

大学入学共通テストで測定される「学力」へのアプローチ ——「数学」に関わるモニター調査の分析から——

○倉元直樹¹，石井裕基²

¹ 東北大学

1. 問題

大学入学共通テスト（以下、共通テストと略記）は、いわゆる高大接続改革の帰結として導入された。従来の大学入試センター試験（以下、センター試験）が『知識・技能』を問う問題が中心であるのに対し、『知識・技能』と『思考力・判断力・表現力』を総合的に評価する試験（中央教育審議会，2016）との位置付けで導入されたため、試験問題の特徴はセンター試験とは異なる。具体的には、数学の問題への会話文の導入等、問題の中に解答に結びつかない情報が多くなった。

改革の経緯は不透明で、測定すべき能力の議論が不十分なまま当初の構想とは全く異なる形式となった（南風原，2023，倉元，2025）。共通テストは構成や内容が不自然に歪んだように見える。物理では試験問題の文字数が増え、処理能力を問う試験と化した（加藤，2025）。英語は構成そのものが一部の分野に偏傾した内容となった（秦野，2025）。共通テストを「思考力を測る」とする言説には確たる根拠が見出せない。

共通テストに記述式を導入するとして公表された当初の「イメージ例」（高大接続システム改革会議，2015）は極めて本格的な記述式問題であった。そこで測定される学力については、文部科学省からの委託事業の枠組等の中でモニター調査が行われた。その結果、数学のイメージ例で測定される学力は、国語の新指導要領（当時）で必要と想定されていた資質・能力に近いと高校生に受け止められており（田中ほか，2018），新奇性の印象が強かった（東北大学，2019）。

本研究は、共通テストで実実施されている数学の問題に対して、測定されている「学力」が受験者側からどのように認知されているか、解明を試みる。

2. 方法

2.1. モニター調査問題と質問紙

共通テストで出題された問題2問²⁾（A問題：令和5年度入試本試験数学IA第2問（2），B問題：同追・再試験数学IA第2問（1））およびそれをセンター試

験風に改変した問題2問（P問題，Q問題）。AまたはP，BまたはQから1題ずつと出題順の組合せで8種類の問題冊子（冊子1 [A,B] ～冊子8 [Q,P]）を作成した。本稿ではA，B，P，Qを大問（小問各2問）と呼ぶ。改変に当たり、正解はオリジナルと同一とした。その結果、やや複雑な数値計算が含まれ、大問Qには大問Bで与えられた具体的な文脈の一部が残った。なお、配点は各大問25点満点、冊子単位では50点満点。

質問紙は田中ほか（2018）で用いられたものに必要最小限の改変を加えた。性別、文系・理系、解答時間、解答状況、大問ごとの印象12項目、解答に必要な資質・能力23項目、自由記述等が含まれる。

2.2. 調査対象者及び調査実施

全国各地の公立私立トップクラスから中堅の進学校16校の2年生。モニター調査問題の範囲は1年次に履修し終えている。実施時期は2024年12月～2025年3月、モニター問題と質問紙は郵送で送付、回収した。実施は学校に依頼、実施クラスごとに8種類の冊子をランダムに調査対象者に割り当てることとした。

モニター問題の解答時間は30分、質問紙への回答時間と配付、回収を含め、実施時間は45分とした。

東北大学高度教養教育・学生支援機構研究倫理審査委員会から2024年11月14日付で承認を得た。

3. 結果

3.1. 調査協力者数と平均点

表1. モニター解答結果（大問別）

大問	受験者数	平均	標準偏差
A問題	1,566	9.20	7.18
B問題	1,543	10.90	8.45
P問題	1,488	10.57	7.58
Q問題	1,511	9.54	7.49
全体	3,054	20.14	13.02

調査協力者は1校当たり29～292名、計3,054名。大問、冊子別の平均点等の概要は表1,2のとおり。

表2 モニター解答結果（冊子別）

冊子	受験者数	平均点	標準偏差
冊子1 [A, B]	406	18.83	12.76
冊子2 [A, Q]	404	18.34	11.55
冊子3 [P, B]	398	21.28	13.70
冊子4 [P, Q]	385	21.17	13.46
冊子5 [B, A]	382	21.01	13.26
冊子6 [Q, A]	374	18.82	12.24
冊子7 [B, P]	357	21.77	14.06
冊子8 [Q, P]	348	20.08	12.52
全体	3,054	20.14	13.02

3.2. モニター調査実施の適切性に関する検証

各大問と全く同じ問題を解いた経験があると回答した者は、A問題が76名(4.9%)、B問題が23名(1.5%)、P問題が23名(1.5%)、Q問題が24名(1.6%)であった。P、Qは本調査用に作題されたので、誤認である。調査用の問題はほとんどの調査対象者には初見であった。A問題の解答経験は成績に多少の影響があった。

全体の解答時間は「1. 短い」「2. やや短い」が75.6%で不足した。第1問の解答時間が平均16.2分、第2問は12.7分であった。第1問に着手しなかった者が34名(1.1%)の一方、第2問は160名(5.3%)に上った。

冊子別に共分散比を算出した。各大問ともで45.2%～54.8%の範囲に収まり、2問のバランスは悪くない。

5区分による設問解答率分析図の分析からも改変による影響は見られなかった。なお、大問B、Qの最後の小問は、下位群(第1～2五分位群)の正答率が中位群(第2～3五分位群)を上回る異常な形状を示した。

3.3. モニター調査成績の分析

大問ごとに問題種別(改変の有無)、出題順、文系・理系、性別を要因として、4元配置の分散分析を行って成績に与える要因について分析した。大問AとP、大問BとQで異なる結果が得られた。

大問AとPでは1次、2次、3次の交互作用は見られず、大問種別(大問A < 大問P, $F[1,2838]=28.97$, $P<.0001$)、出題順(第1問 > 第2問, $F[1,2838]=60.62$, $P<.0001$)、文系・理系(文系 < 理系, $F[1,2838]=184.93$, $P<.0001$)、性別(男子 > 女子, $F[1,2838]=28.97$, $P<.0001$)

の主効果は全て有意であった。

大問BとQでは2次、3次の交互作用は見られなかったが、大問種別(大問B > 大問Q, $F[1,2838]=19.70$, $P<.0001$)、出題順(第1問 > 第2問, $F[1,2838]=93.54$, $P<.0001$)、文系・理系(文系 < 理系, $F[1,2838]=163.41$, $P<.0001$)、性別(男子 > 女子, $F[1,2838]=39.72$, $P<.0001$)の主効果に加えて、大問種別と出題順の交互作用($F[1,2838]=8.86$, $P<.01$)、文系・理系と性別の交互作用($F[1,2838]=13.42$, $P<.001$)、が有意であった。

3.4. 各大問の印象

各大問の印象については、5段階評定のSD法で対になる12種類の形容詞対について、2問の大問それぞれについて2～2の5段階評定による回答を求めた。尺度化に当たって、各形容詞対に対して第1問と第2問の回答の和を取り、探索的因子分析を行った。相関行列のスクリー図を基にガットマン・カイザー基準によって2因子モデルを採用した。主因子法により因子抽出を行い、バリマックス回転を行った。結果を表3に示す。なお、表中に記載しているのは右側(プラス側)に示したことばである。

表3. 大問に対する印象(バリマックス回転後)

項目	第1因子	第2因子	h^2
12. 有意義な	.784	.240	.67
10. 役に立つ	.660	.259	.50
11. 実力がわかる	.642	.183	.45
3. まじめな	.555	.156	.33
9. 上品な	.519	.156	.29
4. 新しい	.381	.015	.15
2. 好きな	.245	.740	.61
7. できそうな	.042	.718	.52
8. 解きたい	.278	.709	.58
6. 素直な	.155	.617	.40
5. 典型的な	.101	.374	.15
1. おもしろい	.490	.514	.50
寄与	2.60	2.55	

東北大学(2019)に倣い、第1因子に負荷の高い6項目を「知識・能力」、第2因子に負荷の高い5項目を「関心・意欲」と命名し、尺度化した。なお、「新奇性」の軸は見出せなかった。2つの尺度の α 信頼性係数は双方とも.79、尺度間相関係数は.41であった。

各尺度および項目の平均値は表4に示す通りである。

表 4. 印象得点平均値

項目	全体	A	P	B	Q
知識・能力尺度	3.36	1.87*	1.48	1.69	1.60
3. まじめな	0.46	0.47	0.42	0.48	0.47
12. 有意義な	0.34	0.36	0.31	0.34	0.33
11. 実力がわかる	0.32	0.33	0.32	0.29	0.32
10. 役に立つ	0.25	0.30*	0.21	0.27	0.20
4. 新しい	0.19	0.28*	0.12	0.19	0.16
9. 上品な	0.11	0.14	0.08	0.11	0.09
関心・意欲尺度	-2.92	-1.44	-1.40	-1.30*	-1.73
5. 典型的な	0.07	0.05	0.09	0.11*	0.03
7. できそうな	-0.21	-0.15	-0.24	-0.17*	-0.29
6. 素直な	-0.28	-0.22*	-0.31	-0.25*	-0.34
2. 好きな	-0.52	-0.56	-0.49	-0.49	-0.55
8. 解きたい	-0.54	-0.57	-0.49	-0.52	-0.58
1. おもしろい	-0.02	0.01	-0.04	-0.02	-0.03

全体として、知識・能力尺度の尺度値が高く、関心・意欲尺度が低い。オリジナル問題と改変問題の間に5%水準で有意な平均値の差が存在する項目を「*」で

示す。改変による影響は、大問 P に比べて大問 A が知識・能力尺度値が有意に高く、大問 Q に比べて大問 B の関心・意欲尺度値が有意に高い結果となった。項目ごとにみると、知識・能力項目については「新しい」「役に立つ」で大問 A の値が有意に高く、関心・意欲項目については「典型的な」「素直な」「できそうな」で大問 B の値が有意に高い結果となった。

3.5. 各大問が測定する資質・能力

資質・能力項目は該当すると感じた項目に「○」を付す形式を取ったが、印象項目と同様の方法を用いて尺度化を行うことにした。結果を表 5 に示す。

第 1 因子に高負荷の 11 項目を「数学的思考力」、第 2 因子に高負荷 8 項目を「教養」、第 3 因子に高負荷の 3 項目を「言語能力」と命名、尺度化した。α信頼性係数は数学的思考力が .81、教養が .72、言語能力が .77、尺度間相関係数は数学的思考能力と教養が .50、言語能力が .43、教養と言語能力が .49 であった。

表 5. 大問が測定する資質・能力（バリマックス回転後）

項目	第 1 因子	第 2 因子	第 3 因子	h^2
14. 粘り強く、柔軟に考え考察を深めていく態度	<u>.560</u>	.069	.144	.34
15. 多様な考えを生かし、問題解決する態度	<u>.542</u>	.162	.124	.34
11. 考えを形成し深める力	<u>.537</u>	.177	.130	.34
8. 言葉や数式によって感じたり想像したりする力	<u>.498</u>	.102	.140	.28
21. 論理的思考力	<u>.498</u>	.165	.107	.29
20. 統計的思考力	<u>.497</u>	.241	.065	.31
9. イメージを言葉や数式にする力	<u>.496</u>	.168	.110	.29
6. 情報を様々な見方でよく調べ、まとめる力	<u>.493</u>	.171	.163	.30
22. 図やグラフを描いたり読んだりする力	<u>.490</u>	.032	.091	.25
10. 言葉や数式を通じて伝える力	<u>.453</u>	.284	.103	.30
0. 数学の基本的概念や原理・法則の体系的理解	<u>.445</u>	-.057	.079	.21
13. 歴史的に継承されてきた言語文化の担い手の自覚	.005	<u>.647</u>	.060	.42
19. コミュニケーション能力	.002	<u>.636</u>	.045	.41
7. 自分の文の構成や表現形式を振り返り評価する力	.198	<u>.503</u>	.162	.32
4. 言語文化に関する理解	.064	<u>.488</u>	.330	.35
18. 表現力	.323	<u>.456</u>	.109	.32
17. 要約力	.280	<u>.367</u>	.214	.26
5. 一般常識や社会規範	.194	<u>.351</u>	.195	.20
12. 自分の感情をコントロールして学びに向かう態度	.194	<u>.335</u>	.200	.19
2. 言葉の特徴や決まりに関する理解	.217	.242	<u>.657</u>	.54
3. 言葉の使い方に関する理解	.156	.260	<u>.623</u>	.48
1. 言葉の働きや役割に関する理解	.278	.146	<u>.614</u>	.48
16. 読解力	.413	.069	.291	.26
寄与	3.39	2.34	1.71	

各尺度の平均値と各項目の反応率を表6に示す。

全体として、数学的思考力尺度各項目の反応率が高く、次いで言語能力、教養が低い結果となった。オリジナル問題と改変問題の間に5%水準で有意な平均値の差が存在する項目を「*」で示す。改変による影響は、尺度値レベルでは大問Pに比べて大問Aが3つとも有意に高く、大問B、Qの間に差は見られなかった。項目ごとにみると、数学的思考力項目では「数学基本理解」「図やグラフ」「感受性」「イメージ表現」「伝達力」、教養項目では「要約力」「一般常識」「言語文化」、言語能力項目では3項目の全てで、大問Aの反応率が有意に高かった。大問BとQの間では、数学的思考力の「粘り強さ」で大問Qの反応率が有意に高く、「統計的思考力」で有意に低かった。

表6. 印象得点平均値

項目	全体	A	P	B	Q
数学的思考力	6.30	6.40*	6.20	6.32	6.28
0. 数学基本理解	0.85	0.87*	0.83	0.85	0.85
22. 図やグラフ	0.76	0.79*	0.75	0.75	0.74
8. 感受性	0.64	0.67*	0.62	0.61	0.63
14. 粘り強さ	0.63	0.64	0.63	0.61	0.64*
9. イメージ表現	0.58	0.64*	0.54	0.58	0.56
21. 論理的思考力	0.56	0.57	0.56	0.56	0.54
6. 情報集約	0.53	0.53	0.53	0.52	0.53
11. 思考形成	0.51	0.52	0.51	0.50	0.51
15. 多様な考え	0.45	0.45	0.46	0.44	0.46
20. 統計的思考力	0.41	0.40	0.41	0.45*	0.39
10. 伝達力	0.38	0.43*	0.35	0.38	0.37
教養	1.25	1.37*	1.13	1.24	1.27
17. 要約力	0.27	0.33*	0.22	0.27	0.27
5. 一般常識	0.23	0.27*	0.19	0.24	0.23
12. 感情統制力	0.23	0.23	0.23	0.23	0.24
18. 表現力	0.19	0.21*	0.17	0.18	0.20
7. 文章評価力	0.13	0.14	0.12	0.12	0.14
4. 言語文化	0.12	0.13*	0.11	0.12	0.11
13. 歴史継承	0.04	0.04	0.03	0.03	0.03
19. コミュ力	0.04	0.04	0.04	0.04	0.04
言語能力	1.19	1.27*	1.11	1.18	1.21
1. 言葉の働き	0.50	0.57*	0.44	0.48	0.50
2. 言葉の特徴	0.36	0.40*	0.34	0.35	0.34
3. 言葉の使い方	0.34	0.38*	0.31	0.34	0.33
16. 読解力	0.73	0.79*	0.68	0.72	0.73

4. 考察

本研究の方法論的特徴は、高大接続改革当初の議論を参照するため、当時の新構想試験問題（イメージ例）を用いたモニター調査と実質的に同一の質問紙を用

いたことにある。その結果、東北大学（2019）で独立の尺度として現れた「新奇性」は再現されなかった。調査対象者にとって共通テスト問題は新奇な問題ではなくなっているのだろう。共通テストとその設計思想が高校生の学習とそこから形成される教科観、学習観に影響を与えていることが示唆された。

全体として問題の個別性の影響が強く表れる結果となった。本研究はセンター試験と対比した共通テストの特徴抽出を目的とした調査であったが、問題の解答に必要な数学的要素を変えずにセンター試験風に改変する作業は意外と困難だった。A問題は改変で正答率が上がったがB問題が逆に下がる結果となったのは、当初の想定とは異なる。何を比較しているのかという問題は残るが、本研究と類似のモニター調査を企画するならば、共通テストからの改編ではなくセンター試験の過去問を利用した方が明瞭な結果が得られたかもしれない。また、難易度予測を誤ったことも調査結果に影響した。平均正答率が4割程度というのは低すぎた。解答時間30分は多くの調査校を得るべく設定されたが、学力を十分に測るには短すぎた。その結果、第1問と第2問の成績に差が生じた。既習の内容であり、オリジナルは公開問題だったが、大学入試を想定した演習には至っていない時期だったのだろう。逆に言えば、テストで測定可能な目に見える学力形成の上で、演習の重要性を再確認する結果となった。

共通テストの学力観を検証することは重要な学術的課題である。個別問題からのアプローチは困難であっても、更なるエビデンスの積み上げが必要である。

注釈

- 1) 文部科学省「大学入学者選抜改革推進委託事業」選定事業人文社会分野（国語科）（平成28～30年度）「個別学力試験『国語』が測定する資質・能力の分析・評価手法に関する研究——記述式問題を中心に——、北海道大学（選定機関）、東北大学・九州大学・長崎大学・大学入試センター（連携機関）。
- 2) 共通テスト問題の利用については大学入試センターからの許可を得た。

謝辞

本調査の構想段階で大野真理子講師、林如玉助教（いずれも東北大学）の協力を得た。本研究はJSPS 科研費 JP21H04409 の助成による研究成果の一部である。