

因果推論により抽出された リアルワールド教育エビデンスの 再現性評価とフィードバック設計

奥村 光貴

京都大学大学院情報学研究科

研究背景：RWEと再現性の課題

リアルワールドエビデンス (RWE)

実際の教育現場で収集されたデータから得られるエビデンス。RCT（ランダム化比較試験）を補完し、現場への適用可能性が高い。

因果推論の適用

観測データから介入なしに因果関係を特定できる強力な手法。しかし、抽出された因果関係の**再現性**については検証が不足している。

再現性の危機 (Replication Crisis)

心理学研究の再現率はわずか36%（OSC, 2015）。教育研究の再現率は他分野よりもさらに低い傾向が指摘されている。

RWEにおける再現性の問題：

- 有意性検定は検出力に依存する
- 効果量差の「等価性」基準が不明確
- データ固有のノイズに影響されやすい

研究目的とリサーチクエスチョン

因果探索の「発見力」とCTの「検証力」を統合し、
信頼性の高いRWEの抽出と適応的フィードバックへの応用を目指す

RQ1

RWEの質保証

因果探索により抽出されたRWEに対し、Correspondence Test (CT) を適用することで、再現性を定量的に評価し、信頼性の高いRWEを抽出できるか。

RQ2

再現性あるRWEの解明

再現性のある因果関係はどのような学習行動と成績変化の間に存在し、それらは学年や年度といった文脈によってどのように変化するのか。

RQ3

実践への応用

再現性あるRWEに基づく解釈から、学習者の内部状態を推論し、その状態に応じた適応的なフィードバック介入の設計指針を提案できるか。

方法(1) : データセットと特徴量生成

分析データセット

中学1-3年生の夏休み学習ログ (6データセット)

学年	2023年度 (n)	2024年度 (n)
中学1年 (J1)	117	112
中学2年 (J2)	109	115
中学3年 (J3)	109	114

目的変数 : 夏休み前後の試験偏差値差 (deviation_diff)

111個の網羅的特徴量

基本的学習指標 (11個)

jump_count , yellow_marker_count , red_marker_count 等

高度な学習指標 (100個)

max_gap_days (最大学習間隔)

wandering_time (迷走時間)

learning_velocity (学習速度) 等

ポイント : 理論に基づく絞り込みを行わず、全特徴量を因果探索の入力とすることで予期せぬ知見の発見を目指す。

方法(2) : 因果探索 (LiNGAM-MMI)

🔍 採用手法

LiNGAM-MMI

- 🛡️ **未観測交絡への耐性:**
学習意欲や家庭環境など、データに含まれない要因の影響を考慮しつつ因果方向を特定。
- ⚡ **高次元データへの対応:**
111個の網羅的な特徴量に対しても、効率的かつロバストに因果構造を推定可能。
- 🔄 **非ガウス性の活用:**
データの分布の歪みを利用して、因果の「向き」を一意に決定。

⚙️ 因果探索プロセス

- 1 特徴量の生成 (111個)
- 2 因果グラフの推定 (LiNGAM-MMI)
- 3 因果ペアの抽出 (直接・間接影響)

※特定の仮説に基づく絞り込みを行わず、網羅的に探索することで予期せぬ因果関係の発見を目指す。

方法(3) : 再現性評価 (Correspondence Test)

DT (Difference Test)

$$H_0: \beta_{\text{diff}} = 0$$

統計的有意性の検定。差があるかを判定。

ET (Equivalence Test)

$$H_0: |\beta_{\text{diff}}| \geq \delta$$

実質的等価性の検定。差が許容範囲内かを判定。

再現性の判定基準 (Steiner & Wong, 2018)

判定区分	DT (有意差)	ET (等価性)	定義と解釈
Equivalent	非有意	有意	再現性あり。統計的な差がなく、かつ差が許容範囲内。
Trivial Difference	有意	有意	統計的な差はあるが、実質的には無視できる（再現成功とみなす）。
Indeterminate	非有意	非有意	判定不能。差があるともないとも言えない（検出力不足など）。
Different	有意	非有意	再現性なし。統計的な差があり、かつ差が許容範囲外。

閾値 δ (Equivalence Margin) の設定

$\delta = 0.2$ (厳格) : 厳密に再現される基礎的な法則を評価

$\delta = 1.0$ (寛容) : 文脈による変動を含む応用的な法則を評価

結果(1) : 全体分析 (普遍的法則)

全学年共通の最重要要因

最大学習間隔 (max_gap_days)

$\beta = -0.16$

Definition

最大学習間隔日数

夏休み期間中において、学習セッションが空いた日数の最大値。学習の「継続性」を測る指標。

Data Interpretation

学習間隔が1標準偏差 (約7.6日) 増加するごとに、偏差値が約1.4ポイント減少する傾向がある。

Insight

学習間隔の拡大は「学習リズムの断絶」を意味する。記憶の減衰や習慣の喪失が成績低下の直接的な要因となっている。

Key Conclusion: 「学習リズムの維持」は偏差値向上のための普遍的な必要条件であり、最も優先的に介入すべき対象である。

結果(2) : 学年別分析（発達段階による転換）

中学初期の「量・計画性」から、中学後期の「質・認知戦略」へと成績向上の鍵が移行する。

中学1年 (Grade J1)

summer_time_concentration

負の影響

Definition

取り組みのムラの大きさ（日別イベント数の標準偏差）

Data Interpretation

ムラが1標準偏差(約1907単位)増えると、偏差値が 約160点 減少。

Insight

学習ペースの不安定さと計画性の欠如が成績低下に直結する。

learning_velocity

正の影響

Data Interpretation

学習速度が1標準偏差(約3.9単位)増えると、偏差値が 約4.1点 向上。

Insight

J1段階では、新しい知識に対する「効率的探索」が成功要因となる。

その他、多動・計画性の欠如が負の影響

(J1 : events_per_day、daily_events_grad、J2 : action_density_grad)

中学3年 (Grade J3)

yellow_marker_count

正の影響

Definition

黄色マーカーの使用回数（能動的な情報の選別）

Data Interpretation

使用回数が1標準偏差(約10回)増えると、偏差値が 約3.8点 向上。

Insight

情報の「能動的選択」という高度な認知戦略が有効に機能している。

jump_count

負の影響

Definition

ページ遷移の合計回数（情報の断片的な参照）

Data Interpretation

回数が1標準偏差(約53回)増えると、偏差値が 約2.3点 減少。

Insight

体系的な理解を伴わない「拾い読み」が深い理解を妨げている。

発達段階の転換：中学初期は「学習リズムの土作り」が、中学後期は「情報の取捨選択と構造化」が成績を左右する主要因となる。

結果(3) : 年度別・縦断的分析 (行動の多義性)

行動のリテラルな意味に囚われず、文脈に応じた「内部状態」の推論が必要である。

🕒 年度別分析 (2024年度)

wandering_time

Definition

迷走時間 : 1分以内に同じページに戻った時間の合計。

Data Interpretation

迷走時間が1標準偏差(**約2110秒**)増えると、偏差値が **約9.2ポイント** 向上。

Insight

困難な問題への「構成的試行錯誤」や「深い処理」というポジティブな没頭状態を示唆している。

👥 縦断的分析 (J2→J3)

red_marker_count

Definition

赤色マーカー回数 : 重要箇所の強調操作。

Data Interpretation

使用回数が1標準偏差(**約6.3回**)増えると、偏差値が **約2.7ポイント** 減少。

Insight

暗記作業への偏重が「わかった気 (Illusion of Competence) 」を生じさせ、深い定着を阻害している。

行動の多義性 : 「迷走」が深い思考を、「強調 (赤マーカー)」が浅い処理を意味する場合がある。再現性あるエビデンスに基づき、これらの行動の背後にある「内部状態」を正しく見極めることが不可欠である。

発達段階に応じた内部状態の転換モデル

中学初期 (J1/J2)

📌 基礎的自己調整学習

学習量・頻度の一貫性や計画性が成績を左右する。

`max_gap_daysummer_time_concentration`

🕒 計画的実行

終盤の駆け込みを避け、安定したペースを維持する能力。

`action_density_grad`



中学後期 (J3)

🧠 情報処理の質

能動的な情報の選別と構造化が鍵となる。

`yellow_marker_countjump_count` (負)

⚠️ 高度な認知的課題

単純な暗記作業による「わかった気」の抑制が重要。

`red_marker_count` (負)

行動の多義性：「迷走時間 (`wandering_time`)」は、リテラルな意味とは異なり、困難な課題への構成的試行錯誤というポジティブな内部状態を示唆する。

。

適応的フィードバック戦略の設計指針

対象状態 / 指標	介入の論理	フィードバック例
学習リズムの維持 max_gap_days	普遍的な基礎法則に基づき、習慣形成と継続をポジティブに強化する。	「これで学習連続日数が9日目ですね！この調子で次も達成しましょう。」
計画性の欠如 summer_time_concentration	中学初期の不安定なペースに対し、学習開始前に計画フェーズを促す。	「今回の学習セッションで、何をどこまで達成したいですか？」
浅い閲覧 / 多動 events_per_day (→learning_velocity)	非生産的な多動を抑制し、成績に正の影響のある「効率的探索」へ誘導する。	「理解すべきポイントは見つかりましたか？次のページに進んでみましょう。」
わかった気 red_marker_count	単純な強調行動による阻害を防ぐため、想起努力を課して処理を深める。	「定着度を確認するため、今マークした内容を要約してみましょう。深い理解を試みることで、この段階で陥りやすい成績の低下（約2.7ポイント分）を防ぎ、知識の定着を確実にすることができます。」
構成的試行錯誤 wandering_time	困難な問題への没頭を肯定し、内発的動機づけを維持・強化する。	「難しい問題に粘り強く取り組んでいますね。その粘り強さが成果につながります！」

結論と今後の展望

本研究の貢献

RWEの質保証フレームワーク

因果探索とCTを統合し、教育データの再現性を定量的に評価する手法を確立した。

多義的な学習行動の解明

「迷走」や「マーカ―」など、文脈によって意味が変化する行動の背後にある内部状態を特定した。

今後の展望

リアルタイム介入への実装

提案したフィードバック戦略を学習支援システムに実装し、その教育的効果を検証する。

分析対象の拡大

他教科や異なる学習環境（対面授業等）への適用を通じ、エビデンスの汎用性を高める。

データ駆動型エビデンスの信頼性を担保し、
個別最適化された教育実践の実現に貢献する。