適応型学習機能を組み込んだ AI ドリルの「比・割合」のビッグデータ解析から見えてくること
Big Data Analysis of Ratio in The Adaptive Learning Drills

益川 弘如†，稲垣 忠‡
Hiroyuki Masukawa, Tadashi Inagaki
†聖心女子大学，‡東北学院大学
University of the Sacred Heart, Tokyo, Tohoku Gakuin University
masukawa@u-sacred-heart.ac.jp

概要
小中学校で一人一台情報端末の整備と共に，AI ドリルを導入する自治体が増えている．AI ドリルで子供たちはいかに学んでいるのか. 小学校 5 年生 54 名が 2020 年 5 月から 2021 年 3 月の間に取り組んだログ記録データ 111,161 件から「比・割合」に関連する「小数のわり算」単元分を抽出，分析した．結果，正答が出せることに主眼が置かれており，意味理解をしっかりしたものにすることは支援対象外だった．認知科学の研究成果を取り入れ，改善サイクルを回すことが望まれる．

キーワード : AI ドリル, 意味理解, ビッグデータ解析

1. はじめに

文部科学省の GIGA スクール構想により，2020 年度末に小中学校に一人一台情報端末が整備された．それに合わせて市販ドリルアプリを導入する自治体が増えている．ドリルアプリの利用によって，子供たちのいかなる力が育まれているのか. 市販ドリルアプリには，従来の紙ベースのものを電子化し配布採点業務を自動化したものと，適応型学習機能を組み込んだいわゆる AI ドリルと呼ばれているものがある．今回は，AI ドリルに焦点をあて，小学校「比・割合」の領域を対象として，いかなる学びが起きていたかをビッグデータ解析を通して推定する．そして分析結果を基に，今後の改善の余地の方向を示す.

2. 背景と目的

稲垣ら[1]は，一人一台情報端末とクラウド環境の整備を見通し，小学校の算数科において，学校及び家庭の双方での AI ドリル活用における実践結果を分析している．宮城県内の公立 T 小学校で，2016 年 6 月から 2017 年 3 月まで，実証研究として一人一台情報端末を貸与し，3年生から6年生までの児童 61 名の学習履歴を分析した．この研究での分析視点は，(1)学校と家庭での利用状況，(2)単元テストの正答率と学習履歴との関連性，であり，(1)学校で共通の問題に取り組み家庭では習熟度に応じた取り組みをしていること，(2)習熟度に応じて取り組む問題の正答率と単元テストの正答率との間に関連があることが示されている．また，単元テスト正答率下位群の児童は，繰り返し取り組む児童と，そうでない児童に分かれ，後者の児童のスコアは低下し，AI ドリル単体では支援しきれていない側面が見えた．これらの結果より，学校で標準レベルの問題を中心に取り組み，家庭で個に応じたレベルの問題に取り組めば，成果が得られるとしている．

しかしこの研究での「成果」とは，単元テストの点数との関連である．認知科学の立場から見ると，ドリル活用の「成果」が，「計算問題が解けるか」「文章題が解けるか」だけでなく，「分からない手続きが分かるようになるか」「概念的理解が深まるか」「多様な文脈の問題に活用できるか」といったプロセスが気になる．

認知科学の研究領域では，古くより Cognitive Tutor をはじめとした知的 CAI の開発･検証が行われてきた．例えば Anderson ら[2]の，理解が難しいとされる幾何学の学習を対象とした Geometry Tutor の開発と実践がある．学校場面への導入により，教師と生徒双方に効果があったとされ，特に教師の役割が「教える」ことから，生徒たちの「なぜそうなのか」の個別質問に答えて回るようになり，より概念的理解に対する支援が可能になったという．Bruer は著書 Schools for Thought[3]で，知的 CAI 研究の教育的に最も重要なのは，コンピュータによるチューターそれ自体ではなく，Geometry Tutor を生み出した学習者モデルであるとしている．システムに組み込まれた学習者の熟達の姿を示すモデルは，カリキュラムの構成や新しい形式の開発，テストに対して重要な意味を持つとした.「私たち研究者が教育者に提示しなければならないのは，こうしたモデルなのである」，とまとめている．

これらを踏まえ本研究では，稲垣ら[1]のデータ解析結果による「AI ドリルに取り組むことには効果がある」という実証の先として，具体的な学習プロセス「AI ド

リルでどういう学びが起きていそうか」に焦点をあて分析を行う．

3. 「比・割合」の領域について

今回，AI ドリルでどういう学びが起きていそうかの分析を進めるにあたり，「比・割合」の領域を対象とすることにした．この領域は，全国学力・学習状況調査でも度々出題され，指導改善の必要性が主張されている．図 1 は，平成 24 年度全国学力・学習状況調査，算数 A の出題問題である（図 1）． (1)は，場面と図とを関連付けて，二つの数量の関係を理解しているかどうかをみるもので（正答 4），(2)は，1 に当たる大きさを求めるために，除法が用いられることを理解しているかどうかをみるものだった（正答 120÷0.6 など）．

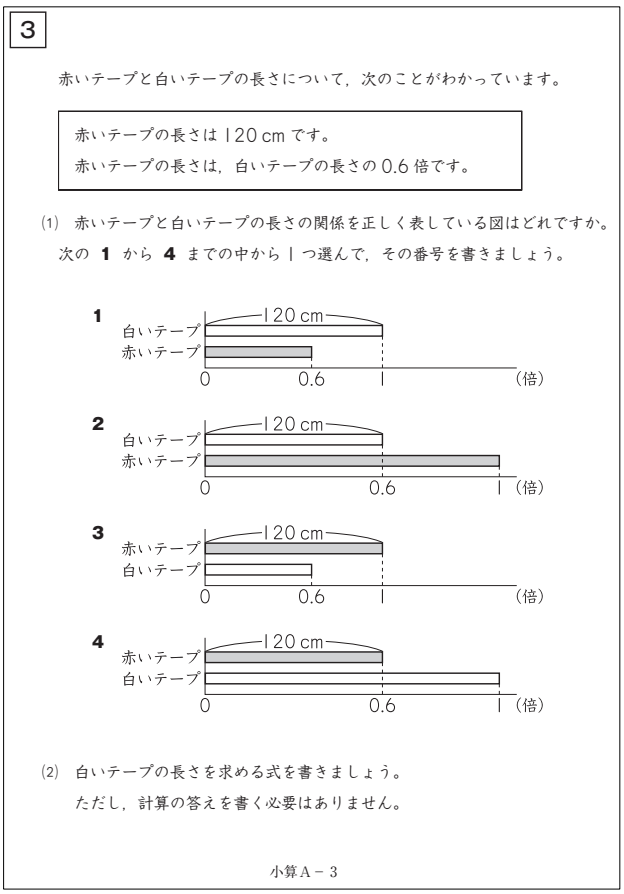

図 1 全国学力・学習状況調査の「比・割合」出題例

国立教育政策研究所の分析報告書[4]によると，(1)の正答率は 34.3%で(2)の正答率は 41.3%だった．また，(1)(2)の両方正答できた児童は，21.9%に留まった．正しい立式は機械的にできても，図のイメージができない児童がいることが分かる．また，最も多かった解答パタンは(32.2%の児童)，立式と図のイメージが両方誤っているが整合した解答，(1)は 3 を選択，(2)では 120×0.6 の掛け算で答えていた．これらより，問題文を十分に読み込まないまま，図と式の組み合わせを何らかの思い込みで選んでいる可能性がある．本来，教える側としては，問題文に対して，問題文を図と対応させていくことで意味を理解し，そしてそれを立式する，というプロセスを想定しているにも関わらず，児童たちは，そのようには学んでいない懸念がある．

普段の授業における指導において，見慣れた計算式や文章題が出たときにもっともらしい手続きで解答する力よりも，比べるとはどういうことなのか，何が基準となって（基準量），何が比較され（比較量），割合は何を示しているのかの意味理解の形成と，解答時にもそれを踏まえたプロセスが求められているところである．この状況に AI ドリルはどれだけ貢献しているのであろうか．

4. 分析対象 AI ドリルの概要

分析対象の AI ドリルは，稲垣ら[1]の研究と同じく，凸版印刷株式会社の適応型学習機能を組み込んだ「やる Key」とした（なお，2021 年 3 月にサービスを終了し，後継サービスは「navima」として展開）．

「やる Key」は，小学校３～６年生の算数を対象に，約 24,400 問が収録されていた．教員が一斉配信したドリルを児童が取り組み，児童の解答の正誤とそれ以前の学習履歴を基に，各児童の習熟状況に応じた問題を自動出題するシステムである．児童の取組状況はグラフ表示され，教員による個に応じた指導に役立てることができるとされていた．

5. 適応型学習機能について

表 1 は，「やる Key」のドリル教材の種別を示したものである．稲垣ら[1]と同様，論文用にラベル付けした．

つまずき判定による，ドリルの遷移パスを図式化したものが図 2 である．教師が児童に一斉提示するスタートポイントは④の標準ドリルで，児童はその標準ドリルの正誤具合によって，次に表示されるドリルが，③の推薦ドリルか，⑤の習熟ドリルかに遷移する．①の解説ドリル以外には，小問単位ですべてに，その問題を解くために必要な複数の知識を「つまずきポイント」として定義しタグ付けされている．解答の正誤判定時に，つまずきポイントをもとに児童が習得できて

いない知識・技能を特定する．つまずきポイントの情報と，児童の過去の学習履歴を合わせて，児童の次に提示するドリルをシステムが判断する仕組みとなっている．例えば，③の推薦ドリルで誤答となった場合，そのつまずきポイントが，該当単元内のつまずきであれば，②の基礎ドリルに遷移する．更にそのつまずきポイントが，前学年で学習した内容だと判断された場合は，前学年の該当する単元の③の推薦ドリルに遷移する．

表 1　「やる Key」のドリル種別

種別	内容
①解説	「基礎」で不正解児童への解説中心の小問提示（教科書の解説）
②基礎	特定のつまずきに対して理解を確認する基礎問題（前学年の単元の確認となる場合も）
③推薦	つまずきに応じて提示される低難易度のドリル
④標準	**教師が選択し児童に最初に提示するドリル，習熟度の算出対象となる**
⑤習熟	「標準」で満点をとった場合に出題される同難易度の問題
⑥発展	「習熟」ドリルで満点をとった場合に出題される発展問題

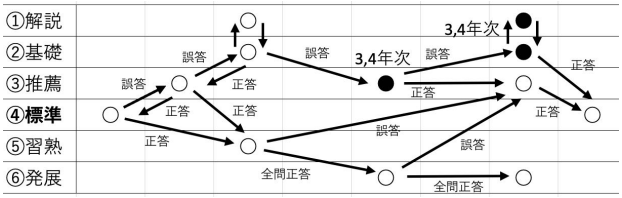

図 2　ドリルの遷移図

6.　出題形式について

「比・割合」の学習を含む５年生の単元は，「小数のかけ算」「小数のわり算」「割合」などがあるが，直接的な「割合」の単元は実施記録がなかったため，前出した全国学力・学習状況調査の問題と関連深い「小数のわり算」の取り組みデータを分析対象とした．この範囲で出題された各小問の出題形式は，単純な「計算問題」，立式ができれば解ける「文章題（図なし）」，図が何を意味するのか考えて立式する必要がある「文章題（図あり）」，②の基礎ドリルでの確認の結果遷移し出題される「3,4 年次の単元問題」，の 4 タイプに分けることができる．それぞれの小問は，１箇所～数箇所の解答欄から構成されている．それぞれの小問出題例を図 3～6 に示す．

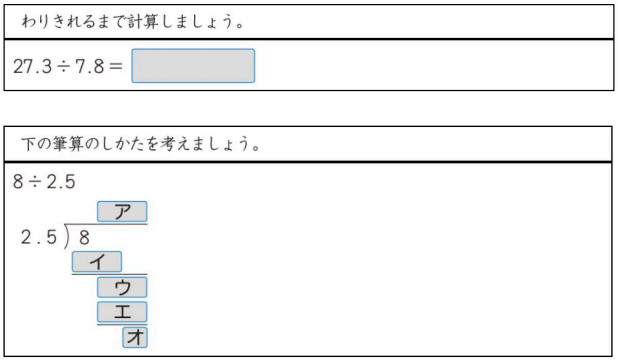

図 3　「計算問題」の出題例

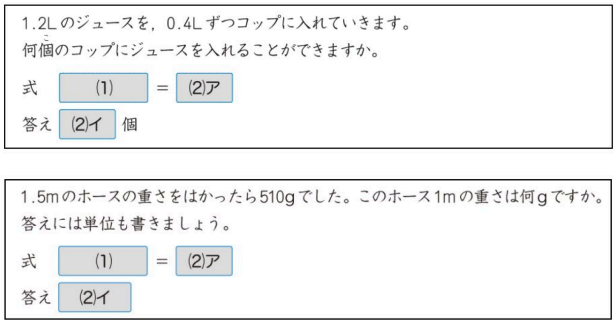

図 4　「文章題（図なし）」の出題例

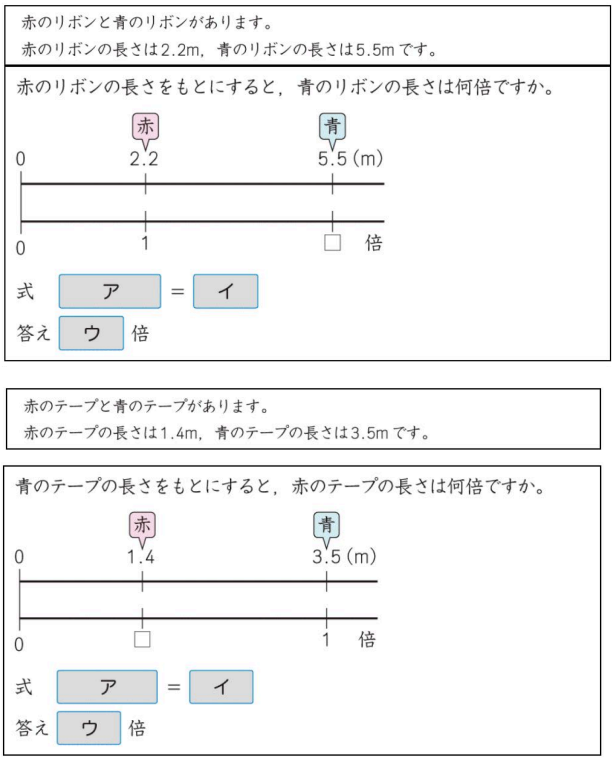

図 5　「文章題（図あり）」の出題例

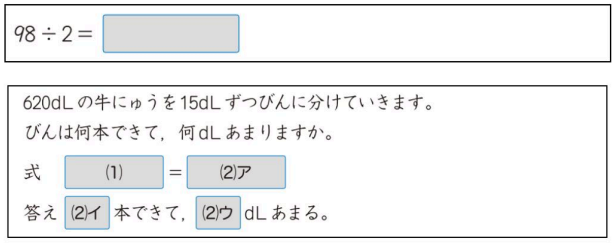

図 6 「3,4 年次の単元問題」の出題例

7. 対象データと分析方法

第 2 著者と凸版印刷株式会社との共同研究で収集されたデータ提供を受け，第 1 著者が主に分析を行った．

データは，宮城県の公立 M 小学校の 5 年生 54 名が一年間にわたって学校と家庭で使用し，2020 年 5 月から 2021 年 3 月の間に記録されたものを対象とした．全部で 111,161 件の解答データがあり，そのうち「小数のわり算」の単元での解答データは 8,971 件だった．8,971 件の中で，1,478 種類の小問が出題された．

児童たちの学力層別の取り組み具合を見ていくため，児童 54 名それぞれが全期間中に取り組んだ全設問の正誤から平均点を出し，上位群(14 名)，中上位群(13 名)，中下位群(13 名)，下位群(14 名)の 4 群に分けた．

本来はドリルとは独立した情報を基に学力を定義したいところだが，今回は AI ドリルに記録されたデータのみで分析を行ったため，稲垣ら[1]の研究成果で得られた単元テストの成績とドリル解答の成績が関連しているという報告に基づき，111,161 件の解答の正誤から各児童の平均正答率を算出し，それを児童の学力と仮に定義した．

設問解答データより児童の学習プロセスの違いについて分析可能な方法は以下の通りだった．

・学力層によって取り組み回数や正誤に差があるか
・学力層によって取り組みドリル種別に差があるか
・学力層によって取り組み出題形式に差があるか

8. 分析結果

小問内の解答欄の解答回数と正誤数を学力層ごとに整理したのが，図 7 である．また，学力層別に一人あたりの平均設問解答回数を出したのが表 2 である．表 2 より，上位ほど解答した回数が多く，下位層は上位層の 1/4 以下しか取り組んでいないことがわかる．また，下位層の誤答割合は 50%を超えており，他の層と比べても誤答割合が高いことが分かる．

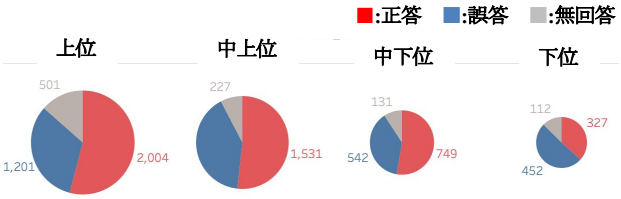

図 7 学力層別取り組み解答回数と正誤の割合

表 2 学力層別の一人あたり平均解答数

	上位	中上位	中下位	下位
平均解答数	264.7	227.6	109.4	63.6

次に，学力層ごとに解いた小問の種類を分類して示したのが表 3 である．児童たちは④の標準ドリルからスタートするが，どの学力層であっても，標準ドリルの誤答に応じて提示される③の推薦ドリルの小問を解いた回数が最も多かったことが分かる．また，⑤の習熟ドリルと⑥の発展ドリルは，上位，中上位層は取り組んでいたが，中下位層は出題機会が減り，下位層は一度も⑥の発展ドリルに取り組んでいなかったことが分かる．例として，上位層の A さんのドリル遷移を図 8 に，中下位層の B さんのドリル遷移を図 9 に示す．

表 3 ドリル種別・学力層別の取り組み小問数

	上位	中上位	中下位	下位
①解説	20	24	4	11
②基礎	16	13	6	3
③推薦	517	470	239	181
④標準	**76**	**76**	**76**	**66**
⑤習熟	177	170	77	25
⑥発展	146	118	61	0

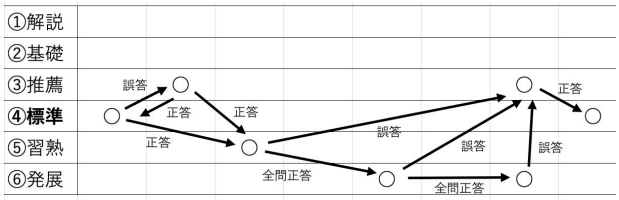

図 8 上位層の A さんのドリル遷移

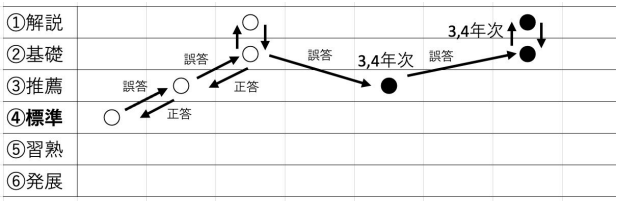

図 9 中下位層の B さんのドリル遷移

さらに，学力層ごとに取り組んだ出題形式に差があるかどうかを調べたところ，「計算問題」「文章題（図なし）」は，全員取り組んでいたが，「文章題（図あり）」の小問に取り組んでいた人は限られており，学力層によって差があった．「文章題（図あり）」に取り組んだ人数を表 4 に示す．中下位層，下位層は，そもそも取り組んだ人数が少なかったことが分かる．

これら「文章題（図あり）」を取り組んだ児童たちの，設問解答の正誤の数を分析したのが図 10 である．これより，中上位層の児童たちの誤答割合が，他の層と比べて高いことが分かる．

表 4　文章題（図あり）に取り組んだ人数

	上位 (n=14)	中上位 (n=13)	中下位 (n=13)	下位 (n=14)
人数	10 名	10 名	4 名	2 名

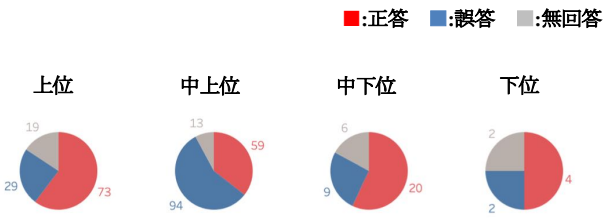

図 10　学力層別文章題（図あり）の正誤の数

9.　結果のまとめと原因の抽出

限られたデータ解析の範囲ではあるが，これら結果より，AI ドリルでどのような学びが起きていそうかを考察する．

上位層は，取り組み回数が多く，③の推薦ドリルに取り組み満点が取れるようになると，⑤の習熟ドリルや⑥の発展ドリルに遷移し，達成感を得ていた可能性が高い．

中上位層は，上位層と同様⑥の発展ドリルまで遷移しているが，途中出題される「文章題（図あり）」では誤答割合が高く，機械的な立式や計算ができても意味理解まで支援されていない可能性がある．

中下位層になると，「文章題（図あり）」を取り組む人数は 13 名中 4 名と大幅に少なく，「計算問題」や「文章題（図なし）」を繰り返すことで，計算間違えをしないような習熟や，機械的な立式への取り組みが主となっている可能性がある．

下位層では，⑤習熟ドリルでは誰一人満点にならず，③推薦ドリルの取り組みが中心となっている．全体的に正答率が低いにも関わらず，①の解説ドリルや②の基礎ドリルの取り組みも少なく，取り組み小問数も少ないということは，途中で諦めてしまっている可能性がある．

これら学力層によって取り組み具合が異なっていた原因として，以下の設計デザインによる影響が考えられるのではないか．

・出題形式：小問の出題形式が計算問題の正誤や文章題の立式に偏っている
・つまずきの判定基準：何の計算手続きができなかったかの判定に偏っている

小問の出題形式の偏りについては，出題された小問（実際に児童たちが解いた小問）をドリル種別と形式別に整理した図 11 を参照されたい．形式別③の推薦ドリルでは，「計算問題」が出題された割合が多く，「文章題（図なし）」は 1/3 程度で，「文章題（図あり）」はわずかである．⑤の習熟ドリルや⑥の発展ドリルでは多少「文章題（図なし）」の出題割合は高まるが，「文章題（図あり）」は少ない．

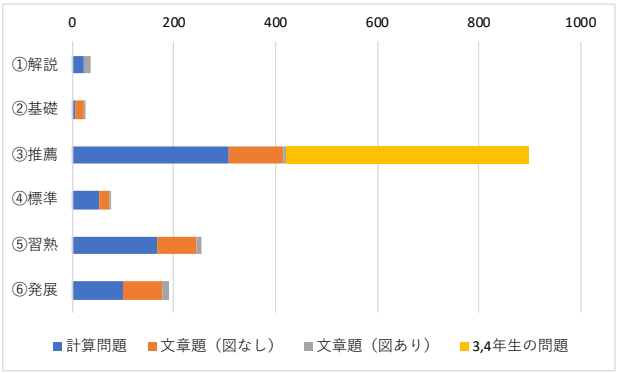

図 11　ドリル種別・形式別の出題された小問数

続いて，つまずきの判定基準が，計算手続きが間違っているか・いないかの表面的判定に偏っている点である．今回分析した AI ドリルは，小問一つ一つに，誤答の原因と想定される内容を「つまずき」としてタグ付けしていた．この「つまずき」判定が，次にいかなるドリルを推奨するかの基準になっていた．今回提供していただいたデータからは，全ての小問に対する「つまづき」情報を自動的に抽出することが困難であった．

そのため，分析過程で同定できた代表的な小問を複数ピックアップ（②〜⑥の各ドリル種別から各出題形

式の小問を選択）し，その小問にタグ付けされていた「つまずき」の定義テキスト情報を表 5 に転載した．抽出した小問の範囲では，②〜⑥のドリル種別によって「つまずき」の定義内容には質的な違いが見られず，出題形式によって「つまずき」の定義が追加されていることが分かった．ただ，追加された定義は表 6 の太字で示したような「文章題（図なし）」の「小数のわり算の立式ができない」という立式に関する判定と，「文章題（図あり）」の「小数でも，基準量が違うときに倍を用いて比べられるとわからない」という，比・割合に関する判定のみであった．

表 5　出題形式別つまずき定義のリストアップ

計算問題	・整数÷小数＝小数の計算ができない ・小数÷小数＝純小数の筆算ができない ・被除数に 0 を付け足して割り進められない ・除数と被除数に同じ数をかけると商が変わらないとわからない
文章題 (図なし)	・小数÷小数、あまりが出せない ・整数÷小数を整数÷整数にして計算できない **・小数のわり算の立式ができない**
文章題 (図あり)	・小数÷小数＝小数の計算ができない ・小数÷小数＝純小数の筆算ができない ・整数÷小数を整数÷整数にして計算できない ・被除数に 0 を付け足して割り進められない ・除数と被除数に同じ数をかけると商が変わらないとわからない ・被除数を付け足してわり進められない **・小数でも，基準量が違うときに倍を用いて比べられるとわからない**

10. 背後に考えられる学習者モデル

本研究では，AI ドリルに記録されたビッグデータ解析によって，「AI ドリルでどういう学びが起きていそうか」を捉えた．児童たちの解答状況の分析からは，すべての学力層の児童たちが，計算問題への解答や図なしの文章題での立式など，問題の意味理解を飛ばして可能な小問に繰り返し取り組んでいたことが見えてきた．また，中上位層であっても，図ありの文章題で誤答割合が高く，意味理解の習熟には繋がっていない様子が見えてきた．

このことは，今回のビッグデータに基づいた解析の限界とも関連する．適応型学習機能は，児童の理解状態よりも，解答結果に依存した推奨に特化されていたため，記録されているビッグデータから解答結果がどうだったかの分析は容易であったが，特定の小問に対する解答結果によって次の解答はどう変容したか，どのようなつまずきが多く見られるのかといったプロセスを捉えるような，児童の理解状況を可視化することが難しかった．このことはすなわち，AI ドリル内に学習者モデルとして意味理解の深さは想定されていないとも捉えることができるだろう．

11. 今後の改善可能性に向けて

今回分析した AI ドリルでは，「習熟＝正答が出せること」に主眼が置かれていたと考えられ，その学習目標に向けて，学習者モデルが最適化されていたのではないだろうか．その背景として，授業では意味理解を深め，その上で AI ドリルに取り組むことが期待されていたのかもしれない．しかし今回の分析では，意味理解を深める学習プロセスを見出すことが難しかった．

今後，AI ドリルが「習熟＝比・割合の意味理解をしっかりしたものにする」に貢献していくことができる改善可能性を提案する．

一つは，問題文と図の整合や，式と図の整合を問うような小問の割合を増やしていくことが第一であろう．比・割合の理解状況を推定できるような小問に児童が取り組む必要がある．もう一つは，そのような比・割合の理解状況を推定できるような小問に対して，いかなる解答や選択を行ったか，想定される学習者の正答・誤答モデルを解答類型として設定しておき，誤答具合によって判定できるようなつまずき判定の設計である．

その上で，新たに仮定した学習者モデルに沿った認知過程を引き出すことが実際できているかどうか，ビッグデータから認知過程を可視するような分析ツール，ダッシュボードの開発も重要となるだろう．そのようなツールを用いながら，子供たちの実態を把握し，それを基につまずき判定基準を見直し続けるような手続きを開発・運用プロセスに組み込むことが有用だと考えられる．そこには認知科学や学習科学の知見が生かされていくだろう．

謝辞

本研究は，凸版印刷株式会社と第２著者との共同研究からのデータ提供により実現した．データ解析には，聖心女子大学研究生田中冴氏の協力を得た．また，科学研究費補助金基盤研究 S (17H06107)ならびに基盤研究 C (18K02906)の支援を受けた．記して感謝する．

文献

[1] 稲垣忠, 大森裕二, 志野奈美子, 阿波弘真, 村上壮, 菊地尚樹,（2019）“学校および家庭における適応学習の実践と評価”, 日本教育工学会, Vol. 42, No. 4, pp. 345-354.
[2] Anderson, J. R., Boyle, C. F. and Yost, G., (1985) “The geometry tutor”, in Proceedings of the Ninth International Joint Conference on Artificial Intelligence.
[3] Bruer, T., J., (1993) “Schools for Thought. A Science of Learning in the Classroom”, Cambridge, MA: The MIT Press.
[4] 国立教育政策研究所, (2012) “平成 24 年度全国学力学習状況調査【小学校】報告書”. https://www.nier.go.jp/12chousakekkahoukoku/03shou_houkokusho.htm