

新共通テスト（イメージ例）が測定する資質・能力の分析（1）

—高校生対象のモニター調査から—

田中光晴，宮本友弘，倉元直樹（東北大学）

文部科学省は新しい大学入試の制度設計を進めている。2020 年度導入予定の新共通テスト「大学入学希望者学力評価テスト（仮称）」で採用する記述式問題の「イメージ例」が 2015 年末に公表された。これに対して様々な議論が起こったが，本研究の目的は，新共通テストに導入予定の記述式問題が何を測ることになるのかについて，当事者となる高校生の視点から分析することである。既存の個別学力試験問題やセンター試験との比較を通し，「イメージ例」が測る資質・能力について，実証的なデータに基づき検討する。

1 問題と目的

文部科学省は新しい大学入試の制度設計を進めている。これに伴い，2020 年度導入予定の新共通テスト「大学入学希望者学力評価テスト（仮称）」で採用する記述式問題の「イメージ例」が 2015 年末に公表された。国語の「例 1」は，多様な見方や考え方が可能な題材として，警察庁事故統計資料に基づいて作成された交通事故の発生件数，負傷者数，死者数のグラフが示され，それを基にした高校生の話し合いの内容について考えを表現する問題であった。高大接続システム改革会議（2016）では，記述式を導入する意義として，より主体的で論理的な思考力・判断力の発揮が期待できるとしている。この「イメージ例」については，専門家会議の委員から「高校教育が変わる」と評価する意見が出た一方で，採点の方法やコストといった課題が指摘されている（例えば，毎日新聞，2015 年 12 月 22 日付）。

テストは，学力を測定するためのものさしとしてのみ存在するわけではない。テストが，教師と学習者，選抜者と被選抜者との関係の中で，社会システムの一つとして実施されるとすれば，テストは単なるものさしとしてだけでなく，各自により様々な意味や価値づけが行なわれる。「テスト」を入試という意味に限って使用すれば，学習者は様々な影響を「テスト」から受けるわけである。この場合，テストの開発者側の判断で測定しようとする特性を十分に測定できているかどうかだけを満たせば良問であるという事にはならない。

「テストがどのような影響を学習者に与えるのか」ということまでも含めて考えるべきである。いわゆるテストの波及効果（washback effect, Anderson & Wall, 1993）として論じられてきた問題である。

テストの開発者が規定する測定内容は，それだけでは表面的妥当性（face validity）と呼ばれるものにす

ぎない。表面的妥当性とは「テスト課題が想定されたテスト得点の解釈やその利用に対して有する外見上の適切性」であり，「受検者や様々な関係者からのテストの受容を促すために努めて築かれる概念」である。しかしながら，「外見上の適切性は（測定された）特性の解釈について保証するものではない。（以上，Kane, 2006: 36, 筆者訳）。したがって，「テストが何を測っているのか」という事実は様々な側面から多角的，かつ，実証的に検証していく必要がある課題なのである。特に，大学入試のようなハイスタークな場面で用いられるテストでは，開発者側が高次の能力を測ろうと企図しても，学習者は想定されていたこととは別の学習行動を起こすことが考えられる。したがって，テストを作成する際には，「学習者にどのように見えるのか」ということを意識する必要がある。

本研究で取り上げる「イメージ例」も例外ではない。高校生の目に，既存の入試問題と比べてどう映ったのかを明らかにする必要がある。「イメージ例」が開発者側の判断で「新しい」タイプの問題だとされていても，受検者が「その課題によって測られている資質・能力は変わらない」と感じていたら，それは見た目だけが新しいということになる。一方，文部科学省が提示する養成したい資質・能力が「イメージ例」でしか測れないのであれば，既存のテストは大いに改善の余地があることになる。

そこで本研究では，既存の大学入試問題（センター試験や個別学力試験）と「イメージ例」に対して，高校生がどのような印象をもち，それらの問題を解くのにどのような資質・能力が必要だと認知しているのかについて比較検討する。したがって，実際にこれらのテストが高校生の資質・能力を測定できているかどうかを問うわけではない。つまり，テストが「実際に何を測っているか」というテスト自体の性能ではなく

「何を測っていると思うか」というテストを受ける側の認知に着目することに本研究の特徴がある。

2 方法

2.1 調査対象

本調査では、調査内容の難易度の問題に十分解答する必要があるため、進学校に協力依頼を出し承諾が得られた1校で実施した¹⁾。対象者は高校2年生の214名で、実施時期は2016年11月である。本報告ではこの調査から得られた結果を報告する。その後2017年2月から3月にかけて同様の調査を全国4か所で行なっている。その結果は別稿としたい²⁾。

2.2 調査内容

2.2.1 試験問題の抽出

国語と数学の試験ともに、①過去のセンター試験(国語の第2問, 数学の①)、②過去の国立大学の個別学力試験(国語の第4問, 数学の②)、③公開されている「大学入学希望者学力評価テスト(仮称)」のイメージ例(国語の第3問, 数学の③)から構成した。問題の選定にあたっては、高等学校で30年以上の教職歴のある国語科と数学科の教員それぞれ1名に依頼した。また、調査対象者の基礎的能力を測定するために平他(1998)が開発した語彙テストから項目反応理論(IRT)の項目パラメータに基づき20項目を選定し、国語の第1問とした。

2.2.2 試験問題評価用質問紙の設計

国語と数学の試験ごとに2種類の質問紙を作成した。試験直後に実施するもので、①解答時間(80分)の評価、②同じ問題の解答経験、③各問題の難易度の評価と解答時間、④無解答問題の有無とその理由、⑤各問題に対する印象(SD法)、⑥各問題を解答するのに必要な資質・能力、⑦意見・感想の自由記述等について尋ねた。

本報告が主に使うデータは上記⑤と⑥である。⑤各問題に対する印象については、SD法を採用し、それぞれの試験問題に対し個々の被験者がそれぞれの好みや経験に応じて感じる感情的意味を測定することにした。前述したように、試験が与える被験者側への影響を被験者の立場から評定することを目的としているためである。SD法では多くの形容詞対を両側に置いた評定尺度群が用いられる。それらの形容詞対は、「良い-悪い」など価値に関係したもののだけでなく「温かい-冷たい」「強い-弱い」など多面的であるのが特色である(大山他, 2005)。本調査では、「つまらない-おもしろい」「嫌いな-好きな」「ふざけた-まじめな」「古い-新しい」「奇抜な-典型的な」

「意地悪な-素直な」「むりそうな-できそうな」「解きたくない-解きたい」「下品な-上品な」「役に立たない-役に立つ」「実力がわからない-実力がわかる」「無意味な-有意義な」の12項目を設定し、5段階尺度で構成した。

また、⑥各問題を解答するのに必要な資質・能力については、「高大接続改革の進捗状況について」(文部科学省, 2016)に示された資質・能力(国語に関する「資質・能力」と中央教育審議会答申(2014年12月22日)の「別添資料4」(言語に関する思考力・判断力・表現力と数に関する思考力・判断力・表現力)から抽出した。抽出した23項目中(後掲, 表3)1, 21, 22, 23の4項目は数学と関連した資質・能力, その他19項目は国語と関連した資質・能力と捉えられる。質問紙では、項目1を除いて同じ項目を設定し、それぞれの大問を解くのに必要と思う資質・能力すべてに○をつけさせるようにした。

2.3 調査の手続き

調査はクラスごとに集団で実施された。まず、国語の試験を実施し(80分)、終了後、質問紙に回答してもらった(10~15分)。休憩後、数学も同様の手順で実施した。

2.4 分析方法

以上の調査から得られたデータのうち、印象については評定値の平均を問題ごとにプロットし、有意差検定を行った。資質・能力については問題とのクロス集計表を作成し、対応分析を適用し、視覚的に各項目の関係を2次元平面上にマッピングした。尚、対応分析においては最大3軸まで取り、もっともそれぞれの項目の布置が解釈しやすい2軸を選んだ。

3 結果

3.1 印象の分析

図1は、国語のプロフィールを示したものである。問題のタイプを要因にして分散分析を行った結果、表1の通りとなった。

第1問(語彙問題)は、他の問題と比べ、より「できそうな」「解きたい」「役に立つ」という印象であった。第2問は、相対的に「嫌いな」「まじめな」「意地悪な」「無理そうな」「解きたくない」問題であるが「実力がわかる」という印象であった。問題の種類が論説文であることを考慮するとマークシート方式という解答形式の印象というよりは「論説文」に対する印象と解釈できよう。第3問(イメージ例)の特徴は「新しい」「奇抜な」「実力がわからない」「無意味な」という印象であった。被験者にとつ

てイメージ例はこれまで国語の問題として出題されるタイプのものとは少し異なるものとして映ったようである。最後に第4問は、「おもしろい」「好きな」「解きたい」「役に立たない」という印象であった。第4問は小説文であったが、第2問同様、記述式問題というよりも内容に対する印象として捉えることが自然である。

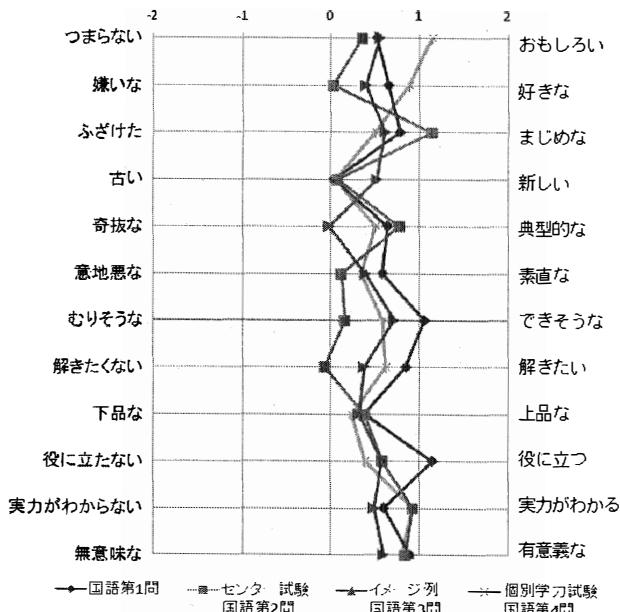


図1 国語の各問題への印象

表1 国語の問題間の印象差

	F	多重比較の結果 (Bonferroni, $p < .05$)
つまらない	32.81**	セ=語=イ<個
嫌いな	32.58**	セ<イ<語<個
ふざけた	42.72**	個=イ<語<セ
古い	26.99**	語=個=セ<イ
奇抜な	39.59**	イ<個=語=セ
意地悪な	14.35**	セ<個=イ=語
むりそうな	53.04**	セ<個=イ<語
解きたくない	38.21**	セ<イ<個<語
下品な	4.69**	個<語=セ, 個=イ, イ=個=セ
役に立たない	46.50**	個<イ=セ<語
実力がわからない	16.79**	イ=語<個=セ
無意味な	8.50**	イ<個=セ=語

** $p < .01$

語:語彙 セ:センター試験 イ:イメージ例 個:個別学力試験

図2は、数学のプロフィールを示したものである。問題のタイプを要因にして分散分析を行った結果、表2の通りとなった。

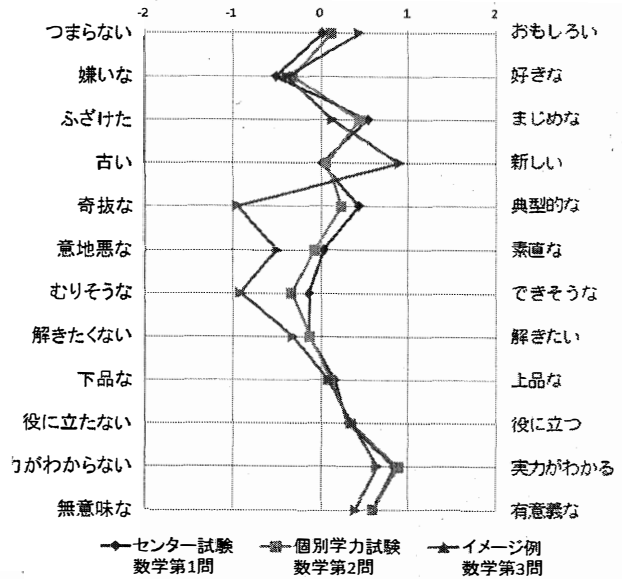


図2 数学の各問題への印象

表2 数学の問題間の印象差

	F	多重比較の結果 (Bonferroni, $p < .05$)
つまらない	20.33**	セ=個<イ
嫌いな	3.02	
ふざけた	29.21**	イ<個=セ
古い	127.12**	セ=個<イ
奇抜な	188.47**	イ<個<セ
意地悪な	39.87**	イ<個=セ
むりそうな	50.60**	イ<個<セ
解きたくない	4.02	
下品な	2.87	
役に立たない	0.29	
実力がわからない	8.54**	イ<セ=個
無意味な	9.81**	イ<個=セ

** $p < .01$

語:語彙 セ:センター試験 イ:イメージ例 個:個別学力試験

第1問と第2問の傾向はほぼ同様である。大きく異なるのは第3問である。第3問は、他の問題と比べ相対的に「おもしろい」「ふざけた」「新しい」「奇抜な」「意地悪な」「むりそうな」「実力がわからない」「無意味な問題」という印象であった。大学入試の過去問や各種教材での勉強に慣れてきた生徒にとって第3問は奇異に映ったことが見て取れる。一方、「嫌いな」「解きたくない」「下品な」「役に立たない」では、問題間で有意差はなかった。いずれの問題も「嫌いな」と印象を与えるのは、国語と異なる点で、数学という科目や数学の各問題で扱った三角関数の領域に対する印象とも解釈できよう。

3.2 資質・能力の分析

資質・能力について当該問題の解答に対して「必要」と答えた調査対象者の比率を表3に示す。「1 数学的概念」は国語で必要とされていない能力・資質である。「2 言葉の働き」～「20 表現力」は国語の新指導要領や新テストなどで必要と構想される資質・能力、「21 統計的思考力」～「23 図やグラフ」は同様に数学に関する資質・能力である。各資質・能力の項目ごとに、相対的に最も多くの調査対象者が○を付した項目をボードで示している。

つづいて、以上の結果を全体として概観するために資質・能力について対応分析の結果を示す(図3, 図4)。ここではより解釈が容易と考えられる、第1軸を横軸(イナーシャ=.205, 寄与率 56.9%)、第3軸(イナーシャ=.059, 寄与率 16.4%)を縦軸に描いた散布図と第2軸(イナーシャ=.067, 寄与率 18.6%)、第3軸の散布図に基づいて解釈を行う。

図3は、第3軸を縦軸に、第1軸を横軸にとったものであり、3つの群に分かれる。第1群では、国

語第1問(語彙問題)と言語文化、言葉の特徴、言葉の働き、歴史継承、言葉の使い方、一般常識と近接していることがわかる。

第2群には、数学の全問題と国語第2問、国語第4問が布置されており、近接する資質・能力は論理的思考力、感受性、粘り強さ、感情統制力、読解力、イメージ表現、多様な考え、思考形成、伝達力、要約力である。第3群は国語第3問が布置されたが、資質・能力は表現力、文章評価力、情報集約、統計的思考力、コミュニケーションである。

図4は、第2軸を縦軸に、第3軸を横軸にとったものであり、大きく2群に分かれる。第1群は、数学の各問題が布置されているが、第3問が、言葉の働きや言葉の特徴、多様な考えと近接していることが興味深い。一方第2群は、国語の第3問を除いた問題が布置され、歴史継承、言語文化、思考形成、伝達力、表現力、言葉の使い方が近接する資質・能力として認知されていることがわかった。

表3 「資質・能力」が必要と回答した調査対象者の比率

	国語				数学		
	第1問	第2問	第3問	第4問	第1問	第2問	第3問
1 数学的概念					96.3%	82.2%	62.6%
2 言葉の働き	90.7%	53.7%	21.5%	36.9%	22.9%	29.0%	54.7%
3 言葉の特徴	76.2%	40.7%	15.4%	24.3%	19.6%	27.6%	36.0%
4 言葉の使い方	76.6%	44.9%	45.3%	42.1%	9.8%	23.4%	41.1%
5 言語文化	49.1%	36.0%	12.6%	23.8%	2.8%	4.7%	12.6%
6 一般常識	68.2%	52.8%	68.2%	24.3%	8.9%	6.5%	30.8%
7 情報集約	0.5%	36.9%	92.1%	29.9%	21.5%	24.8%	65.0%
8 文章評価力	0.9%	28.0%	79.9%	82.2%	6.1%	31.8%	36.4%
9 感受性	22.0%	32.7%	33.6%	87.4%	46.7%	63.6%	74.3%
10 イメージ表現	6.5%	25.7%	69.2%	87.4%	40.2%	61.7%	78.5%
11 伝達力	7.9%	33.6%	83.6%	80.8%	17.3%	67.8%	72.4%
12 思考形成	2.3%	56.1%	71.5%	63.1%	31.3%	55.1%	68.2%
13 感情統制力	10.7%	35.5%	22.0%	48.1%	22.0%	22.9%	29.9%
14 歴史継承	34.6%	28.0%	9.3%	16.8%	3.3%	4.7%	9.3%
15 粘り強さ	4.7%	69.6%	52.8%	64.5%	51.9%	65.0%	80.4%
16 多様な考え	4.2%	45.8%	67.8%	30.8%	43.0%	51.4%	78.5%
17 読解力	3.3%	97.7%	40.2%	92.5%	21.0%	34.1%	70.6%
18 要約力	1.4%	65.0%	48.1%	77.6%	2.3%	8.9%	21.0%
19 表現力	3.3%	36.0%	81.3%	85.5%	4.2%	33.2%	41.1%
20 コミュニケーション	11.2%	10.3%	54.7%	32.7%	1.4%	5.6%	9.8%
21 統計的思考力	0.9%	16.4%	87.9%	7.5%	11.7%	16.4%	38.3%
22 論理的思考力	1.4%	83.6%	40.2%	49.1%	40.2%	58.4%	62.1%
23 図やグラフ	0.0%	3.7%	87.9%	1.9%	53.3%	41.1%	76.6%

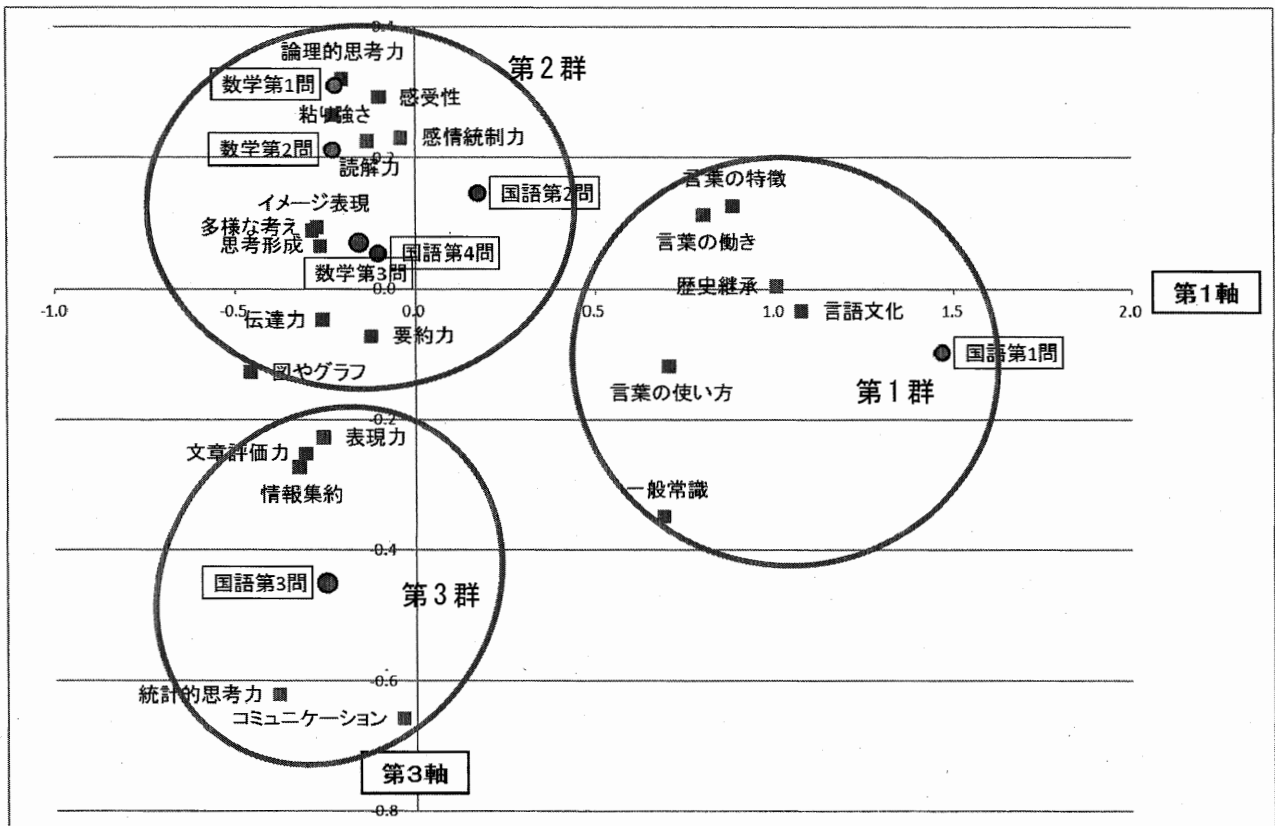


図3 資質・能力と各問題への印象のマッピング（1-3軸）

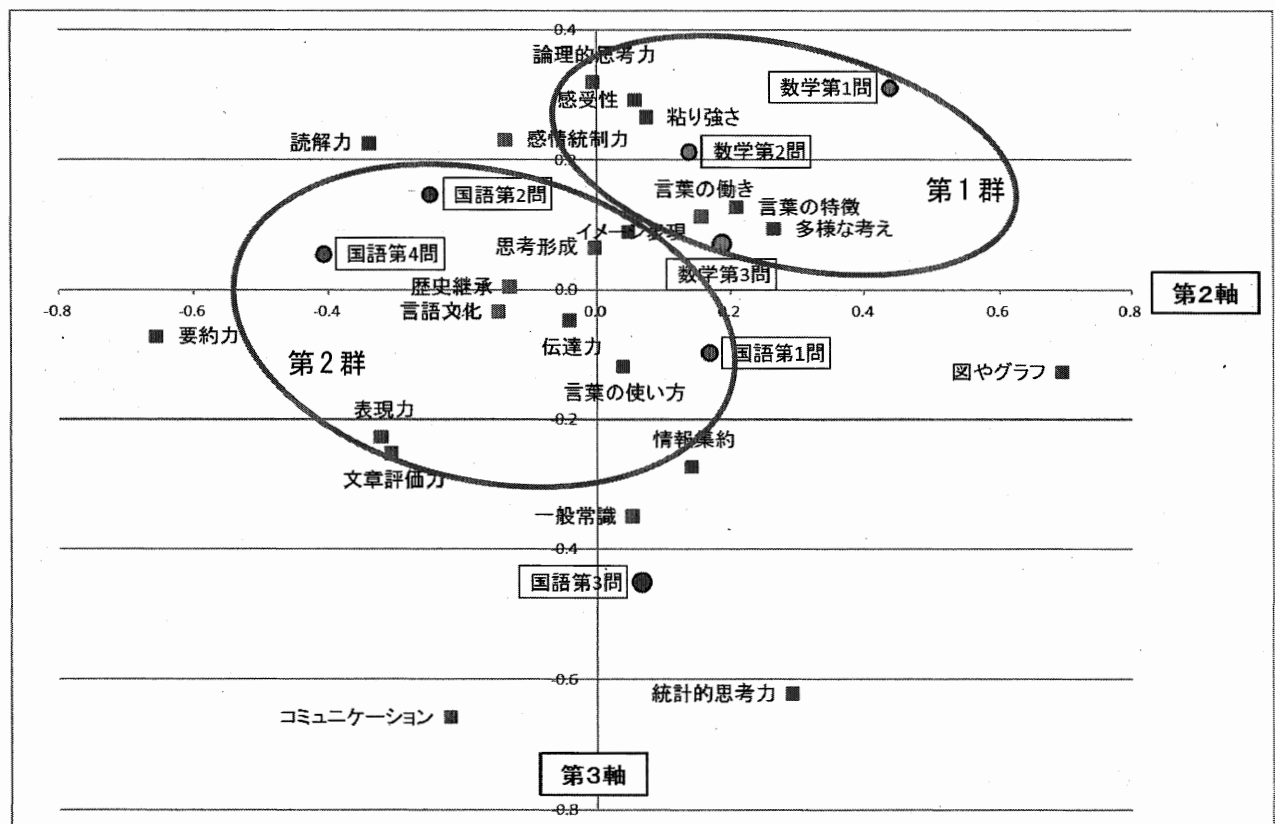


図4 資質・能力と各問題への印象のマッピング（2-3軸）

4. 考察

本研究の分析から、各試験問題は高校生によって次のように受けとめられていることがわかった。

まず、資質・能力の項目を主に国語で必要となる要素から作成したにもかかわらず、数学の問題でもある程度測られていると判断された。数学第3問(イメージ例)は、これまでにない新しい奇抜な問題として映った一方で、資質・能力の観点からは、他の問題とそれほど異なったものとは認知されておらず、しいて言えば国語の問題を解くのに必要な資質・能力と近いものとして捉えられていた。このことは、数学でも国語で培おうとする資質・能力を使わないと解けない問題を設定することが可能であるということを示す一方で、数学として出された問題が実は既存の「数学」の問題に比べ相対的に国語的な能力が必要だと認識されてしまっているということの意味する。

国語の第3問(イメージ例)への印象は若干の奇抜さはあるものの既存の問題とほぼ同様である。ただ、資質・能力との関係でみると、どうやら既存の国語とは少し異なった問題として映ったようである。具体的には、論理的思考力ではなく統計的思考力と近接していることを考えると、この問題もやはり従来の「国語」を解くのに必要な資質・能力とは異なった資質・能力が必要な問題であると認知されていた。さらに論理的思考力とは離れた場所に布置されているのも興味深い。論理的思考力はむしろ数学と近かった。第3問(イメージ例)の内容が統計的であったため統計的思考力と近い。従来の「国語」の問題の方がより国語で測ろうとする資質・能力が試されていると高校生に認識されたことは興味深い点である。

一方、国語第1問(語彙問題)は項目反応理論(IRT)のモデルに典型的に適合する、問題文の存在しない単純な構造の5肢選択の設問であった。それでも、必ずしも高校生からネガティブな印象を持たれたわけではない。むしろ、言語文化、歴史継承、言葉の特徴、言葉の働き、言葉の使い方といった資質・能力を測るのに適した設問と評価されていた。受検者の視点に立てば、どのような形式の問題であっても、単なる知識や記憶力を超えた何かの資質・能力を測定する問題として機能する可能性があるかと判断されたことになる。この結果は、テストを出題する側にとって、試験問題の形式の多様性を確保するうえで参考になる。

しかし、本研究に課題が無いわけではない。数学はいずれの問題もイメージ例に合わせ三角関数を使用した問題を用意したが、国語の問題への印象は問題の内容に左右された可能性が否めない。第1問を除けば、

第2問(センター試験)は論説文、第3問(イメージ例)は統計を使った説明文の解釈、第4問(東北大個別試験問題)は小説文であった。傾向として、論説文は国語の問題としてはオーソドックスであるが、難しいと捉えられがちであり、統計を使った説明文についてもどちらかと言えば論説文に近いものである。一方第4問の小説文は読み物としておもしろいという印象を持ちやすい。さらに、国語という教科の中に含まれる古典に関する設問がなかったことも結果に影響していることが考えられる。

また、試験のできが印象に左右していることも考えられるため、成績との関連や解くのにかかった時間などとの関係にも目を配る必要がある。指導する側(教員)の印象も分析に加えてみたいが、これらの点については引き続き今後の課題としたい。

注

- 1) 本調査は東北大学大学院教育情報学研究倫理審査委員会の承認を受けて実施したものである(承認番号:教情研倫第16-005)。
- 2) 他の調査は大学入学者選抜改革推進委託事業「問題の出題形式と測定する資質・能力の関係に関するテスト理論・測定学に基づく分析」(代表:北海道大学)の一環として行なわれている。

文献

- Alderson, J. & Wall, D. (1993). Does Washback Exist?, *Applied Linguistics*, 14, 115-129.
- 中央教育審議会答申(2014).「新しい時代にふさわしい高大接続の実現に向けた高等学校教育, 大学教育, 大学入学者選抜の一体的改革について」
<http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/tou shin/_icsFiles/fieldfile/2015/01/14/1354191.pdf> (2017年4月20日アクセス)
- Kane, M.T. (2006). Validation, In R. L. Brennan (ed.), *Educational Measurement 4th edition*, Praeger Publishers, 17-64.
- 高大接続システム改革会議(2016).「最終報告」<http://www.mext.go.jp/component/b_menu/shingi/toushin/_icsFiles/fieldfile/2016/06/02/1369232_01_2.pdf> (2017年4月20日アクセス)
- 文部科学省(2016).「高大接続改革の進捗状況について1」
<http://www.mext.go.jp/b_menu/houdou/28/08/1376777.htm> (2017年4月12日アクセス)
- 大山正・岩脇三良・宮埜壽夫(2005).『心理学研究法—データ収集・分析から論文作成まで』サイエンス社
- 平直樹・前川眞一・小野博・林部英雄・内田照久(1998).「日本語基礎能力テストの項目プールの作成」独立行政法人大学入試センター『研究紀要』28, 1-12.

謝辞

本研究は JSPS 科研費 JP16H02051 の助成を受けた。