ていた。加えて、探究学習に力を入れても、自己肯定感に関わる項目「自分には、よいところがあると思いますか」が全国平均と比べ毎年低いという結果になっており、探究学習は非認知能力育成にもつながっていないのではないか、との疑念も払しょくできなかった。

そこで私たちDS-EFAはX自治体の教育委員会と連携し、全国学力・学習状況調査の結果を共同で分析することにした。学力値については、「平均ゾーンシステム」を利活用し、非認知能力等については、13章で紹介した合成尺度により、全国を目安としてX自治体の成果を見積もることにした。

その結果、学力値については、国語は小6中3とも、全国学力・学習状況調査が新型コロナウイルス感染症で中止となった令和2年度を除く過去3年度で、全国との差がほとんどないことがわかった。効果量ではほんのわずかではあるが、学力向上も見られた。「学力が全国平均を下回る」というのは事実の一面でしかなかった。算数・数学はグラフと効果量から、確かに年々全国との差が開いている結果となったが、その差は効果量としては小さな差であることも判明した。

また、文部科学省からX自治体にフィードバックされた分析結果の国語と算数・数学のヒストグラムを、全国と比較したところ、算数・数学では、学力の最下位層の割合が全国よりも少ないことから、X自治体では最下層の底上げができていたことが判明した。対照的に学力の下位層の割合が全国よりも多く、学力の上位層や最上位層が全国よりも少ないことから、特に中位層の学力向上に焦点を絞った取り組みが必要であるとの認識にいたった。基礎基本の反復練習よりも、応用的あるいは発展的な問題解決や深い思考力の育成が今後の課題というところである。

非認知「徳」スコアは、小中ともに、令和3、4年度で全国平均よりも若干高いことがわかった。確かに自己肯定感の項目だけをみると全国平均を下回るが、関連する8項目をトータルすれば、全国平均を上回ったのである。これが合成尺度の強みである。特に、「人の役に立つ人間になりたいですか」と「地域や社会をよくするために何をすべきかを考えることがありますか」の項目が全国平均を大きく上回り、合成スコアを押し上げていた。地域密着型探究学習の成果がこの数値でも確認できた。

対話・探究学習スコアは、「平均値差検定システム」でグラフにすると、全国平均を圧倒的に上回っていた。これも協働での調べ学習や発表や地域発信を核とする探究学習の成果である。自治体規模でこれほど全国と差が開くのは、自治体をあげて取り組んできた成果に他ならない。13章で見たように、対話・探究学習スコアは学力と弱い相関があり、非認知「徳」スコアとは中程度の相関があるという全国の傾向からすると、探究学習に力を入れてきたことが、X自治体の児童生徒の学力と非認知「徳」スコアを押し上げていると強く推測できる。逆に、これまで積み上げてきた探究学習をやめるならば、X自治体の学力と非認知「徳」能

力はかえって低下してしまうと予想される。

一方、対話・探究学習スコアが全国平均よりもかなり高い割には、特に算数・数学の学力が伸び悩んでいることからすれば、探究学習を算数・数学の学力につながる内容に、今後バージョンアップしていくという課題も浮上した。

授業充実度は、小中ともに、また国語と算数・数学ともに、全国平均よりも若干高いこともわかった。

以上の分析結果から、X自治体の学校教育は比較的良く機能してきたという総合評価となった。そして、探究学習をやめるのではなく、進化（深化）させるという教育施策が適当であるとの共通認識にいたった。

15. むすびに代えて

本稿では、私たちの教育データサイエンスの紆余曲折の歩みを、現象学を手助けとして哲学的に省察してみた。現象学は、私たちが何をしているかの自己理解を深めてくれる。本稿で特に検討したのは、私たちの出発点となるリサーチクエスションの一つ、「全国平均正答率との2～5ポイント差は大問題なのか、それとも小問題なのか」である。全国平均正答率とのポイント差は、全国学力・学習状況調査の結果が話題になるとき、教育関係者や報道機関が最も注目する数値である。このポイント差は、教育関係者や報道機関によって生じられる数値であり、2～5ポイント差は少なくとも小問題ではないと前提にされてきた数値である。この自明性を、私たちは問いに付したわけである。

私たちの問いは、「全国平均正答率との2～5ポイント差には本当に意味があるのか」という問いでもある。同時にこの問いは、「このポイント差には意味がないかもしれない」という疑念を含んでいる。だからこそ、ポイント差の意味の有無を検証したわけである。本稿のここまでの到達点から見直すと、この問いは、意味の空洞化への危機意識を含んでいたことがわかる。この点で、私たちはフッサールの危機意識を大なり小なり共有していた。

フッサールにとって、ガリレオにはじまる科学技術の深刻な問題は、自然が数学化されることにあるのではなく、科学技術が完全性をめざし無限に進行することにあるのでもなく、生活世界に理念的世界がすりかえられることにあるのでもない。個や集団の人間を「数学化」することは、いかにも「非人間的」に映るかもしれない。しかし、数学を扱えるのが人間だけだとするなら、自然や人間の数学化はきわめて「人間的」ではないだろうか。質問項目に回答するのも人間、それを集計し分析し可視化し評価するのも人間となれば、数値やグラフに人間が現れないはずはないだろう。

科学技術の最大の危険性は、意味の空洞化を全世界にあまねく拡散させ増長させることにある。危機書の冒頭を思い出そう。近代人の全世界観は実証科学によって徹底的に規定され、「真の人間性にとって決定的な意味をもつ問題」が見失われてしまったこと、つまり人間にとって決定的な意味が空洞化されていることが、フッサールの危機意識だったと解釈できる。意味の空洞化の行きつく先は、「人間の生存全体に意味はない」である。もちろん本稿はこれほどの大問題を相手取ったわけではない。ただ、意味の空洞化を放置すれば、予想外の大きな人類的危機を招きかねないし、歴史的に招いてきたし、現在も招いていることを、現象学から学んだとは言える。

本稿におけるポイント差の意味の有無の検証は、帰無仮説検定の検討を経て、効果量にいたり、 $d=0.50$ を基準値とすればさほどの意味はない、という結論にいたった。ただし、この結論は絶対的な答えではなく、仮説（問い）にとどまり続ける。このことを私たちは自覚している。私たちの結論は仮説にとどまり続けるからこそ、科学的なのである。ポパーによれば、巧妙な答えを与えるのが占星術であり、問いであり続けるのが科学である。占星術や神話を、ポパーは「疑似科学」と言い換え、「疑似科学的態度のほうが、発生論的に言って、科学的態度よりも原初的であり、これに先立っている」とし、「科学は神話と共に始まり、神話の批判と共に始まるのでなくてはならない」という¹⁵⁵。

¹⁵⁵ ポパー, 1980, p.85.

この洞察もまた、私たちの教育データサイエンスの自己理解を促進する。教育現場や報道機関にも疑似科学はあまねく流布している。ポイント差はその氷山の一角にすぎない。私たちのデータサイエンスは、その疑似科学の批判であった。同時に、この疑似科学こそが、私たちのデータサイエンスの意味への問いの豊かな源泉である。それゆえ、私たちにとって「疑似」という言葉は、決して価値を貶める表現ではない。

例えば、私たちのデータサイエンスの出発点となるもう一つのリサーチクエスションは、「全国学力・学習状況調査の結果を経年比較できないか？」である。この問いは、統計学的な関心からではなく、全国学力・学習状況調査の結果を使って、個々の学校とか自治体の取り組みの成果や課題を測定できないか、という実践的な関心から生まれたものである。ここにも疑似科学があり、多くの自治体や学校では、過去の年度の全国平均とのポイント差で疑似的に経年比較がされてきた。その批判的発展が、私たちのデータサイエンスのもう一つの挑戦であった¹⁵⁶。ここでも疑似科学は、私たちの科学の苗床になっている。それゆえ、私たちの教育データサイエンスは、教育現場での疑似科学の批判から発生するという意味で発生論的性質をもつ。

この科学では、真理観の転換が起きている。あらゆる学問は真理探究であるが、多くの場合、真理探究の結果、なにがしかの真理を得たと主張する。デカルトが「われ考える、ゆえにわれ在り (コギト・エルゴ・スム)」を発見し、これを絶対的に確実な真理としたように。しかし、フッサールやポパーによれば、科学は仮説であり続け、何が真であることは立証できず、偽であると反証できるだけである。それゆえ、こうした科学の真理探究の結果言えることは、それは真ではないと反証するか、反証できないため今のところ真と仮定しよう、でしかない。真理を得たとは公言しない。

こうして転換された真理観に立つなら、現象学の科学性もまた浮き彫りになる。ガリレオの科学技術に「自然の数学化」を発見し、「人間に知覚される自然」という観察結果によってそれを反証・反駁するとき、フッサール現象学は科学とみなすことができる。フッサール現象学が科学であるならば、それは反証可能性に開かれているはずである。とするならば、知覚される自然と知覚する意識作用という志向性もまた、絶対的真理ではなく、一つの仮説ということになるだろう。もしもそうだとするならば、現象学という謎に満ちた哲学運動を科学的に理解する道が拓けてくる。フッサール自身もフッサールに共感する哲学者集団も、志向性を核とするフッサールのパラダイム内で自己批判を繰り返し、フッサール現象学を彫琢している。あるパラダイム内での自己批判—ターンならこれを「通常科学」と評するだろう—が現象学運動の一つの流れとなる。さらにもう一つ、別の流れがある。フッサール現象学を根本から批判し、全く新しい問いと答えの無限の連鎖の地平(新パラダイム)を拓くハイデガーと彼に共感する哲学者集団の現象学運動—さしずめ「科学革命」—である。知覚を土台とする生活世界は、「……のため」という有意義性のネットワークでもある。あるいはさらにハイデガーを根本的に批判する現象学運動もある。例えば、ユルゲン・ハーバーマスが発見したように、生活世界はコミュニケーション的行為の世界でもある。

ある時期以降ハイデガーが自身の思考を現象学とも哲学ともみなさなくなったとはいえ、現象学の創設者には、この2人を数えなくてはならない。この2人がワン・セットのお手本となり、のちの現象学運動が展開したと解釈できる。それゆえ、現象学は通常科学と科学革命の2タイプを原型(プロトタイプ)とするダイナミックな科学運動であり続けると解釈してみたい。ガリレオやニュートンの古典物理学が場の理論や相対性理論のような現代物理学へとパラダイム変革したとしても、物理学は物理学であり続ける。現象学運動の変革は、物理学の次のような変革と同じではないだろうか。

新しい理論は、古い理論のメリットと限界を示し、一段高い水準から旧来の諸概念を私たちに再獲得させるのである。これは電気流体や場の理論にとってだけでなく、物理学的理論のあらゆる変革

¹⁵⁶ その検証については、拙著(田端, 2022)を参照いただきたい。

—それらの変革がいかに革命的に見えようとも—にとつての真実である。¹⁵⁷

たとえるなら、新たな理論を創造することは、古い納屋を取りこわしてその跡に摩天楼を建てるようなものではない。むしろ、山を登るにつれ、新しく一段と広い展望が開け、出発点とその豊かな周囲世界との予期せぬ関連を発見するのに似ている。¹⁵⁸

同様に、フッサール現象学がいかにパラダイムシフトしようと、現象学は現象学であり続ける。それがいかに革新的であろうと、新しい現象学は古い現象学の諸概念を一段高い水準あるいは別の水準から再獲得することを可能にする。「私たちの出発点は依然として存在し、眺めることができる」¹⁵⁹。新しい地点から出発点の予想外の豊かさが再発見できるなら、その新たな地点はまぎれもなく現象学であろう。現象学は、そのすべての言明においてではないとしても、基本において科学であるからこそ、神話や魔術でないことはもとより、マルクスの歴史論やフロイトの精神分析やアドラーの個人心理学のような非科学とも一線を画する。こうした仮説を置いて、本稿のむすびに代えたい。

【付記1】本稿は科学研究費助成事業、基盤研究B「グローバル世界を視野とする学力・非認知能力の効果的学校モデル」（2020-2022年度、課題番号：20H01667、研究代表者：田端健人）の研究成果の一部である。

【付記2】本稿で適宜明記したとおり、本稿の一部は、文部科学省の個票データ等の貸与による成果である。また、文部科学省総合教育政策局が作成・公表した資料と、本研究が独自に作成・加工した資料とは、本稿の説明のとおり明確に区別している。

【付記3】「個票データ等の貸与利用規約」に則り、本稿は公表以前のしかるべき時期に、文部科学省総合教育政策局による事前確認を受けている。

¹⁵⁷ Einstein, A. and Infeld, L., 2007, *The Evolution of Physics: From Early Concepts to Relativity and Quanta, Touchstone*, New York, pp.151-152. アインシュタイン・インフェルト, 2010『物理学はいかに創られたか 上巻』石原純訳, 岩波新書, pp.174-175.

¹⁵⁸ Einstein and Infeld, 2007, p.152. アインシュタイン・インフェルト, 2010, p.175.

¹⁵⁹ Einstein and Infeld, 2007, p.152. アインシュタイン・インフェルト, 2010, pp.175-176.

資料：全国とトップ都道府県およびボトム都道府県の平均値差の効果量 (R4-H28, H21-H19)

※ $d \geq 0.50$ を黄色でハイライトした。

R4	EL_Jpn	EL_Math	JH_Jpn	JH_Math
問題数	14	16	14	14
トップ平均正答	9.9	11.0	10.3	8.1
トップ標準偏差	2.9	3.3	2.8	3.5
トップ人数	5,928	8,588	8,970	8,974
全国平均正答数	9.2	10.1	9.7	7.2
全国標準偏差	3.3	3.6	2.9	3.6
全国人数	965,308.0	965,431	891,820	891,913
ボトム平均正答	8.8	9.5	9.0	5.9
ボトム標準偏差	3.2	3.7	3.2	3.5
ボトム人数	14,451	14,452	13,208	13,210
トップ-全国 平均値差	0.7	0.9	0.6	0.9
トップ-全国 効果量d	0.23	0.29	0.21	0.25
相関係数r	0.11	0.14	0.10	0.12
ボトム-全国 平均値差	0.4	0.6	0.7	1.3
ボトム-全国 効果量d	0.12	0.16	0.23	0.37
相関係数r	0.06	0.08	0.12	0.18

R3	EL_Jpn	EL_Math	JH_Jpn	JH_Math
問題数	14	16	14	16
トップ平均正答	10.0	11.9	9.6	10.1
トップ標準偏差	2.8	3.1	2.8	3.6
トップ人数	8,986	8,989	9,031	9,037
全国平均正答数	9.1	11.2	9.0	9.1
全国標準偏差	3.1	3.5	2.8	3.7
全国人数	993,975	994,101	903,157	903,253
ボトム平均正答	8.6	10.8	8.4	8.2
ボトム標準偏差	3.3	3.4	3.0	3.8
ボトム人数	10,386	5,418	13,561	13,556
トップ-全国 平均値差	0.9	0.7	0.6	1.0
トップ-全国 効果量d	0.31	0.21	0.21	0.27
相関係数r	0.15	0.10	0.10	0.13
ボトム-全国 平均値差	-0.5	-0.4	-0.6	-0.9
ボトム-全国 効果量d	0.16	0.12	0.21	0.24
相関係数r	0.08	0.05	0.10	0.12

H31	EL_Jpn	EL_Math	JH_Jpn	JH_Math
問題数	14	14	10	16
トップ平均正答	10.4	10.0	7.8	10.6
トップ標準偏差	2.9	2.8	2.1	4.0
トップ人数	6,960	9,843	7,242	6,564
全国平均正答数	8.9	9.3	7.3	9.6
全国標準偏差	3.4	3.1	2.4	4.2
全国人数	1,028,203	1,028,177	938,797	938,887
ボトム平均正答	8.2	9.0	6.8	8.4
ボトム標準偏差	3.5	3.1	2.5	4.3
ボトム人数	67,141	38,837	14,044	14,042
トップ-全国 平均値差	1.5	0.7	0.5	1.0
トップ-全国 効果量d	0.48	0.24	0.22	0.24
相関係数r	0.23	0.12	0.11	0.12
ボトム-全国 平均値差	-0.7	-0.3	-0.5	-1.2
ボトム-全国 効果量d	0.20	0.10	0.20	0.28
相関係数r	0.10	0.05	0.10	0.14

H30	EL_Jpn_A	EL_Jpn_B	EL_Math_A	EL_Math_B	JH_Jpn_A	JH_Jpn_B	JH_Math_A	JH_Math_B
問題数	12	8	14	10	32	9	36	14
トップ平均正答	9.2	4.9	9.5	5.9	25.6	5.9	25.9	7.4
トップ標準偏差	2.4	1.9	3.0	2.5	4.4	1.6	7.4	3.4
トップ人数	7,255	7,253	9,619	9,616	7,364	7,365	6,668	6,666
全国平均正答数	8.5	4.4	8.9	5.1	24.3	5.5	23.8	6.6
全国標準偏差	2.7	1.9	3.2	2.7	5.3	2.0	8.1	3.5
全国人数	1,030,025	1,029,799	1,030,013	1,029,847	966,764	966,786	966,969	966,908
ボトム平均正答	8.1	4.2	8.5	4.9	22.9	5.2	21.3	5.5
ボトム標準偏差	2.8	1.9	3.2	2.6	5.7	2.1	8.4	3.3
ボトム人数	66,893	66,885	13,198	13,204	14,063	12,412	14,064	14,059
トップ-全国 平均値差	0.7	0.5	0.6	0.8	1.3	0.4	2.1	0.8
トップ-全国 効果量d	0.27	0.26	0.19	0.31	0.27	0.22	0.27	0.23
相関係数r	0.13	0.13	0.09	0.15	0.13	0.11	0.13	0.11
ボトム-全国 平均値差	-0.4	-0.2	-0.4	-0.2	-1.4	-0.3	-2.5	-1.1
ボトム-全国 効果量d	0.15	0.11	0.13	0.08	0.25	0.15	0.30	0.32
相関係数r	0.07	0.05	0.06	0.04	0.12	0.07	0.15	0.16

H29	EL_Jpn_A	EL_Jpn_B	EL_Math_A	EL_Math_B	JH_Jpn_A	JH_Jpn_B	JH_Math_A	JH_Math_B
問題数	15	9	15	11	32	9	36	15
トップ平均正答	12.0	5.8	12.7	5.8	26.3	7.0	26.2	8.2
トップ標準偏差	2.3	2.0	2.6	2.6	4.7	1.9	7.5	3.3
トップ人数	7,183	7,182	9,460	9,451	7,656	7,656	6,938	6,938
全国平均正答数	11.2	5.2	11.8	5.1	24.8	6.5	23.3	7.2
全国標準偏差	2.8	2.2	3.1	2.6	5.7	2.3	8.5	3.3
全国人数	999,683	999,498	999,723	999,451	982,511	982,619	982,811	982,786
ボトム平均正答	10.8	4.9	11.4	4.7	23.0	6.0	20.7	6.3
ボトム標準偏差	2.9	2.2	3.2	2.5	6.1	2.4	8.7	3.0
ボトム人数	68,770	68,745	12,756	5,402	14,418	14,413	14,413	14,411
トップ-全国 平均値差	0.8	0.6	0.9	0.7	1.5	0.5	2.9	1.0
トップ-全国 効果量d	0.31	0.32	0.29	0.27	0.29	0.24	0.36	0.30
相関係数r	0.15	0.16	0.14	0.13	0.14	0.12	0.18	0.15
ボトム-全国 平均値差	-0.4	-0.3	-0.4	-0.4	-1.8	-0.5	-2.6	-0.9
ボトム-全国 効果量d	0.14	0.14	0.13	0.16	0.31	0.21	0.30	0.29
相関係数r	0.07	0.07	0.06	0.08	0.15	0.10	0.15	0.14

H28	EL_Jpn_A	EL_Jpn_B	EL_Math_A	EL_Math_B	JH_Jpn_A	JH_Jpn_B	JH_Math_A	JH_Math_B
問題数	15	10	16	13	33	9	36	15
トップ平均正答	11.8	6.4	13.2	7.0	26.1	6.5	25.0	7.6
トップ標準偏差	2.9	2.3	2.9	2.7	4.5	2.0	7.5	3.6
トップ人数	9,674	7,431	9,675	9,665	7,831	7,828	7,169	7,167
全国平均正答数	10.9	5.8	12.4	6.1	25.0	6.0	22.4	6.6
全国標準偏差	3.1	2.4	3.4	2.8	5.4	2.3	8.3	3.5
全国人数	1,021,905	1,021,635	1,021,910	1,021,637	996,188	996,365	996,502	996,578
ボトム平均正答	10.5	5.5	12.0	5.8	23.5	5.7	19.5	5.6
ボトム標準偏差	3.0	2.5	3.5	2.8	5.7	2.4	8.2	3.2
ボトム人数	7,572	71,895	40,276	40,262	14,731	70,335	14,733	14,736
トップ-全国 平均値差	0.9	0.6	0.8	0.9	1.1	0.5	2.6	1.0
トップ-全国 効果量d	0.30	0.26	0.25	0.33	0.22	0.23	0.33	0.28
相関係数r	0.15	0.13	0.12	0.16	0.11	0.11	0.16	0.14
ボトム-全国 平均値差	-0.4	-0.3	-0.4	-0.3	-1.5	-0.3	-2.9	-1.0
ボトム-全国 効果量d	0.13	0.12	0.12	0.12	0.27	0.13	0.35	0.30
相関係数r	0.06	0.06	0.06	0.06	0.13	0.06	0.17	0.15

H21	EL_Jpn_A	EL_Jpn_B	EL_Math_A	EL_Math_B	JH_Jpn_A	JH_Jpn_B	JH_Math_A	JH_Math_B
問題数	18	10	18	14	33	11	33	15
トップ平均正答	13.6	6.0	15.5	8.9	27.1	9.0	23.3	9.8
トップ標準偏差	2.9	2.0	2.7	3.1	5.1	2.2	6.7	3.7
トップ人数	9,200	9,200	9,200	9,199	10,076	10,074	7,847	7,851
全国平均正答数	12.6	5.1	14.2	7.7	25.4	8.2	20.7	8.5
全国標準偏差	3.4	2.3	3.4	3.3	6.1	2.7	7.6	3.9
全国人数	1,137,808	1,137,598	1,137,844	1,137,624	1,033,291	1,033,703	1,033,885	1,033,909
ボトム平均正答	11.6	4.6	13.3	6.9	22.9	7.5	17.0	6.8
ボトム標準偏差	3.6	2.3	3.6	3.4	6.6	2.9	7.0	3.6
ボトム人数	15,862	15,858	45,920	15,856	15,522	15,523	15,524	15,528
トップ-全国 平均値差	1.0	0.9	1.3	1.2	1.7	0.8	2.6	1.3
トップ-全国 効果量d	0.32	0.42	0.42	0.38	0.33	0.33	0.36	0.34
相関係数r	0.16	0.21	0.21	0.19	0.16	0.16	0.18	0.17
ボトム-全国 平均値差	-1.0	-0.5	-0.9	-0.8	-2.5	-0.7	-3.7	-1.7
ボトム-全国 効果量d	0.29	0.22	0.26	0.24	0.39	0.25	0.51	0.45
相関係数r	0.14	0.11	0.13	0.12	0.19	0.12	0.25	0.22

H20	EL_Jpn_A	EL_Jpn_B	EL_Math_A	EL_Math_B	JH_Jpn_A	JH_Jpn_B	JH_Math_A	JH_Math_B
問題数	18	12	19	13	34	10	36	15
トップ平均正答数	13.4	7.6	15.3	7.7	26.7	6.7	26.0	8.8
トップ標準偏差	3.4	2.9	3.0	2.8	4.8	2.3	7.9	3.5
トップ人数	9,496	9,494	9,497	9,494	9,837	9,835	7,692	7,691
全国平均正答数	11.8	6.1	13.7	6.7	25.0	6.1	22.7	7.4
全国標準偏差	4.0	3.1	3.7	3.0	5.8	2.5	9.0	3.7
全国人数	1,147,582	1,147,406	1,147,633	1,147,475	1,028,550	1,028,841	1,029,014	1,029,104
ボトム平均正答数	10.4	5.4	12.6	5.9	23.1	5.4	17.9	5.7
ボトム標準偏差	4.1	3.2	3.8	2.9	6.1	2.5	8.7	3.3
ボトム人数	16,503	16,502	16,503	16,496	15,521	15,528	15,519	15,513
トップ-全国 平均値差	1.6	1.5	1.6	1.0	1.7	0.6	3.3	1.4
トップ-全国 効果量d	0.43	0.50	0.48	0.35	0.32	0.25	0.39	0.39
相関係数r	0.21	0.24	0.23	0.17	0.16	0.12	0.19	0.19
ボトム-全国 平均値差	-1.4	-0.7	-1.1	-0.8	-1.9	-0.7	-4.8	-1.7
ボトム-全国 効果量d	0.35	0.22	0.29	0.27	0.32	0.28	0.54	0.49
相関係数r	0.17	0.11	0.14	0.13	0.16	0.14	0.26	0.24

H19	EL_Jpn_A	EL_Jpn_B	EL_Math_A	EL_Math_B	JH_Jpn_A	JH_Jpn_B	JH_Math_A	JH_Math_B
問題数	18	10	19	14	37	10	36	17
トップ平均正答	15.5	6.9	16.8	9.6	31.7	7.7	28.9	11.5
トップ標準偏差	2.3	2.3	2.7	2.7	4.9	2.2	7.1	4.0
トップ人数	9,616	9,615	9,616	9,614	9,438	9,441	7,754	7,756
全国平均正答数	14.7	6.2	15.6	8.9	30.2	7.2	25.9	10.3
全国標準偏差	2.8	2.6	3.4	3.0	5.5	2.4	8.7	4.3
全国人数	1,125,575	1,125,422	1,125,585	1,125,622	1,022,518	1,023,009	1,023,316	1,023,516
ボトム平均正答	13.8	5.3	14.5	7.6	27.5	6.4	20.6	8.1
ボトム標準偏差	3.1	2.7	3.7	3.0	6.3	2.6	8.9	4.2
ボトム人数	16,093	16,090	16,093	16,087	15,745	15,737	15,749	15,749
トップ-全国 平均値差	0.8	0.7	1.2	0.7	1.5	0.5	3.0	1.2
トップ-全国 効果量d	0.31	0.29	0.39	0.25	0.29	0.22	0.38	0.29
相関係数r	0.15	0.14	0.19	0.12	0.14	0.11	0.19	0.14
ボトム-全国 平均値差	-0.9	-0.9	-1.1	-1.3	-2.7	-0.8	-5.3	-2.2
ボトム-全国 効果量d	0.31	0.34	0.31	0.43	0.46	0.32	0.60	0.52
相関係数r	0.15	0.17	0.15	0.21	0.22	0.16	0.29	0.25