

全国学力・学習状況調査結果の長期的分析等に係る
調査研究の請負

報告書 ポイント版

2026年3月

三菱UFJリサーチ&コンサルティング

世界が進むチカラになる。



本調査研究実施の背景と目的・実施内容

背景（現状）

- 全国学力・学習状況調査（以下「全国学調」という。）データについて
 - 文部科学省が保有する、大規模かつ長期にわたる貴重なデータである。
 - 一方、経年での利用を前提とした調査設計でなかったことから、**長期的な分析に当たっては、調査項目の整理やデータクレンジング※の上、データを経年的に接続する必要がある。**

目的

- 以下のノウハウ及び留意点を蓄積し、これらの知見を政府全体に共有することで各府省等におけるEBPMの推進につなげること。
 - 各府省等が自らの業務においてデータを正確に扱うために求められる**データクレンジングに関するポイント**
 - 今回のようなビッグデータを長期的に分析する際に必要となる、**個票データを経年的に分析する手法**など

実施内容

- 文部科学省等と共同で、全国学調データという大規模かつ長期にわたるデータを学校・市町村単位で経年的に分析できるように整理する。
- 当該データを用いて経年分析を行うとともに特徴的な傾向を示す学校・市町村等を把握する。

※本調査研究におけるデータクレンジングとは、データ内の誤りや欠損、重複、形式の不統一等を修正することで、異なる時点のデータの正確な接続を可能にし、データ分析、データの活用等に適した形式に整理する作業を指す。

本調査研究の全体像

- 本調査研究では、以下のとおり**全国学調データのクレンジング及び接続を行うとともに、得られたノウハウを「手引き」として整理した。**
また、経年データ作成後、**経年的な実態把握や取組等の効果分析を行った。**

【使用するデータ】 全国学調悉皆調査個票データ（H19年度～R6年度の計16回分） ※1

- ・ 対象：国・公・私立学校の小学校第6学年（約2万校、約100万人）、中学校第3学年（約1万校、約100万人）の原則として全児童生徒
- ・ 内容：「教科に関する調査」、「質問調査（児童生徒に対する調査、学校に対する調査）」の結果

全国学調データのクレンジング・接続

1 調査項目をつなげるためのデータクレンジング

調査項目の年度間の変更を一覧で整理し、変数IDによって各年度の調査項目を対応させる。

2 調査対象をつなげるためのデータクレンジング

学校等の統廃合や各年度のデータの不整合に留意し、各年度の接続キーを統一し、経年でデータを接続する。

「データクレンジング及び経年データ作成に関する手引き」の作成

全国学調データのクレンジング及び接続の結果得られたノウハウを他のデータにも活用できるように整理した。

全国学調データの分析

1 基礎集計

基本統計量の算出により分析に用いる変数の特徴を把握する。標準化スコアの属性（学校種別等）別の推移を可視化する。

2 経年変化の分析（固定効果モデルを用いた分析）

質問調査項目と学力等のアウトカムの関連を探索的に分析し、アウトカムに対して有効な取組を把握する。

3 特徴的な傾向を示す学校・市町村の把握

経年データを用いて、SES（社会経済的地位） ※2が低い環境にありながら、高い成果を挙げている学校・市町村を抽出する。

4 外部データの接続の検討

全国学調データと接続する意義があると考えられる外部データや分析方針を検討する。

※1 貸与データの全体像については文部科学省HPを参照（https://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/sonota/1386492.htm）

※2 SES（社会経済的地位）とは、社会的・経済的な位置づけを表す指標であり、子どもの場合、親の所得や学歴、職業、教育環境などから測定されることが多い。

全国学調データのクレンジング・接続のプロセスと得られた示唆

基本的なクレンジング

調査項目をつなげるためのデータクレンジング

調査対象をつなげるためのデータクレンジング

1

データクレンジング結果イメージの検討

- データの構造を把握
- データ容量の大きさ及び分析内容を踏まえ具体的に作成するデータを検討

※本調査研究においては、基本的なクレンジングを実施済みの単年度ごとのデータを受領

2

調査項目の接続方法を検討

- 年度間で設問文や選択肢が異なる調査項目の接続方針を検討
- 年度間で異なる質問番号を対応させる調査項目マスタを作成

3

調査項目に変数IDを採番

- 調査項目ごとに変数IDを設定しデータセットに割当て
- 調査対象による調査項目の違いや、判断・変更箇所の追跡可能性確保に留意

4

接続キーの設定

- 全年度データに接続キーとして学校コードを割当て
- 学校名や学校コードに誤りが含まれているデータを修正

5

経年データの作成

- ロング型データセットとして経年データを作成
- 接続キーが正しく紐づいているかチェック
- 基本統計量やデータ数の確認も実施

6

データクレンジングのプロセスの記録

- データクレンジング時のプロセスを記録し、年度ごとの調査項目を追跡できるようにリストを作成

プロセスの概要

押さえておくべきポイント

- 分析やデータの活用を見据えて結果イメージを持っておくことがデータクレンジングの省力化につながる。

- 人の目で確認するための調査項目マスタを作成することで、全体像がわかりやすくなる。

- 今後新たにデータを追加接続する場合を見据えて、長期的な目線での変数IDの採番が必要となる。

- 他の調査等でも共通して使われる接続キーを設定することで、外部データとの接続が容易になる。

- ただデータを接続するのではなく、エラーがないか複数人で確認することが重要である。

- 記録を残しておくことで、後からの追跡や修正が容易になる。

ボトルネックになりやすいポイント

- 実際のデータを処理する過程で初めて把握できる課題もあるため、作業の進捗に応じてゴールイメージを適宜見直し、精緻化を要する場合がある。

- 調査時点間で調査項目の設問文や選択肢等の変更がある場合、マスタ作成の過程でその差異をどのように整理・統合するかの方針を検討する必要が生じる。

- 調査時点間でデータを正しく接続できるように、データを確認することが必要であり、データに含まれる調査項目の数が多い場合、確認に時間を要する場合がある。

- 調査時点間で接続キーが統一されていない場合や、接続したい他の調査等と同一のコードが使用されていない場合、一貫した接続キーの設定に著しくコストがかかる場合がある。

- データに含まれるエラーに気づかずに分析をした場合、想定される含意が得られない可能性がある。

- データが大きい場合、修正が必要な点が多く、記録に抜け漏れが発生しやすい。また、データの追加の際、記録が後回しになり抜け漏れが発生しやすい。

この結果得られたノウハウを他のデータにも活用できるよう「データクレンジング及び経年データ作成に関する手引き」に整理

全国学調データの分析概要とその結果

- クレンジング及び経年接続した全国学調データを用いた分析・検討概要とその結果は以下のとおりである。
- 本調査研究で用いるH19年度～R6年度のデータは、経年での比較を前提とした調査設計とはなっておらず、正答率による経年比較は適当ではないため、分析の際には、正答率の代替として正答率を標準化したスコア（標準化スコア）を用いることで集団内での相対的な位置をアウトカムとした。

	概要	結果	解釈・示唆
基礎集計	<ul style="list-style-type: none"> ■ 基本統計量の算出、標準化スコアの推移の可視化によってデータの特徴を概観する。 	<ul style="list-style-type: none"> ■ 経年で属性間における順の入れ替わりは見られなかった。 	<ul style="list-style-type: none"> ■ 経年で、属性ごとの標準化スコアの推移の差が安定的であり、国内における属性による構造的要因が経年で変化していないことが示唆される。
経年変化の分析 (固定効果モデルを用いた分析)	<ul style="list-style-type: none"> ■ 固定効果モデルを用いて学校単位に経年接続したデータを分析し、教育の時系列の変化、教育の効率性、教育格差是正を検証する。 	<ul style="list-style-type: none"> ■ コロナ禍前後でSESと標準化スコアの関係に一貫した明確な傾向は見られなかった。 ■ 学校内の取組のうち、特に「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」は標準化スコアとの間に正の関連をもつ。 ■ 学校内の学力の散らばりと取組の関係について、小学校におけるICTの使用頻度と標準偏差に負の関連が見られた。 	<ul style="list-style-type: none"> ■ 先行研究を支持しない結果であり、データの制約や影響している可能性がある。IRTスコアを用いて経年で分析することでより実態に即した結果を得ることができる可能性がある。 ■ 「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」を推進することは、学校全体の学力向上に有効な取組である可能性がある。 ■ ICTの使用頻度を高めることは、学校内学力格差の是正に有効な取組である可能性がある。
特徴的な傾向を示す学校・市町村等の把握	<ul style="list-style-type: none"> ■ 市町村単位に経年接続したデータを用いて、SES指標が一定以下の水準の市町村のうち、高い水準で成果が推移している・成果が伸びている市町村を抽出する。 ■ 抽出された市町村から、同様の特徴が見られる学校を抽出する。 	<ul style="list-style-type: none"> ■ SESが低いものの①継続して高い成果を示す市町村は全部で163市町村、②成果の伸びを示す市町村は全部で422市町村抽出した。 ■ 上記の市町村のうち、抽出対象となった市町村から、①について1,013校、②について497校の学校を抽出した。 	<ul style="list-style-type: none"> ■ SESが低い中でも高い成果を示す学校・市町村がデータで示された。 ■ ①、②で異なる市町村が多く抽出されていることから、高い成果を維持している・伸びを示している市町村はそれぞれ異なる特徴を持つことが考えられる。
外部データの接続の検討	<ul style="list-style-type: none"> ■ 全国学調データに接続する意義があると考えられる外部データや分析方針を検討する。 	<ul style="list-style-type: none"> ■ 外部データの接続に関する留意点等を整理した。 ■ 調査主体や調査目的等、実施条件が異なるデータを接続する際は、誤りが生じないよう留意が必要である。 	<ul style="list-style-type: none"> ■ 各調査で取得できるデータや項目には限りがあるが、外部データと接続することによって、各調査項目を活用でき、より有効な分析が可能になると考えられる。