

大学におけるデータサイエンス教育開始時の アンケート結果を用いた追跡調査

——大学初年次におけるデータサイエンスリテラシー科目の成績との関係——

平井 佑樹, 鈴木 治郎 (信州大学)

信州大学では2023年度より全学必修科目「データサイエンス (DS) リテラシー」を開講しており, 科目開始時に高等学校情報科やDSに関するアンケート調査を実施している。本研究では, 2024年度終了までに得られた約3,300名分の回答を用い, 回答内容とDSリテラシーの成績に関係があるか否かを調査した。その結果, (1) 情報科で利用した教科書, (2) 情報科の授業に対する好意度, (3) DSを教養として学ぶ重要性に関する意識, (4) DSが今後の進路において重要かどうかに関する意識, それぞれと成績に係り性は見られなかったものの, 上記(2), (3), (4)を説明変数とする重回帰分析の結果, 上記(3), (4)が成績に有意に影響していることが明らかになった。

キーワード: 追跡調査, 高大接続, 高等学校情報科, データサイエンス教育

1 はじめに

内閣府の統合イノベーション戦略推進会議で決定された「AI戦略2019」を受け, 各大学等でデータサイエンス (以下, DS) 教育が積極的に行われている。このDS教育では, 「数理・データサイエンス・AI教育強化拠点コンソーシアム」が発表したモデルカリキュラムをベースとする大学が多く, リテラシーレベルでは, 文理問わずすべての大学及び高等専門学校生が, 大学の学士課程等にて初級レベルの数理・DS・AIに関する知識やスキル等を習得することを目指している。

本稿執筆時点で最新のリテラシーレベルのモデルカリキュラム (数理・データサイエンス・AI教育強化拠点コンソーシアム, 2024) では, 高等学校「情報I」とモデルカリキュラムとの関係性が明示され, 現行のDS教育は, 高等学校情報科との結びつきが強いことが示された。しかしながら, 高等学校学習指導要領解説 (文部科学省, 2018) では「生徒や学校, 地域の実態を適切に把握し, 教育内容や時間の配分, 必要な人的・物的体制の確保, 教育課程の実施状況に基づく改善などを通して, 教育活動の質を向上させ, 学習の効果の最大化を図るカリキュラム・マネジメントに努めることが求められる」と示されていることから, 各高等学校等で実施されている教育内容が異なり, 大学へ入学してくる学生の習熟度も大きく異なる可能性がある。また, 大橋 (2022) は, 旧課程の情報科でDSを扱っているのは「情報の科学」であるものの, 教科書によってはDSにまで踏み込む内容が少なく, AIについての記載も少ないことを述べている。

このような背景を踏まえ, 信州大学では2023年度より学部1年次生を対象とした全学必修科目として

開講している「データサイエンスリテラシー (以下, DSリテラシー)」の履修者を対象として, 科目開始時に高等学校情報科やDSに関するアンケートを実施している。その目的は科目内容の検討や自己点検・評価 (平井ほか, 2024) であり, 現在も継続して実施している。一方, 平井ほか (2021) の報告を踏まえ, このアンケートの回答からDSリテラシーの成績を予測できないかと考え, 本研究の実施に至った。このような予測を実現できれば, たとえば, 成績が低くなりそうな学生に対して重点的に学習支援等が可能になる。本稿では, アンケートの回答内容と成績との関連性について調査した結果を報告し, どのような要素が成績の高低に寄与しているのかについて検討する。

2 本研究の位置づけ

2.1 DSリテラシーの概要および先行研究

DSリテラシー (1単位科目) は全8回で構成されており, 前述のモデルカリキュラム (リテラシーレベル) を踏まえて学習内容を決定した。達成目標は「DSの見方に沿って情報を捉えることができる。モデルカリキュラム『基礎』においては, データの図示の中に代表値や散布度などの統計量を読み取れる」であり, これからの社会のあり方を考えてもらうなど, DSの基礎を築いていくために不可欠な統計学, 情報科学, 数学の位置づけを理解してもらうようにした。そのため, いわゆる文系学生でも無理なく学習できるよう, 資料の読解を中心とした構成とした。なお, DSリテラシーを含む「信州データサイエンスプログラム (リテラシーレベル)」は, 文部科学省の「数理・データサイエンス・AI教育プログラム認定制度」における認定を2023年8月に受けている。

DSリテラシーは、非同期オンライン環境で、各年度の前期前半または後期前半に実施し、本学7学部(人文, 教育, 経法, 理, 工, 農, 繊維)と2学科(医[医], 医[保健])の計9部局それぞれに担当教員が1名ついたうえで、部局単位で教育・学習を進めるようにした。なお、前期・後期のどちらで履修するかは部局ごとに決められており、各学生の履修する時期はあらかじめ決まっている。

DSリテラシーの成績は次の3つでつけられる：

- ・小テスト(10点満点×7回)
- ・記述課題(20点満点)
- ・授業に対する良い質問の投稿や受講前・受講後アンケートへの回答による加点(20点満点)

これらの満点を合計すると110点となり、60点以上の獲得で単位修得となる。なお、100点以上獲得した場合は100点となる。

先行研究(平井ほか, 2024)は、2023年度前期前半の履修者約900名から得られたデータをもとにして、DSリテラシー開講の成果を明らかにし、科目運営に関する今後の課題を洗い出すことを目的として実施された。本研究で利用している科目開始時のアンケート(受講前アンケートとして学習管理システム上で実施)は先行研究で考案されたものであり、科目内容の検討や自己点検・評価目的で実施された。そのため、質問項目は本研究の目的である成績との関係性を調査するために設計されたものではないが、本研究の成績データとの関連を探索的に調査する目的で利用した。

2.2 関連研究

アンケート調査等で学生の習熟度や理解度を把握し、それを用いた追跡調査を実施している大学がいくつかある。たとえば、武田ほか(2022)は「数理・データサイエンス入門」の履修開始直後に、情報機器の操作レベルや数理・DSについて現時点で有する知識等を把握し、履修終了直前に科目を通して獲得した知識や今後獲得したい知識等を把握している。また、Gherghel・仲田(2024)は「データサイエンス入門」履修者のDSへの興味、DS科目の受講歴、および数学力について把握しており、分析の結果から、DSへの興味と数学力が課題提出率と成績に対して有意に正の影響を与えることを明らかにしている。

一方、DS教育に限定した追跡調査ではないものの、平・大久保(2018)は「高等学校での評定平均は、入学後の全学共通科目、専門科目、全科目のGPAと弱い相関を持つ」ことや「本学の志望順位や、本学を

積極的に選択したという意識、学部・学科への志望を自発的に行ったという意識が影響しうる」ことを述べている。また、天野・樽松(2025)は「主体的学習姿勢と適度な自習習慣を身につけた学生の成績推移がよく、大学名よりも学部・学科の専門性、自分の興味関心との一致度が入学後の成績に強く影響することが推定された」と述べている。平井ほか(2021)も、高等学校各科目に対する自己評価理解度と学部初年次成績に弱い正の相関が認められたことを明らかにした。

本研究の内容に最も近い、前述のGherghel・仲田(2024)の研究では、人文科学を専攻する学生を対象として調査した結果を示している。本研究では、人文科学も含む複数の分野(領域)の学生に調査しており、より一般的な関係性について明らかにすることを試みている。

3 本研究で利用するデータ

3.1 科目開始時のアンケート

本研究で利用するアンケート項目は、前述の受講前アンケートで尋ねた次の4項目である。

- ① 高等学校情報科で用いた教科書のタイトル(4択。1:「社会と情報」を含む、2:「情報の科学」を含む、3:上記2つの両方とも、4:上記以外)
- ② 情報科の授業は好きだったか(5択。1:嫌い、2:どちらかと言えば嫌い、3:どちらでもない、4:どちらかと言えば好き、5:好き)
- ③ 大学生時代の教養教育として、DSを学ぶことは重要だと思うか(5択。1:思わない、2:どちらかと言えば思わない、3:どちらでもない、4:どちらかと言えば思う、5:思う)
- ④ DSに関する知識・理解があなたの今後の進路において重要だと思うか(5択。選択肢は③と同じ)

このアンケートでDSは、北川・竹村(2021)が示している定義にしたがって、「データを処理・分析し、データから有益な情報を取り出す方法論」としている。

2024年度までの履修者4,084名のうち、約8割の3,289名から回答を得た。ここで回収したものは、各履修者が入学した年度の前期または後期の前半において回答したものであり、再履修時に回答したものは含まれていない。また、信州大学生以外(例:先取り履修した高校生)は履修者や回答者に含まれていない。

3,289名の所属学部(内訳)は次のとおりである: 人文286, 教育425, 経法333, 理275, 医423, 工767, 農307, 繊維473。また、前述の②③④の平均

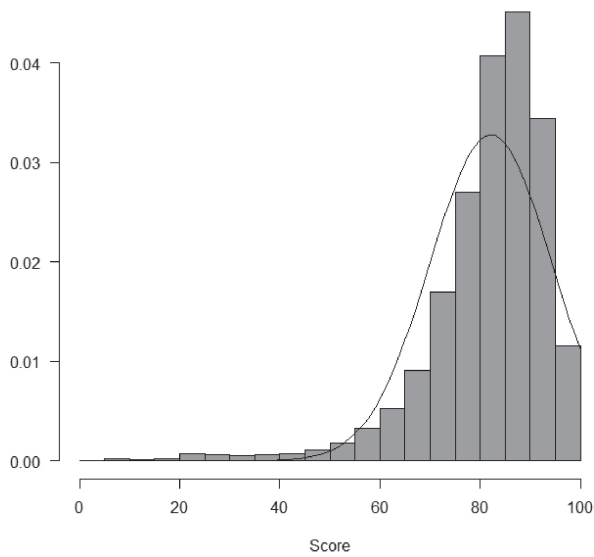


図 1 DS リテラシーの成績分布および正規分布曲線

値と標準偏差は次のとおりである。

- ・ 質問項目②：平均値 3.41, 標準偏差 1.05
- ・ 質問項目③：平均値 4.56, 標準偏差 0.68
- ・ 質問項目④：平均値 4.47, 標準偏差 0.76

3.2 DS リテラシーの成績

成績について、本学では小数点以下の入力も認められており、一部の部局ではそれが含まれる成績もあった。しかし、卒業判定等に利用される成績（秀, 優, 良, 可, 不可）では小数点以下は考慮されず、あらかじめ切り捨てた状態で成績を入力している部局もあったことから、本研究でも切り捨てて利用した。

図 1 に DS リテラシーの成績分布（平均値 82.1, 標準偏差 12.1）を示す。この図は統計分析ソフトウェア R version 4.3.1 を用いて作成（図 2 から図 5 も同様）したものであり、ヒストグラムおよび正規分布曲線が示されている。本研究ではこの図から、DS リテラシーの成績は概ね正規分布にしたがっていると判断し、以降の分析も同じく R を用いて実施した。

4 分析結果

分析では、まず 3.1 節で示した①から④のそれぞれに対して、選択肢ごとの回答数や成績の平均値・標準偏差を求め、選択肢間で平均点に有意な差があるかどうかを確認した。具体的には、3.2 節で述べた分布の正規性を前提として一元配置分散分析を行った。その後、成績を目的変数、3.1 節で示した②から④の選択肢を説明変数とする重回帰分析を行い、説明変数のそれぞれが、目的変数にどの程度影響しているのかを確認した。なお、各検定の有意水準は 5% としている。

表 1 情報科で利用した教科書との関係性
(選択肢ごとの回答数・成績の平均値・標準偏差)

	選択肢			
	1	2	3	4
回答数	1,589 (48%)	1,048 (32%)	73 (2%)	579 (18%)
平均値	82.4	81.9	83.9	81.5
標準偏差	11.8	12.2	8.6	13.5

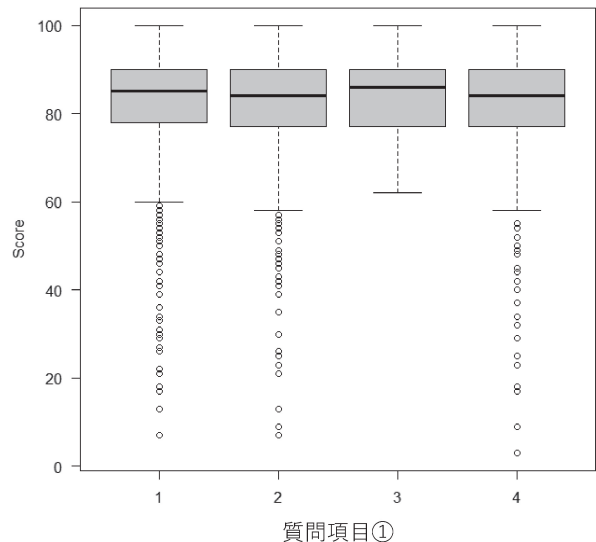


図 2 情報科で利用した教科書との関係性
(選択肢ごとの成績分布)

表 1 から表 4 で示している選択肢番号、および、図 2 から図 5 における横軸の選択肢番号は 3.1 節で示したものと対応している。また、図 3 から図 5 のキャプションで示している r 値はスピアマンの順位相関係数、p 値は無相関検定の結果である。

4.1 情報科で利用した教科書との関係性

標記の結果を表 1 および図 2 に示す。本アンケートでは、「社会と情報」を含む教科書を利用した学生が約 5 割、「情報と科学」が約 3 割となった。赤澤ほか (2024) が報告している教科書採択率（「社会と情報」80.9%、「情報の科学」19.1%）と比較すると、「情報の科学」を履修した学生の割合がやや大きいことが窺える。一方、「上記以外」と回答した学生が約 2 割いた。本研究ではこの要因を明らかにしていないものの、専門学科等で開設されている情報科の科目を履修した学生や、使用した教科書が分からない学生がこれを選択したと考えている。

選択肢間の平均値の比較において、「上記 2 つの両方とも」を選択した学生の平均値が他と比較してやや高く、かつ標準偏差もやや小さいように読み取れる。

表2 情報科の授業に対する好意度との関係性
(選択肢ごとの回答数・成績の平均値・標準偏差)

	選択肢				
	1	2	3	4	5
回答数	140 (4%)	517 (16%)	958 (29%)	1,188 (36%)	486 (15%)
平均値	81.4	82.3	81.8	82.6	81.7
標準偏差	11.9	11.0	12.2	12.1	13.5

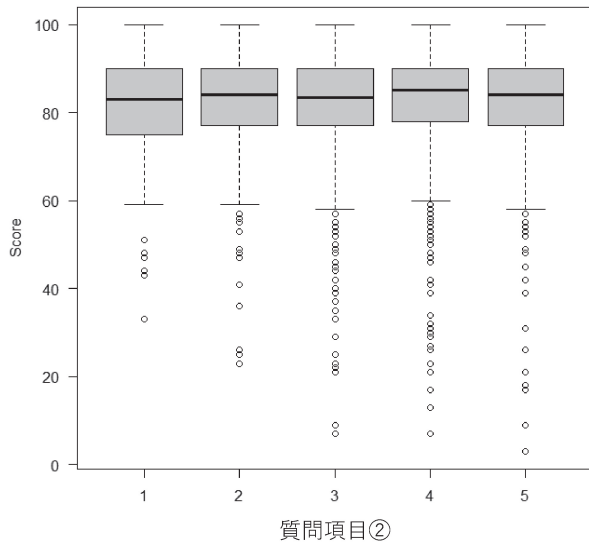


図3 情報科の授業に対する好意度との関係性
(選択肢ごとの成績分布, $r = 0.03$, $p = 0.09$)

しかし、一元配置分散分析において有意な差はみられなかった ($F(3, 3285) = 1.46$, $p = 0.23$)。

4.2 情報科の授業に対する好意度との関係性

標記の結果を表2および図3に示す。本アンケートでは、「好き」側の回答(選択肢4・5)が約5割、「嫌い」側の回答(選択肢1・2)が約2割となった。本アンケートでは、その選択肢を選んだ理由までは明らかにしていないものの、高等学校情報科の授業に対して、好意的な学生が比較的多いことが窺える。

選択肢間の平均値の比較において、図3を見ると、好意度が高いほど(=選択肢が5に近づくほど)分布の範囲が大きくなっているように見える。そのため、選択肢間で何らかの差があるように読み取れるものの、一元配置分散分析において有意な差はみられなかった ($F(4, 3284) = 1.00$, $p = 0.41$)。

1節で述べたように、各高等学校等で実施されている教育内容が異なり、学び方の違い(例:対面同期と非同期オンライン)等があることから、好意度との関連性が見られなかったのではないかと考えている。

表3 教養として学ぶ重要性に関する意識との関係性
(選択肢ごとの回答数・成績の平均値・標準偏差)

	選択肢				
	1	2	3	4	5
回答数	21 (1%)	42 (1%)	109 (3%)	1,015 (31%)	2,102 (64%)
平均値	80.8	84.6	81.4	82.5	82.0
標準偏差	13.1	9.5	15.5	12.2	12.0

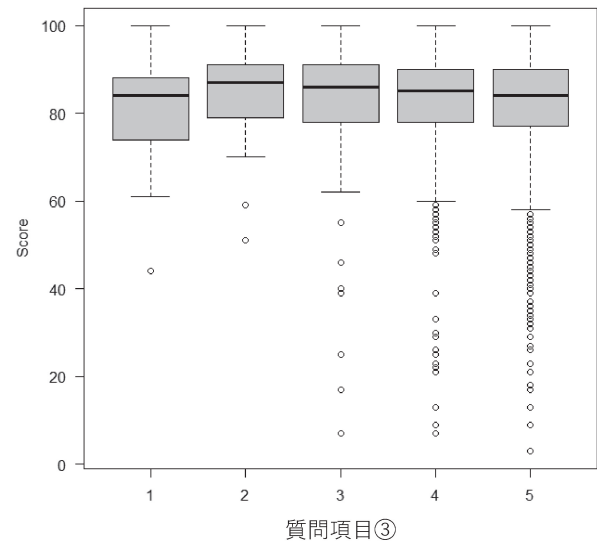


図4 教養として学ぶ重要性に関する意識との関係性
(選択肢ごとの成績分布, $r = -0.03$, $p = 0.0502$)

4.3 教養として学ぶ重要性に関する意識との関係性

標記の結果を表3および図4に示す。表3を見ると、9割を超える学生が、「思う」側(選択肢4・5)を選択していることが分かる。このことから、教養としてDSを学ぶ重要性については、本学内で浸透していることが窺える。重要性を認識していれば、意欲的に学習に取り組み、それが成績として表れると予想していたものの、表3で示した平均値や標準偏差では、選択肢間で極端な差は見られず、一元配置分散分析においても有意な差は認められなかった ($F(4, 3284) = 0.94$, $p = 0.44$)。

既報(平井ほか, 2024)では、入学当初に何となく「DSは重要だ」と思っている学生がいる可能性について言及した。この影響や2.1節で述べた学習環境等の影響について別途調べる必要はあるものの、この分析の結果から、学習する重要性に関する認識がそのまま成績に結び付くとは限らないことが示唆された。

4.4 今後の進路における重要度との関係性

標記の結果を表4および図5に示す。3.3節で示した項目と同様に、9割を超える学生が「思う」側を選

表 4 今後の進路における重要度との関係性
(選択肢ごとの回答数・成績の平均値・標準偏差)

	選択肢				
	1	2	3	4	5
回答数	25 (1%)	82 (2%)	154 (5%)	1,066 (32%)	1,962 (60%)
平均値	78.0	81.2	79.8	82.2	82.4
標準偏差	14.6	13.8	14.5	12.0	11.9

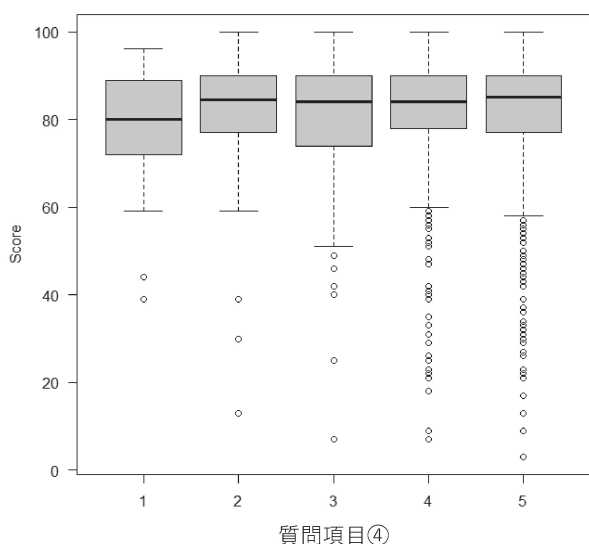


図 5 今後の進路における重要度との関係性
(選択肢ごとの成績分布, $r = 0.02$, $p = 0.16$)

択した。このことから、今後の進路においてもDSに関する知識や理解が重要であると認識していることが分かる。本項目では、一元配置分散分析において有意差が認められた ($F(4, 3284) = 2.41$, $p = 0.047$)。しかしながら、Tukey 法による多重比較では、有意差は認められなかった (p 値が最も小さかったのは選択肢 3 と 5 の間で 0.09)。

4.5 成績を目的変数とする重回帰分析

情報科の授業に対する好意度 (4.2 節)、教養として学ぶ重要性に関する意識 (4.3 節)、今後の進路における重要度 (4.4 節) それぞれを間隔尺度とみなして説明変数とし、成績を目的変数として重回帰分析を行った結果を表 5 に示す。ここで、VIF (Variance

Inflation Factor) は説明変数間の多重共線性を判断するための値であり、大塚 (2023) が示した基準にしたがえば、5 を目安としてそれを下回れば多重共線性は存在しないと判断できる。なお、調整済み R 二乗値は 0.003 ($F(3, 3285) = 3.91$, $p = 0.009$) であった。

この結果、回帰モデルとして説明力が高いとは言えないものの、モデルは妥当であることが認められた。また、切片の値を基準として、教養として学ぶ重要性は負の方向に、進路における重要度は正の方向に有意に影響を与えていることが認められた。一方、授業に対する好意度はここでも有意とはならなかった。

4.6 考察

1 節で述べたように、Gherghel・仲田 (2024) は DS への興味と数学力が課題提出率と成績に対して有意に正の影響を与えることを明らかにした。この研究における「興味」と本研究における「重要度」を単純に比較することはできないものの、DS に対する意識が成績に何らかの影響を与えていることが本研究の結果からも示唆された。

一方、粟津 (2024) は、データサイエンスに対する態度と数学に対する態度の関係性について約 800 名の回答結果を分析しており、「数学の必要性を認識しなかったり、学習意欲を示さない学習者でも、データサイエンスには肯定的な態度を示す可能性がある」ことや「データサイエンスでは、必要性を認識すれば学習意欲に結び付きやすいかもしれない」ことを述べている。本研究では、数学に対する態度との関係性については明らかにしていないものの、重回帰分析結果から、教養として学ぶ重要性和、具体的な進路における重要性で違いが表れた。

1 節で述べた、成績が低くなりそうな学生に対する学習支援について考えるならば、「教養として学ぶ重要性」について「思う」側を回答した学生が対象となる。しかし、この中には、

- ・平井ほか (2024) が言及しているような、単に何となく「DS は重要だ」と思っている学生
- ・重要だと思っていて、基礎学力も問題ないが、非同期オンライン環境での学習に慣れていない学生

表 5 重回帰分析の結果

	偏回帰係数	95% 信頼区間	t 値	p 値	VIF
授業に対する好意度	0.05	- 0.36 ~ 0.46	0.22	0.82	1.08
教養として学ぶ重要性	- 0.88	- 1.60 ~ - 0.17	- 2.43	0.01	1.37
進路における重要度	1.05	0.41 ~ 1.69	3.23	< 0.01	1.38
切片	81.31	78.20 ~ 84.41	51.36	< 0.01	-

・同様に、自己管理（小テスト受験期限の管理等）ができていない学生

などがあることが予想され、タイプによって支援内容を変えることが必要になるかもしれない。これについては今後の発展的な課題となり、具体的にはパス解析や一般化線形モデルによるモデル比較等を検討することにより、目的変数に対する要因間の複合的な影響をより詳細に解明できる可能性がある。

前述のモデルカリキュラム（数理・データサイエンス・AI教育強化拠点コンソーシアム，2024）では、カリキュラム実施にあたっての基本的な考え方の1つとして、“数理・データサイエンス・AIを活用することの「楽しさ」や「学ぶことの意義」を重点的に教え、学生に好奇心や関心を高く持ってもらう魅力的かつ特色ある教育を行う。数理・データサイエンス・AIを活用することが「好き」な人材を育成し、それが自分・他人を含めて、次の学修への意欲、動機付けになるような「学びの相乗効果」を生み出すことを狙う”と述べている。ここでの「次の学修への意欲、動機付け」と本研究における「進路における重要度」を単純に同等と見なすことはできないものの、モデルカリキュラムでの考え方を実現できれば、DSリテラシーではより良い成果を得ることができる可能性が本研究の結果から示唆された。

5 おわりに

本研究では、DSリテラシーを開講した2023年度から2024年度終了までに得られた約3,300名分のアンケート回答結果を用い、回答内容とDSリテラシーの成績に関係があるか否かを調査した。さらに多くの学生に調査したり、3.2節で示した学部等（分野）とは別の分野に所属する学生に調査したりすることで結果が変わる可能性があるものの、本研究の結果から、DSを教養として学ぶ重要性に関する意識やDSが今後の進路において重要かどうかに関する意識が成績に有意に影響していることが明らかになった。

2025年度からは、いわゆる新学習指導要領のもとで教育を受けてきた学生が履修するため、本研究で示唆された結果とは異なる傾向が明らかになる可能性がある。今後は、この可能性について検証するとともに、「進路における重要度」を高めてもらうための教育内容について検討を進める予定である。

謝辞

本研究はJSPS 科研費 JP24K06280 の助成を受けて実施した。また、本研究は信州大学「全学教育セン

ター等におけるヒトを対象とした教育研究に関する倫理委員会」での承認（承認番号第 R06-23）を受けて実施した。アンケートに回答いただいた学生の皆様、および本アンケート調査の実施に協力いただいた太田家健佑氏に御礼申し上げる。

参考文献

- 赤澤紀子・赤池英夫・柴田雄登・角田博保・中山泰一（2024）.「情報科教科書に現れる用語の変遷：情報ABCから情報I・IIまで」『情報処理学会論文誌：教育とコンピュータ』 **10**(1), 13-24.
- 天野哲彦・樽松理樹（2025）.「入学前の学習習慣や志望度と入学後の成績推移について」『大学入試研究ジャーナル』 **35**, 333-338.
- 栗津俊二（2024）.「数学が嫌いな学生・生徒は、データサイエンスも嫌いなのか。」『実践女子大学人間社会学部紀要』 **20**, 1-8.
- Gherghel Claudia・仲田知弘（2024）.「データサイエンス入門におけるアンケート調査と成績への影響」『日本教育工学会研究報告集』 **2024**(4), 181-186.
- 平井佑樹・高野嘉寿彦・小山茂喜・平野吉直（2021）.「高等学校における数学および理科の履修状況に関するアンケートの分析(3)―学部初年次の学業成績との関係―」『大学入試研究ジャーナル』 **31**, 91-96.
- 平井佑樹・鈴木治郎・太田家健佑（2024）.「信州大学全学必修科目「データサイエンスリテラシー」の開講とその成果」『情報処理学会論文誌：デジタルプラクティス』 **5**(4), 11-23.
- 北川源四郎・竹村彰通（編）（2021）.『教養としてのデータサイエンス』講談社.
- 文部科学省（2018年4月）.「高等学校学習指導要領（平成30年告示）解説 情報編」文部科学省
https://www.mext.go.jp/content/1407073_11_1_2.pdf（2025年4月2日）.
- 大橋真也（2022）.「高等学校共通教科「情報」におけるデータサイエンス―新学習指導要領解説およびその他の資料から見えること―」『コンピュータ & エデュケーション』 **52**, 18-23.
- 大塚芳嵩（2023）.「中級の統計解析（4）～重回帰分析～」『日本緑化工学会誌』 **48**(4), 613-618.
- 数理・データサイエンス・AI教育強化拠点コンソーシアム（2024年2月22日）.「数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム～データ思考の涵養～」数理・データサイエンス・AI教育強化拠点コンソーシアム
http://www.mi.u-tokyo.ac.jp/consortium/model_

literacy.html (2025 年 4 月 2 日).

平知宏・大久保敦 (2018). 「大学での学業成績に与える
高校評定、学びの意識の影響：潜在成長曲線モデルを
用いた検討」『大阪市立大学大学教育』 **16**(1), 16-25.

武田将季・坂本孝文・須藤智・滑田明暢・小西達裕 (2022).
「静岡大学における数理・データサイエンス教育の展
開と成果：「数理・データサイエンス入門」における
学習状況と習得意欲からみた教育効果の検証」『静岡
大学教育研究』 **18**, 87-96.