

大規模国際学力調査におけるログファイル及びプロセスデータの 利活用に関する研究事例について

– 2019年国際会合（Opportunity versus Challenge : Exploring Usage of Log-File and Process Data in International Large Scale Assessment）の議論を中心に –
Research Trends on the Usage of Log-File and Process Data in International Large Scale Assessments

猿田 かほる
SARUTA Kahoru

Abstract

In recent years, with the development of ICT (Information and Communication Technology), the shift to ICT in education has become an urgent issue in Japan. In this context, the "Expert Committee on the Utilization of Educational Data" has been discussing the utilization of educational data as a pressing matter in Japan. As CBT is now able to collect the log data of test takers, which could not be collected through the conventional Paper Based Testing (PBT), there is an urgent need for the usage of data.

Therefore, this paper reports on the "Log-File and Process Data Meeting and Training" that the author attended in 2019, and summarizes the current status of discussions on the usage of log-data in international large-scale academic surveys.

In assessments conducted through CBT, log-file data is inevitably generated. It is necessary to define in advance what to use as process data in accordance with the purpose of utilizing the data. In other words, we should not just utilize process data in the dark, but should carefully consider the issues and problems involved in usage of the process data. In addition, the results of analysis and research using process data cannot be simply generalized, and abstract discussions based on the results should be avoided. In addition, it is necessary to consider ethical issues when utilizing the process data.

1. はじめに

近年、我が国では高度情報技術の進展に伴って教育の ICT 化が急務として進められている。2019 年 12 月 19 日に「GIGA スクール実現推進本部」が設置された。GIGA スクール構想は、Society 5.0 時代を生きる子供たちにとって、教育における ICT を基盤とした先端技術等の効果的な活用が求められる一方で、現在の学校 ICT 環境の整備は遅れており、自治体間の格差も大きい。また、令和時代のスタンダードな学校像として、全国一律の ICT 環境整備が急務であることを背景としており、1 人 1 台端末及び高速大容量の通信ネットワークを一体的に整備するとともに、並行してクラウド活用推進、ICT 機器の整備調達体制の構築、利活用優良事例の普及、利活用の PDCA サイクル徹底等を進めることで、多様な子供たちを誰一人取り残すことのない、公正に個別最適化された学びを全国の学校現場で持続的に実現させることが目指されている(文部科学省, 2019)。

また、こうした動きを踏まえ、全国学力・学習状況調査のコンピュータ使用型調査 (Computer Based Testing、以下 CBT) 化について、「全国的な学力調査の CBT 化検討ワーキンググループ」において、専門的・技術的な観点から検討が行われている最中である。

大規模国際学力調査に焦点化すると、OECD が実施する「生徒の学習到達度調査」(Programme for International Student Assessment、以下 PISA) では、2015 年調査から CBT へ移行 (国立教育政策研究所, 2019)¹、IEA (国際教育到達度評価学会) が実施する「国際数学・理科教育動向調査」(Trends in International Mathematics and Science Study、以下 TIMSS) では 2023 年調査から完全 CBT 移行を予定している (国立教育政策研究所, 2020)²。また、OECD が実施する「国際成人力調査」(Programme for the International Assessment of Adult Competencies、以下 PIAAC) においても、2011 年調査から CBT が導入されている (国立教育政策研究所, 2013)³。

日本国内において教育の ICT 化が進められている中で、並行して教育データの利活用についても喫緊の課題として、「教育データの利活用に関する有識者会議」において、議論が進められている。従来の筆記型調査 (Paper Based Testing、以下 PBT) では収集できなかったテスト受検者のログデータが、CBT では収集可能になったことにより、そのデータの利活用の目的や方法を整理し、利活用の在り方を検討することも急務である。近年、LA (Learning Analytics) や EDM (Educational Data Mining) の研究領域において、MOOC (大規模公開オンライン講義) や LMS (Learning Management System)、教育用デジタルコンテンツ等の学習ログに関する分析・研究 (例えば、野田・金子, 2020; 小柳, 2019; 李, 2018; 田口ら, 2017; 緒方ら, 2016a; 緒方ら,

1 日本は、PISA2009 年調査時から、国際オプションである 40 分間の CBT に参加しており、PISA2015 年調査からは全面 CBT 移行した。PISA2018 年調査に参加した 79 か国・地域のうち PBT で実施した国は 9 か国・地域で、それ以外は CBT で実施した (国立教育政策研究所, 2019)。

2 TIMSS2019 年から、基本的に CBT 実施となり、並行して PBT 実施も認められている。そのため、日本は PBT で調査を実施した (国立教育政策研究所, 2020)。

3 PIAAC では、調査対象者の自宅等において、専用のコンピュータを用いて対面方式で調査が実施される。背景調査では、調査員が質問項目を読み上げ、回答を入力する。読解力、数的思考力、IT を活用した問題解決能力の調査は、対象者本人がコンピュータを用いて解答する。しかし、①背景調査において「コンピュータを使った経験がない」と回答した場合、②自らコンピュータ調査を拒否した場合、③コンピュータの導入試験 (ICT コア) で「不合格」になった場合は、PBT で実施する。PBT の場合、IT を活用した問題解決能力の調査は行われない (国立教育政策研究所, 2013)。

2016b；山田，2015；久保田，2015 など）の蓄積は活発に行われてきているが、大規模国際学力調査でのログデータの利活用については、日本国内においてほとんど先行研究が行われていないのが現状である。

そこで本稿では、筆者が2019年5月に参加した「ログファイル及びプロセスデータ会合・研修」の報告をすることで、大規模国際学力調査におけるログデータの利活用に関して国際的にどのような議論がなされているのか、現状についての論点をまとめる。

2. ログファイル及びプロセスデータ概念とその利活用について

まず「ログファイル及びプロセスデータ会合・研修」の概要について紹介し、そこで議論されたログファイルやプロセスデータとは何かという概念や定義について解説する。

2.1. ログファイル及びプロセスデータ会合・研修

The Educational Research Centre（ダブリン：アイルランド、以下 ERC）、Educational Testing Service（プリンストン：アメリカ、以下 ETS）の共同開催による「ログファイル及びプロセスデータ会合・研修（Opportunity versus Challenge: Exploring Usage of Log-File and Process Data in International Large Scale Assessments）」（以下、「会合・研修」）は、2019年5月16日から17日にアイルランド・ダブリンの Hotel Riu Plaza the Gresham を会場に開催された⁴。「会合・研修」には主催機関である ETS、ERC 関係者をはじめ、各国の教育機関関係者、研究者、PISA 及び PIAAC 等の調査実施機関関係者ら約 60 名が集まった。1 日目と 2 日目の午前で開催されたログファイル及びプロセスデータに関する会合では、ログファイル及びプロセスデータに関する概念・定義、研究分野の今後の方向性及び優先事項、方法論的・理論的アプローチ、倫理的問題、プロセスデータの活用分析・研究事例等について報告・説明を受け、議論をするというものであった。2 日目午後開催されたログファイル及びプロセスデータに関する研修では、PISA2015 年予備調査のデータから ETS が作成した生徒の解答プロセスに関するデータベースを用いて、事前に参加者から収集したリサーチクエスチョンの中から主催者によって抽出された 4 つに大別されたリサーチクエスチョン（①生徒が項目において、初動までにかかった時間のデータをどのように利用するか。国によって類似パターンがあるかどうか。また、性別による違いがあるか。②生徒の初動までに時間をかけるパターンと成績にはどのような関係があるか。③1 つの大問の中で、生徒たちはどのような方略で問題項目を解いているか。彼らの問題を解く順序について。また、国や性別による違いはあるか。④反応行動パターンは大問間で一貫しているか。）に沿ってプロセスデータ分析について、演習形式で実際の手順等を確認し、それに基づいて検討を行った。

2.2. ログファイルとは

2.2.1. ログファイルの定義

Stephen 氏の基調講演（Provansik, 2019a）によると、ログファイルとは CBT において記録され

4 会合・研修の一部資料は、ERC のホームページからダウンロードできる（<https://www.erc.ie/programme-of-work/independent-research/student-performance-in-pisa-at-different-times-of-the-year/opportunity-versus-challenge/>）。なお、本稿に掲載されている会合・研修の内容は全て、発表者らから公開許可を得ている。

た全てのデータのことを指す。すなわち、受検者が入力を行った順序や速度（クリックやキーストロークなど）から、調査を実施するために使用されたデバイスのVPN⁵まで、ありとあらゆるものがログファイルとして収集される。一方、受検者がテスト問題に取り組む過程（プロセス）を反映した実証的なデータのことをプロセスデータと言う。認知的及び非認知的、特に心理学的な構成要素を反映している。ログファイルの中で、そのままプロセスデータとして扱うことができるものもあれば、ログファイルを変換又は処理することでプロセスデータとして扱うことができる場合もある。つまり、調査中に収集される全てのデータがログファイルであり、その中の一部がプロセスデータなのである。換言すると、「ログファイルはプロセスデータのデータソース」（Provansik, 2021 : 2）ということである。ログファイルとは、CBT 開発の過程で生成された副産物であり、また、プロセスデータ自体はログファイルの意図的な活用（ログファイルからの抽出）のことを意味する（同上）。

例えば、PISA2012年調査における読解力のCBTパイロット調査の結果をもとにした分析（OECD, 2016）においては、生徒のナビゲーション行動が以下のような手順でログファイルから変換した上で、プロセスデータとして扱われている。調査問題の中で、各課題に関してナビゲーションシーケンスに含まれるステップ（異なるページへつながるタブやリンクをクリックすること）の数は課題を通して合計される。この数を全体的なブラウジング活動指標に変換するために、同じデジタル読解力の問題を与えられた生徒における順位を示すパーセントイル順位が計算されている。その際には、すべての参加国からの生徒に関する重み付けなしのプールされた分布が使用されている（同上）。

しかしながら、ログファイルとプロセスデータの線引きは曖昧で、何をプロセスデータと定義するかによって、その線引きが変化する。さらに、分析で使用するために「プロセスデータから新しい変数を作成することは複雑であり、教育測定分野全体では、この複雑さが理解され始めたばかりである。生データと採点された解答／回答の品質保証のためのガイドラインは数多く存在するが、プロセスデータ変数の作成とその品質保証のための実践基準に関する文献は限られている」（OECD, 2020 : 2）のが現状である。

2.2.2. 教育測定の変遷

教育測定における問題開発、問題形式、テスト設計、テストの組立て、アクセシビリティタイミング、経路、妥当性、フィードバックの9つの側面について、過去、現在、そして未来でどのように変化してきたか、またこれから変化するであろうか、その在り方の変遷を整理したものが図表1である（Provansik, 2019a）。この図表では、現在の大規模国際学力調査がどのような方向に向かっているのか、その流れが示されている。

例えば、従来のPBTでは、テスト設計は1ページ目から最終ページまで試験中に変わることがなく静的で、タイミングや経路といったものは測定できなかったが、現在PISAなどで行われているCBTでは、タイミングや経路といったログファイルが収集できるようになり、受検者によってテスト問題が変わるなどテスト設計及び組立てが半静的、半自動化になっている。そして未来の調査では、受検者のデータによって調査問題を自動的に開発してゆくことができ、またロ

5 VPNとは、「Virtual Private Network」の略で、通信事業者のネットワークやインターネットなどの公衆ネットワーク上で作られる、仮想的な専用ネットワークの総称（JPNIC, 2017）。

図表 1. 教育測定の変遷

	過 去	現 在	未 来
問題開発 Item Development	労働集約型 Labor Intensive	労働集約型 Labor Intensive	自動化 Automatized
問題形式 Item types	一般的 Generic	強化された Enhanced	実生活 Real-life
テスト設計 Test design	静的 Static	半静的 Semi-static	データドリブン Data-driven
テストの組立て Test assembly	労働集約型 Labor Intensive	半自動化 Semi-automatized	自動化 Automatized
アクセシビリティ Accessibility	限定的 Limited	ユニバーサル・デザイン Universal design	適応型 Adaptive
タイミング Timing	測定不能 Not measurable	測定可能 Measured	活用可能 Used
経路 Pathways	観測不能 Not observable	観測可能 Observable	モデル化 Modeled
妥当性 Validity	コンテンツ/コアベース Content/core-based	コンストラクトベース Construct based	プロセスベース Process based
フィードバック Feedback	総括的評価 Summative	総括的評価 Summative	診断的評価 Diagnostic

出典：Provansik (2019a)
日本語訳は筆者による。

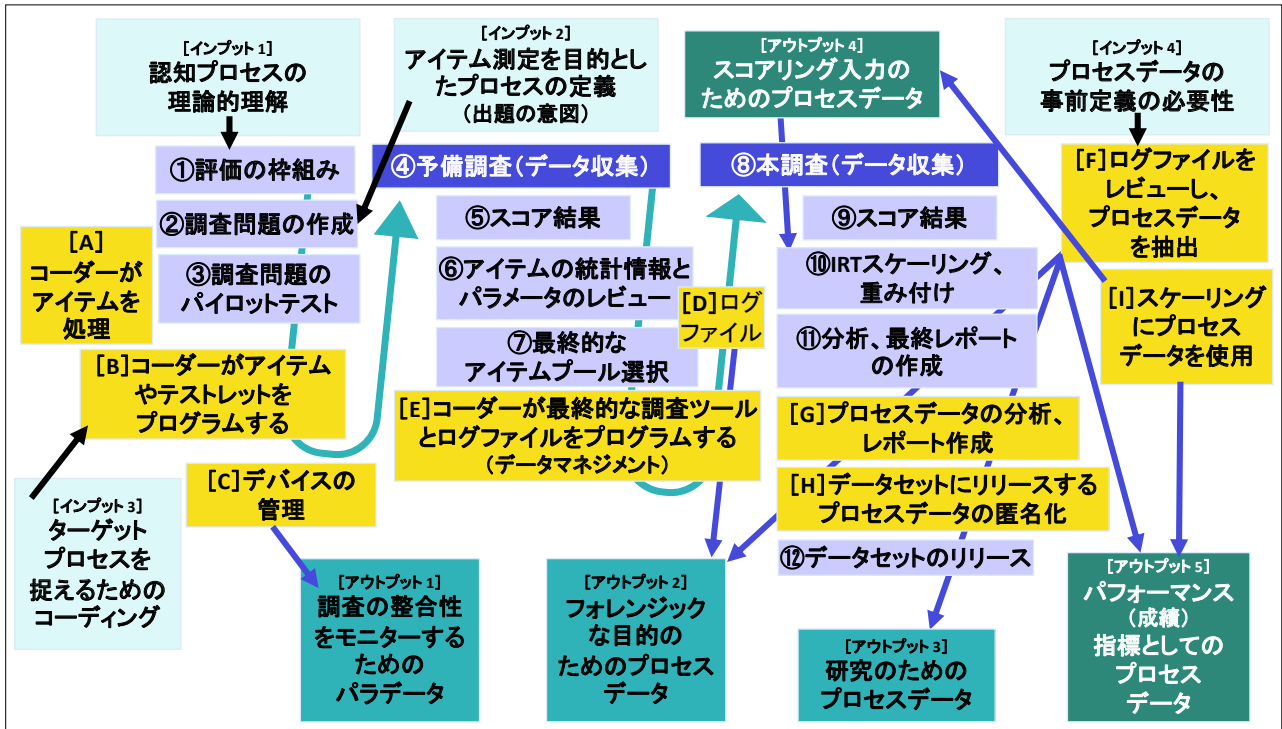
ログファイルが収集できるだけでなく、効果的な利活用が可能となり、またその結果から受検者等へのアドバイスやフィードバックとして返してゆくことができるようになる、と見込まれている。

2.2.3. ログファイル及びプロセスデータが生成・活用される過程

CBT 開発の一連の流れを整理したものが図表 2 である (Provansik, 2021)。図表 2 の中の「①評価の枠組み」「②調査問題の作成」「③調査問題のパイロットテスト」「④予備調査 (データ収集)」「⑤スコア結果」「⑥アイテムの統計処理とパラメータのレビュー」「⑦最終的なアイテムプール選択」「⑧本調査 (データ収集)」「⑨スコア結果」「⑩IRT スケーリング、重み付け」「⑪分析、最終レポートの作成」「⑫データセットのリリース」という一連の流れは、PBT 開発における一連の流れと同様である。

PBT から CBT への移行によって追加されたプロセスが、図表 2 の中の [A] から [I] である。まず、コーダーが評価の枠組みに書き込まれた調査項目を HTML 等の何らかのフォーマットでデジタル処理する ([A])。次に、コーダーがアイテムやテストレットをデータ収集システムにプログラムする ([B])。また、この CBT 開発の中ではデバイスの管理が必要である ([C])。予備調査 (データ収集) の後には、ログファイルをアイテムの統計情報やパラメータとともに確認し、最終的なアイテムプール選択に役立てる ([D])。コーダーは、最終的な調査ツールとログファイルをプログラムする必要がある ([E])。そして、本調査 (データ収集) の後には、ログファイルを確認し、プロセスデータを抽出する ([F])。抽出されたプロセスデータは、レポート作成や分析に使用することができるが ([G])、データセットで公開する場合はプロセスデータを匿名化する必要がある ([H])。またさらに、プロセスデータはスケーリングに利用される ([I])。

図表2. CBT 開発の一連の流れ



出典：Provansik (2021)
日本語訳は筆者による。

CBT 開発の一連の流れの中における、プロセスデータのインプットとアウトプット、すなわちプロセスデータの生成と活用のプロセスを図表2でそれぞれ、[インプット1] から [インプット4]、[アウトプット1] から [アウトプット5] として示している。プロセスデータの最初のインプットは、「①評価の枠組み」で定義された調査内容に関する認知プロセスの理論的理解の必要性である [インプット1]。次に、アイテム測定を目的としたプロセスの定義（すなわち出題の意図）を策定する必要がある [インプット2]。つまり、調査問題の作成自体が、単なる正答／誤答の定義にとどまらず、測定されるべきすべての認知プロセスを特定し、定義することが重要である。その上で、調査で測定されるターゲットプロセスを把握するためのコーディングが必要となる [インプット3]。これは、プロセスの定義の一貫性だけでなく、プロセスデータを分析する際にどのような内容を取得しているのかを把握できるようにするために、プロセスデータについて変数名を使用するなどの標準化された方法でコーディング（及びラベリング）する。最後に、ログファイルから何を抽出すべきかを明確にし、それをサポートするためのコーディングを効率的に行うためには、プロセスデータを事前に定義する必要がある [インプット4]。プロセスデータのインプットを整理すると、①明確な評価の枠組みの設定、②その評価の枠組みに基づいて吟味された上での作問、③測りたい側面についての明確な判断基準に沿った採点、そして④収集したログファイルの中から事前に定義されたプロセスデータの抽出、この一連の過程・要素が肝要である。

一方、プロセスデータのアウトプット、すなわち収集可能なプロセスデータを検討するにはまず、調査を実施するデバイス自体の使用によって得られるパラデータが必要である [アウトプット1]。このパラデータによって、調査問題の配信に関する問題を特定するための情報を取得し、調査の整合性をモニターすることができる。次に、アイテムの信頼性を検証・確認するために、

ログファイル及びプロセスデータはフォレンジック⁶な目的のために使用される [アウトプット 2]。そして研究目的のプロセスデータ [アウトプット 3] は、プロセスデータのアウトプットとして最も注目されており、パフォーマンス（成績）指標としてのプロセスデータに関連する思考のパターンを特定するなどの活用方法が期待されている。最後に、プロセスデータの活用方法としてスコアリングプロセスに入力されるプロセスデータ [アウトプット 4] とプロセスデータ自体をパフォーマンス（成績）指標とするプロセスデータ [アウトプット 5] がある。さらに、スケーリングにプロセスデータを使用することでスコアリング入力するためのプロセスデータが得られた結果、それが項目反応理論（Item Response Theory、以下 IRT）⁷ スケーリングに影響を与える。

上述のように、ログファイル及びプロセスデータは、調査の対象となる領域を定義する評価の枠組みの開発を起点としたテスト開発から、予備調査や本調査の実施、採点、統計処理、分析・レポートの作成、最終データセットの公開、さらに調査実施後に得られたデータの分析に基づくテストの改良といった、CBT 開発の一連の流れの中で必然的に生成されるものである（図表 2 参照）。ログファイルは、プロセスデータを抽出するために意図的に開発されたものではなく、CBT 開発に伴って生成された副産物なのである（Provansik, 2021）。そして、その生成されたプロセスデータは最終的なスケーリングのために利用される。PISA のように、仮説すなわち測りたい能力（コンピテンシー）があらかじめ設定されており、またそれを検証するためのプロセスデータがある場合は、適切なプロセスを記録するための項目、例えば「アイテム測定を目的としたプロセスの定義（出題の意図）」や「ターゲットプロセスを捉えるためのコーディング（採点）」といった項目をテストに事前に組み込む必要がある。

2.2.4. ログファイル及びプロセスデータの例

ここでは、具体的に大規模国際学力調査におけるログファイルとはどのようなものかについて、以下で例を挙げて説明する。

まず、PISA、PIAAC、TIMSS などの定量的調査では、採点された解答／回答プロセス、解答／回答プロセス経路、タイミングデータ及びアクションデータ等のログファイルが収集されている。

一方 TIMSS1999 調査の理科授業ビデオ研究等の定性的調査においても、ログファイルは収集されている（Roth ら, 2006）。この場合、授業記録等（例えば、割り当てられた作業の種類（「クラス全体の作業」、「個人の作業」、「ペア／パートナーの作業」、「小グループの作業」）のカテゴリを持つ作業等々）、語数などが、コード化され記録されている（同上）。

以下では、PISA を例に具体的な、公開されているログファイル及びプロセスデータの内容について説明する。

PISA の場合、OECD の PISA ホームページから調査結果のデータベース（SPSS 形式又は SAS

6 フォレンジックのもともとの意味は「法廷の」や「法医学の」、「法的に有効な」などといった意味で、IT 業界ではコンピュータや記録媒体に含まれた法的証拠にかかわるデジタル的な法科学の一分野をデジタルフォレンジックと呼ぶ。主に用いられるのは、ハードディスクや USB メモリといったメディアの中に保存されたファイルを法的な証拠として利用するといった場面で、具体的な作業の流れとしては、まずメディアの収集を行って完全なコピーを作成し、そのコピーに対して解析や行い、その結果を報告するというまでがデジタルフォレンジックの一連の流れであるとされている（NTT コミュニケーションズ、発行年不明）。

7 IRT とは、ある項目が正しく解答／回答される確率（正答率）は、直接観察できない潜在的な特性や能力の関数であるという概念に基づいた、心理測定理論である（American Psychological Association, 発行年不明）。

形式) がダウンロードできる。公開データベースの中に含まれるプロセスデータは、「解答／回答プロセスに関するあらゆるデータとして定義することができる。一般的に、そのようなデータは、解答／回答プロセスのタイミングに関連するデータのような比較的基本的なものから、調査中の生徒の視線データ (eye-tracking data) のような比較的洗練されたものまで様々である」(OECD, 2020)。

プロセスデータは、各生徒の調査セッションのログファイルにタイムスタンプ付きのテストイベントが書き込まれることで取得される。以下の図表3は、PISA の調査配信システム (Student Delivery System、以下 SDS) で記録されるすべてのイベントを示している。PISA では、すべてのイベントやイベントの詳細がログに記録されるわけではなく、ログに記録されるイベントは、項目の種類やコンピュータ画面上の表示によって異なる。例えば、キープレスはログに記録されるが、どのキーを押したかは記録されない。そのため、反応行動の種類 (例えば、テキスト作成と編集) を区別することはできない。また、1つの画面に2つの問いが表示されることがあるため、プロセスデータと個々の問いを常に明確に結びつけることはできない。また、ログファイルが大きく、ノイズが多い場合があることにも留意する必要がある。例えば、PISA では、技術的な問題のためにテスト管理者が生徒を別のコンピュータに移動させる必要があり、SDS を完全に再起動するか、又は中断した場所に早送りして対処するケースがあった。このような場合、ログファイルを修正したり、マージしたりすることは容易ではない (同上)。

図表3. PISA の SDS に記録されたイベントログのリスト (アルファベット順)

1	breakoff (ブレイクオフ)	12	paste (ペースト)
2	change (変更)	13	PSReadyToScore (採点準備)
3	click (クリック)	14	PSscored (採点完了)
4	copy (コピー)	15	question Loaded (問題のロード)
5	cut (カット)	16	run_simulation (シミュレーションの実行)
6	dblclick (ダブルクリック)	17	score Now Result (得点 (現在の結果))
7	drop (ドロップ)	18	scoring (採点中)
8	keypress (キープレス)	19	selectionevent (セレクションイベント)
9	move (移動)	20	selectWord (ワードの選択)
10	onItemBegin (問いの解答開始)	21	stimulusAndQuestionLoaded (課題文と問題のロード)
11	onItemEnd (問いの解答終了)	22	stimulusLoaded (課題文のロード)

出典：OECD (2020), Table1. 括弧内の日本語訳は筆者による。

2.3. ログファイル及びプロセスデータの利活用に関する分析・研究の事例

ここでは、「会合・研修」での分析・研究の事例の報告と PISA のプロセスデータを利用した分析・研究の事例について整理する。

2.3.1. 「会合・研修」で報告された分析・研究の事例

「会合・研修」で行われたプロセスデータを用いた分析・研究の報告の内容は、①受検者の行動や認知戦略の理解に関する分析・研究と②“操作時間”を変数として使用した分析・研究の2つに大別される。

① 受検者の行動や認知戦略の理解に関する分析・研究では、以下のような報告が行われた。

- ・生徒がどのように考え、どのように行動するかについて具体的な情報を用いて、指導と学習を改善するための研究 (Saldivia, 2019 ; Frank, 2019 など)。
- ・高成績者 (ハイパフォーマー) と低成績者 (ローパフォーマー)、又は初心者とエキスパートの戦略を区別する要因をより理解するための研究 (Greiff, 2019 ; He, 2019 ; Santos, 2019 など)。
- ・動機づけと成績との関係をより理解するための研究 (Vergolini, 2019 など)。

② また、“操作時間”を変数として使用した分析・研究では、以下のような内容を明らかにした研究の報告が行われた。

- ・努力尺度－時間データから作成された代理指標について (Vergolini, 2019 など)。
- ・作業時間 (操作時間) は問題項目から一般化して抽象化することができるが、正答に至るまでの時間と誤答に至るまでの時間との間に差がないこと (He, 2019)。
- ・時間に基づいて成功と失敗を区別することはできないこと (Santos, 2019)。
- ・反応時間のデータが、①生徒の習熟度／スキル、②テスト領域、③テスト (使用) 言語、④問題項目の認知的要求 (例：熟練度の低い受検者は同じ速度で読み進めるが、熟練度の高い読者はより難易度の高い問題 (項目) について読むのにより時間がかかる) に作用すること (Yamamoto, 2019)。

2.3.2. PISA のプロセスデータを利用した分析・研究の事例

PISA では、2015 年調査以降 CBT に移行し、プロセスデータが利用可能になったことで、様々なトピックについての研究が可能になった。PISA のプロセスデータを用いた調査・研究のトピックとしては、以下のようなものが挙げられる。

- ・欠落データの処理 (Lu & Wang, 2020 ; Ulitzsch ら, 2019)。
- ・方略 (Greiff & Avvisati, 2015 ; He ら, 2018)。
- ・生徒の努力とエンゲージメント (Anaya & Zamarro, 2020 ; Michaelides ら, 2020 ; Ulitzsch ら, 2019)。
- ・応答時間のスケーリング (Shin ら, 2020)。
- ・生徒の探究心のプロファイル (Teig ら, 2020)、等。
- ・しかし、上記のトピックのような現状行われているほとんどの研究では、問題 (項目) の比較的小さなサブセットについて実施されている (von Davier ら, 2019)。

3. ログファイル及びプロセスデータの利活用に関する諸課題

「会合・研修」では、大規模国際学力調査におけるログファイル及びプロセスデータの利活用に関する具体的な研究・分析の報告とともに、ログファイル及びプロセスデータを利活用する上での課題や問題についても提示、議論が行われた。以下では、その諸課題を①方法論的・理論的課題、②不適切な利活用に関する課題、そして③倫理的課題の3つの側面に分けて整理する。

3.1. 方法論的・理論的課題

「会合・研修」では、ログファイル及びプロセスデータを活用した分析を行う上で、ログファイル及びプロセスデータをどのように使用するかに関する方法論的・理論的枠組み及びその方

向性を明確にすることが重要であり、さらにそのためにはそもそもログファイル及びプロセスデータを何とみるか、境界線はどこかといった定義についても明確にする必要があるということが指摘された。具体的には、方法論的・理論的課題として以下のようなことが示された（Provansik, 2019b）。

- ・調査受検者の努力度合い（努力指標）を図るための変数として、各項目への応答時間や合計時間などの“操作時間”を用いることは妥当であるかどうか、十分に検証する必要がある。
- ・プロセスデータを利活用するためには、事前にデータを整理したり新たな変数を作成したりする等の複雑な作業を行う必要がある。したがって、ログファイルをプロセスデータに変換するためのガイドラインや企画の策定が必要である。
- ・ログファイル及びプロセスデータ分析を行った上で、更にそれを定性的なデータで補完することも必要である。（例：調査受検者に対して、どのように考えてテストへの解答を行ったかをインタビューするなど）

“情報の生成 (Information generation)” (正しい結論にたどりつくために正しい手順を踏む) と “情報の解釈 (Information interpretation)” (結論を正答とみなす) を区別することが重要である。

- ・プロセスデータと研究の質を向上させるために、CBT テスト開発と分析との間に相互作用的なフィードバックループが不可欠である。

3.2. 不適切な利活用に関する課題

3.1. で上述したように、「会合・研修」では、ログファイル及びプロセスデータの活用にあたっては、方法論的・理論的課題があることが指摘されており、ログファイル及びプロセスデータの使用にあたってのガイドラインや規格の策定についてはまだ確立されておらず現在議論の最中であるが、ログファイル及びプロセスデータの活用にあたっては、最低限、以下の3つの事項については避けるべきであるとの見解が示された（Provansik, 2019b）。

- ・分析・研究の結果明らかとなった、高成績に関する方略（行動パターン等）を全生徒に教えるべき方略であると結論づけること。
- ・授業や診断的な評価を行うことによって、生徒が新しい考え方ややり方を試すための構造化されていない機会を失うような方法でプロセスデータを利活用すること。
- ・ログファイル及びプロセスデータ分析の結果を、単純に一般化すること。また、1つの項目から全ての項目へ、又は1つのプロセスから複数のプロセスへの過度に一般化すること。特に、PISA 等の大規模国際学力調査では、分野、測定する側面、国、言語等々によって全く異なるため、ログファイル及びプロセスデータ分析結果による乱暴な議論は避けるべきである。

3.3. 倫理的課題

「会合・研修」の Murchan の報告（Murchan & Siddiq, 2019）では、CBT は効率性や柔軟性、採点の一貫性があること、そして受検者のエンゲージメントをモデル化できる可能性があるという利点がある一方で、ログファイル及びプロセスデータの使用にあたっては、以下のような倫理的な課題があるということが指摘された。

- ・生徒の解答／回答を倫理的に問題のある方法で使うこと（例えば、生徒個人が特定されるような方法で情報を公開すること）及び解答／回答から生徒を識別するような目的（例えば、

デジタルペンによる生徒の筆跡を再現し、それを公開する等) のために、ログファイル及びプロセスデータを利用すべきではない。

- ・調査の受検者が、調査を受検すること自体には同意していたとしても、プロセスデータがその後の分析と解釈に使用されることに対しては同意しているとは限らない。

4. おわりに

本稿のまとめとして、大規模国際学力調査におけるログファイル及びプロセスデータの利活用に関する現状を整理する。CBT で実施される調査においては、ログファイルデータは必然的に生成されるものである。そして、その中で何をプロセスデータとするかは、そのデータの利活用の目的に沿って事前に定義することが必要である。すなわち、ただ闇雲にプロセスデータを活用するのではなく、プロセスデータを利活用する上での諸課題や問題点等に留意した上で、慎重に検討してゆく必要がある。また、プロセスデータを用いた分析・研究の結果は、単純に一般化できるものではなく、結果を受けた乱暴な議論は避けるべきである。また、プロセスデータの利活用の際には倫理的問題にも留意する必要がある。

以上の「会合・研修」で報告・検討された内容を踏まえると、大規模国際学力調査におけるログファイル及びプロセスデータの利活用に関する研究分野は国際的にもまだ黎明期にあり、今正に分析・研究が蓄積されている段階であると言えよう。今後も引き続きこの分野における分析・研究の動向を注視していきたい。

参考文献

- American Psychological Association. (発行年不明). item response theory (IRT). *APA Dictionary of Psychology*.
<<https://dictionary.apa.org/item-response-theory>> (2021年4月28日アクセス).
- Anaya, L., & Zamarro, G. (2020). *The role of student effort on performance in PISA: Revisiting the gender gap in achievement*. Education Reform Faculty and Graduate Students Publications.
<<https://scholarworks.uark.edu/edrepub/116>> (2021年1月15日アクセス).
- Goldhammer, F., Hahnel, C., Kroehne, U., & Zehner, F. (2019) Some considerations in validating the interpretation of process indicators, Opportunity versus Challenge: Exploring Usage of Log-File and Process Data in International Large Scale Assessments (会合資料：セッション4).
- Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers & Education*, 91, pp.92-105.
- Greiff, S. (2019) Log File Analyses, Transversal Skills & Learning Analytics: A good Match, But Not a Perfect One, Opportunity versus Challenge: Exploring Usage of Log-File and Process Data in International Large Scale Assessments (会合資料：セッション2、発表1).
- He, Q., von Davier, M., & Han, Z. (2018). Exploring process data in problem-solving items in computer-based large-scale assessments. In H. Jiao, R. W. Lissitz, & A. Van Wie (Eds.), *Data analytics and psychometrics: Informing assessment practices* (pp. 53-76). Information Age Publishing.
- He, Q. (2019) Exploring Sequence-Based Approaches Using Process Data in Large-Scale Assessments, Opportunity versus Challenge: Exploring Usage of Log-File and Process Data in International Large Scale Assessments (会合資料：セッション2、発表2).

JPNIC (2017) 「VPN とは」『ニュースレター No.67』

<<https://www.nic.ad.jp/ja/newsletter/No67/0800.html>> (2021年1月15日アクセス).

国立教育政策研究所編 (2013) 『成人スキルの国際比較－ OECD 国際成人力調査 (PIAAC) 報告書』 明石書店.

国立教育政策研究所 (2019) 『生きるための知識と技能 7－ OECD 生徒の学習到達度調査 (PISA) 2018 年調査 国際結果報告書－』 明石書店.

国立教育政策研究所 (2020) 『国際数学・理科教育動向調査 (TIMSS2019) 国内報告書』.

小柳和喜雄 (2019) 「学習ログを活用した個別最適化学習の評価に関する試行研究」『日本教育学会大会研究発表要項』 78, pp.119-120.

久保田真一郎 (2015) 「LMS における学習ログと学習スタイルとの関係性の検討」『情報処理学会研究報告』 pp.1-6.

李凱 (2018) 「教育ビッグデータを用いた学習活動及び質問難易度の可視化に関する研究」『情報学研究』 pp.78-82.

Lu, J., & Wang, C. (2020). A response time process model for not-reached and omitted items. *Journal of Educational Measurement*, 57, pp.584-620

Michaelides, M. P., Ivanova, M., & Nicolaou, C. (2020). The relationship between response-time effort and accuracy in PISA science multiple choice items. *International Journal of Testing*, 20 (3), pp.1-19.

文部科学省 (2019) 「【資料 3-2】 GIGA スクール構想の実現」

<https://www.mext.go.jp/content/20191219-mxt_syoto01_000003363_11.pdf> (2021年1月15日アクセス).

Murchan, D., & Siddiq, F. (2019) Ethical considerations involving data analytics in educational assessment: A systematic literature review, Opportunity versus Challenge: Exploring Usage of Log-File and Process Data in International Large Scale Assessments (会合資料：セッション 3).

NTT コミュニケーションズ (発行年不明) 「フォレンジックとは」『ICT Business Online』

<<https://www.ntt.com/bizon/glossary/j-h/forensic.html>> (2021年1月15日アクセス).

野田健夫・金子真隆 (2020) 「Moodle 上での数学学習に関するログデータが物語ること」『日本科学教育学会第 44 回年会論文集』 pp.213-216.

OECD 編著、国立教育政策研究所監訳 (2016) 『21 世紀の ICT 学習環境 生徒・コンピュータ・学習を結び付ける』 明石書店.

OECD (2020) Annex K: Uses and Reporting of Process Data, *PISA 2018 Technical report*.

<<http://www.oecd.org/pisa/data/pisa2018technicalreport/PISA2018-TechReport-Annex-K.pdf>> (2021年1月15日アクセス).

緒方広明・殷成久・毛利考佑・大井京・島田敬士・大久保文哉・山田政寛・小島健太郎 (2016a) 「教育ビッグデータの利活用に向けた学習ログの蓄積と分析」『教育システム情報学会誌』 33 (2), pp.58-66.

緒方広明・殷成久・大井京・大久保文哉・島田敬士・小島健太郎・山田政寛 (2016b) 「デジタル教材の閲覧ログを利用したアクティブラーナーの学習行動の分析」『基幹教育紀要誌』 2, pp.48-60.

Provansik, S. (2019a) Process Data- the New Frontier for Assessment Development: Rich New Soil or a Quixotonic Quest?, Opportunity versus Challenge: Exploring Usage of Log-File and Process Data in International Large Scale Assessments (会合資料：基調講演).

Provansnik, S. (2019b) Closing Remarks, Opportunity versus Challenge: Exploring Usage of Log-File and Process Data in International Large Scale Assessments (会合資料：閉会挨拶).

Provansnik, S. (2021) Process data, the new frontier for assessment development: rich new soil or a quixotic quest? *Large-*

scale Assessments in Education, 9, pp.1-17.

- Roth, K. J., Druker, S. L., Garnier, H., Lemmens, M., Chen, C., Kawanaka, T., Okamoto, Y., Rasmussen, D., Trubacova, S., Warvi, D., Gonzales, P., Stigler, J., and Gallimore, R. (2006). *Teaching Science in Five Countries: Results from the TIMSS 1999 Video Study*. (NCES 2006-011). U.S. Department of Education. Washington, DC: National Center for Education Statistics. <<http://nces.ed.gov/>> (2021年1月15日アクセス).
- Saldivia, L. (2019) Using national assessment process data to identify construct relevant variables and infer cognitive processes, Opportunity versus Challenge: Exploring Usage of Log-File and Process Data in International Large Scale Assessments (会合資料：セッション5、発表1).
- Salmeron, L. (2019) Selection of irrelevant navigation paths during digital reading: Underlying cognitive processes, Opportunity versus Challenge: Exploring Usage of Log-File and Process Data in International Large Scale Assessments (会合資料：セッション5、発表2).
- Santos, R. D. (2019) Assessment in the age of Data Science: the case of interactive items tested in France, Opportunity versus Challenge: Exploring Usage of Log-File and Process Data in International Large Scale Assessments (会合資料：セッション6).
- Shin, H. J., Kerzabi, E., Joo, S. H., Robin, F., & Yamamoto, K. (2020). Comparability of response time scales in PISA. *Psychological Test and Assessment Modeling*, 62(1), 107-135.
- 田口真奈・後藤崇志・毛利隆夫・飯吉透 (2017) 「<センター教員・共同研究論考> MOOCの行動ログと質問紙を組み合わせたデータ分析の設計」『京都大学高等教育研究』23, pp.109-118.
- Teig, N., Scherer, R., & Kjærnsli, M. (2020). Identifying patterns of students' performance on simulated inquiry tasks using PISA 2015 log-file data. *Journal of Research in Science Teaching*, 57, pp.1400-1429.
- Ulitzsch, E., von Davier, M., & Pohl, S. (2019). A hierarchical latent response model for inferences about examinee engagement in terms of guessing and item-level non-response. *British Journal of Mathematical and Statistical Psychology*, 73, pp.83-112.
- Vergolini, L. (2019) Effort and achievement of 15-year-olds in PISA 2015 across EU member states, Opportunity versus Challenge: Exploring Usage of Log-File and Process Data in International Large Scale Assessments (会合資料：セッション1、発表2).
- von Davier, M., Khorramdel, L., He, Q., Shin, H. J., & Chen, H. (2019). Developments in psychometric population models for technology-based large-scale assessments: an overview of challenges and opportunities. *Journal of Educational and Behavioral Statistics*, 44(6), pp.671-705.
- 山田恒夫 (2015) 「MOOCと学習解析：教育革新のための情報基盤に向けて」『情報処理学会論文誌教育とコンピュータ』11(4), pp.1-11.
- Yamamoto, K. (2019) Using Timing Information Associated with Response Data in Large Scale Assessment, Opportunity versus Challenge: Exploring Usage of Log-File and Process Data in International Large Scale Assessments (会合資料：セッション1、発表1).

(受理日：令和3年3月10日)