

新共通テスト（イメージ例）が測定する資質・能力の分析（2）

——センター試験及び個別学力試験の遂行状況との比較から——

宮本 友弘（東北大学）、田中 光晴（文部科学省）、庄司 強（東北大学）

本研究の目的は、新共通テストにおいて出題が予定される、国語の新しい記述式問題（「大学入学希望者学力評価テスト（仮称）」のイメージ例）が測定する資質・能力について実証的に検討することであった。従来のセンター試験、国立大学の個別学力試験を比較対象にして、1,130 名の高校生に調査を実施した。遂行状況の分析から、イメージ例は、他の問題とは異なる資質・能力を測定していることが示唆された。また、高校生は、すべての問題の成績が良好（オールマイティ）、イメージ例が不得意、イメージ例が得意、といった 3 タイプに分かれる可能性が見出された。

1 問題

2020 年度から実施される「大学入学共通テスト」（以下、新共通テスト）では、国語において記述式問題が出題される。その背景には、図 1 のような発想がある。この図は、高大接続システム改革会議の「最終報告」（2016）に示されたもので、「知識・技能」「思考力・判断力・表現力」とそれら进行评估する方法のイメージ例とされる。そこでは、「採点可能性」と「評価の対象となる能力」という 2 次元上に、多肢選択式から小論文までの多様な回答形式が布置されている。そして、従来のセンター試験は、主に多肢選択式で構成されているため、「思考力・判断力・表現力」が十分に評価されていないこと、一方、新共通テスト（当時

は「大学入学希望者学力評価テスト（仮称）」においては、条件付きではあるが、記述式問題を導入することによって、「思考力・判断力・表現力」の評価を補完できることを示唆している。

また、新共通テストの国語の記述式問題では、「複数の情報を統合し構造化して新しい考えをまとめる能力やその過程や結果を表現する能力」（高大接続システム改革会議、2016）を適切に評価することが求められている。この能力を測定するために、題材には、従来の評論や小説だけでなく、日常生活でみかける文章、会話文、図表などが複数使用され、設問では、他者の考えを推論させたり、比較させたりする等、工夫が重ねられてきた。その結果、「大学入学希望者学力評価

テスト（仮称）」のイメージ例（2015 年）、「大学入学共通テスト（仮称）」記述式問題のモデル問題例（2017 年）と、これまでにない新しい記述式問題が公表されるに至った。

本研究では、こうした新しい記述式問題の、テストとしての性能について、テストを受ける側、すなわち、高校生の遂行状況や反応に基づいて分析することを目的にしている。なぜなら、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）を示すにすぎないからである（田中・倉



図 1 「知識・技能」「思考力・判断力・表現力」とそれら进行评估する方法のイメージ例（高大接続システム改革会議、2016 より）

元・宮本, 2018)。

これまで、新しい記述式問題のプロトタイプともいえる、「大学入学希望者学力評価テスト(仮称)」のイメージ例を取り上げ、高校生の遂行状況や反応を調査した(宮本・倉元・田中, 2017; 田中・宮本・倉元, 2018)。その際、図 1 の対角線上に布置された、従来の大学入試センター試験(以下、センター試験)、国立大学の個別学力試験(以下、個別試験)を比較対象とした。主な結果は次の通りである。

①平均得点率は、センター試験、イメージ例、個別試験の順で高く、図 1 が想定する布置に符合する可能性が示唆された。

②センター試験、イメージ例、個別試験はいずれも、SD法(Semantic Differential method)による印象の「知識・能力」(付表 1 参照)では同程度のポジティブな印象を与え、また、解答に必要な資質・能力のうち「深く学ぶ力」(付表 2 参照)においても、同程度に必要であると認知されていた。

③イメージ例は、他よりも SD 法による印象の「新奇性」(付表 1 参照)が強い印象を与え、解答に必要な資質・能力のうち「情報活用力」(付表 2 参照)が必要であると認知されていた。

以上を踏まえ、本研究では、先の調査データのうち、十分に検討されてこなかった各問題の遂行状況に焦点を当て、3つの問題を比較検討し、イメージ例が測定する資質・能力について考察する。

2 方法

2.1 調査参加者

調査協力校は、所在地域を代表する進学校 5 校であった。総計 1,130 名の 2 年生が調査に参加した。参加者の性別と文系・理系別の人数は、表 1 の通りであった。なお、分析では記入漏れにより有効データ数に変動があった。

2.2 調査内容

2.2.1 試験問題

試験問題は、表 2 に示した 3 つの大問から構成された。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を専門とする教員 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ほか(1998)が開発した語彙テストから IRT の項目パラメタに基づき 20 項目を選定し、第 1 問とした(解答は五肢択一式)。

2.2.2 質問紙

試験の遂行状況及び試験問題に対する認知・評価を尋ねるもので、主に次の内容から構成された。

①解答の遂行状況 解答時間の評価(「1:短い」「2:やや短い」「3:ちょうどよい」「4:やや長い」「5:長い」の 5 段階評定)、各大問の難易度の評価(「1:易しい」「2:やや易しい」「3:普通」「4:やや難しい」「5:難しい」の 5 段階評定)と解答の所要時間(分)。

②各問題に対する印象(付表 1) SD 法による評定尺度を 12 項目用意した。-2~+2 の 5 段階評定とした。

③各問題の解答に必要な資質・能力(付表 2)「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(p.25)を参考に 22 項目を作成した。大問ごとに各資質・能力が、解答にあたって必要かどうかを判断させた。

2.2.3 手続き

調査はクラスごとに実施された。試験の制限時間は 80 分であった¹。試験中には、各大問の解答に要した時間を問題用紙に記録するよう指示した。試験終了後、質問紙に回答してもらった(10~15 分)。

なお、本調査は、東北大学教育情報学研究所研究倫理審査委員会の承認を受けて実施した。

表 1 分析対象者の性別と文系・理系別の人数(人)

	文系	理系	その他	計
男子	188	389	7	584
女子	268	265	1	534
無回答	1	3	8	12
計	457	657	16	1,130

表 2 使用された試験問題

出題形式	出典	問題の概要
選択式	平成 22 年度大学入試センター試験本試験問題	評論問題 題材:岩井克人「資本主義と『人間』」
新しい記述式	「大学入学希望者学力評価テスト(仮称)」の記述式問題イメージ例	問題イメージ<1> 題材:警視庁事故統計資料
従来の記述式	平成 23 年度東北大学前期日程試験問題	小説問題 題材:小川洋子「キリコさんの失敗」

表 3 解答時間(80 分)に対する評価(%)(N=1,120)

短い	やや短い	ちょうどよい	やや長い	長い
1.6	8.1	46.9	30.4	13.0

3 結果・考察

3.1 試験の遂行状況

3.1.1 解答時間

全体の解答時間については、表3の通り、「短い」「やや短い」は1割未満であることから、80分という設定は、調査参加者が全問題を解答するにはあたっては、おおむね適切であったと判断できる。

3.1.2 試験の難易度と所要時間

表4は、各問題の難易度及び所要時間の平均と標準偏差を性別と文系・理系別に示したものである。それぞれについて、性別×文系・理系×問題による3要因の分散分析を行った。

まず、難易度では、試験問題要因の主効果が有意であり、効果量も大きかった($F(2, 2124)=259.39, p<.01, \eta_p^2=.196$)²。多重比較(Bonferroni, $p<.05$, 以下同)によれば、イメージ例<個別試験<センター試験であった(図2)。また、性別要因の主効果も有意であったが、効果量はきわめて小さかった($F(1, 1062)=8.84, p<.01, \eta_p^2=.008$)。

一方、所要時間においても、試験問題要因の主効果が有意であり、効果量も大きかった($F(2, 1940)=809.29, p<.01, \eta_p^2=.455$)。多重比較によれば、イメージ例<センター試験<個別試験であった(図3)。また、性別要因についても、困難度と同様に、有意であったものの、効果量はきわめて小さかった($F(1, 970)=9.12, p<.01, \eta_p^2=.009$)。

以上の結果から、センター試験は個別試験よりも難しかったようであるが、解答に要した時間は個別試験の方が長かった。イメージ例は、他の2問に比べ、易しくかつ短時間で解答できたようである。こうした傾向には、性別や文系・理系による違いはほとんどないといえる。

3.2 試験の成績

3.2.1 性別・文系・理系別による比較

表5は、各問題の得点率を性別と文系・理系別に示したものである。問題ごとに、性別×文系・理系による2要因の分散分析を行った。その結果、センター試験、イメージ例では有意な効果は見られなかった。個別試験では、性別要因の効果が有意であったが、効果量は小さかった($F(1, 1094)=13.06, \eta_p^2=.012$)。

以上から、各問題の成績には性別や文系・理系の違いによる影響は小さいといえる。上記3.1.2の結果も踏まえ、以後の分析では、性別、文系・理系を込みにして進めることとする。

表4 性別・文系・理系別による各問題の難易度及び所要時間の平均(M)と標準偏差(SD)

			男子		女子	
			文系	理系	文系	理系
難易度	センター試験	M	3.50	3.47	3.58	3.64
		SD	0.98	0.89	0.82	0.80
	イメージ例	M	2.76	2.78	2.83	2.88
		SD	1.02	0.88	0.93	0.81
	個別試験	M	3.25	3.35	3.45	3.46
		SD	0.96	0.88	0.86	0.84
所要時間	センター試験	M	21.50	21.42	21.63	21.87
		SD	7.11	5.77	5.86	5.96
	イメージ例	M	14.48	14.09	14.99	14.75
		SD	5.51	4.79	4.49	4.23
	個別試験	M	24.88	7.28	26.32	26.18
		SD	25.35	6.77	6.60	6.27

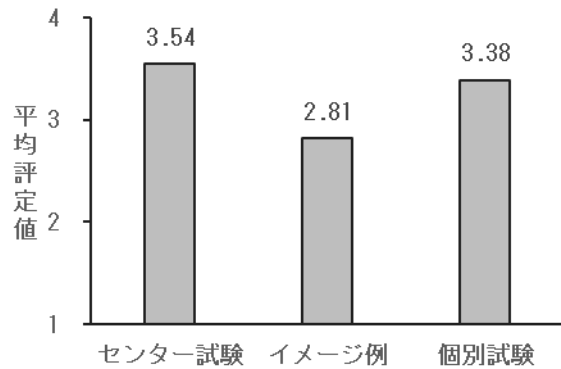


図2 各問題の難易度

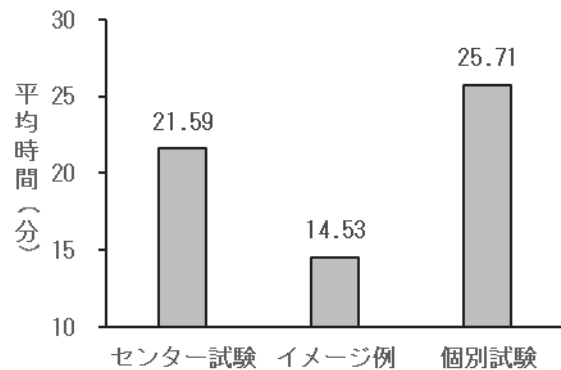


図3 各問題の所要時間

3.2.2 成績と難易度、所要時間の関連

表6は、問題ごとに、得点率、難易度、所要時間の相互の相関を示したものである。

得点率と難易度は、どの問題においても有意な負の相関であったが、相関の強さをみると、個別試験においては非常に弱かった。得点率と所要時間の相関は、

センター試験では有意な負の相関、個別試験では有意な正の相関がみられ、イメージ例では有意な相関はみられなかった。難易度と所要時間は、すべての問題で有意な正の相関であった。

以上の通り、センター試験では、所要時間が短いほど成績が良く、対照的に、個別試験では、所要時間が長いほど成績が良かった。このことから、成績が優秀なものほど、多肢選択式では、迅速で正確な判断を、一方、記述式では、時間をかけじっくり吟味した解答構築を行っているのではないかと考えられる。テストの解答形式によって、そこに要する情報処理プロセスが異なっており、結果として、所要時間に反映されることが推測される。

3.2.3 各問題の成績の関連性

各問題の成績の関連性をみるために、語彙問題も含め、得点率について相互の相関係数を求めた(表7)。

いずれも正の有意な相関であった。相関の強さをみると、語彙、センター試験、個別試験の相互の相関は相対的に強く、一方、イメージ例と他の得点率の相関が相対的に弱い。I-T (項目全体)相関を求めると(表8)、イメージ例が、他の3つの問題に比べて低かった。

以上から、イメージ例が測定している資質・能力は、国語の基礎となる語彙能力や、これまでのセンター試験、個別試験が測定してきた資質・能力とは異質である可能性が示唆された。

3.3 各問題の成績と印象、解答に必要な資質・能力との関連

問題ごとに、宮本・倉元・田中(2017)において構成した、印象の3つの下位尺度の得点、解答に必要な資質・能力の4つの下位尺度の得点を求め、得点率との関連性を調べた。表9の通り、センター試験における得点率と印象の「関心・意欲」及び「知識・能力」の間でみられた正の有意な相関以外は、有意でないか、有意であってもきわめて弱い相関であった。

以上から、問題に対する認知・評価と、問題の成績の関連性はほとんどないことが示唆された。

3.4 各問題の成績による受検者の分類

3つの問題の得点率を使用して、クラスター分析(Ward 法)を行った。デンドログラムから、3クラスター(以下、CL1, CL2, CL3)が妥当と判断した。各クラスターの人数は、CL1が474名、CL2が333名、CL3が286名であった。なお、性別、文系・理系別による人数の偏りは有意ではなかった。

表5 性別・文系・理系別による各問題の得点率の平均(M)と標準偏差(SD)

		男子		女子	
		文系	理系	文系	理系
センター試験	M	58.55	62.35	62.85	62.32
	SD	23.48	21.89	19.62	20.68
イメージ例	M	51.35	52.97	53.26	54.70
	SD	27.16	26.09	25.27	23.76
個別試験	M	49.24	49.41	52.72	52.74
	SD	15.88	15.33	15.13	14.19

表6 各問題の得点率、難易度、所要時間の相関

		得点率	難易度	所要時間
センター試験	得点率	-	-.226**	-.126**
	難易度		-	.281**
イメージ例	得点率	-	-.131**	-.060
	難易度		-	.275**
個別試験	得点率	-	-.064*	.145**
	難易度		-	.251**

** p<.01 * p<.05

表7 各問題の得点率相互の相関係数

	語彙	センター試験	イメージ例	個別試験
語彙	—	.397**	.195**	.329**
センター試験		—	.229**	.325**
イメージ例			—	.252**

** p<.01

表8 I-T 相関の結果

	I-T 相関
語彙	.418**
センター試験	.429**
イメージ例	.292**
個別試験	.413**

** p<.01

表9 各問題の得点率と印象及び解答に必要な資質・能力の下位尺度得点の相関

		センター試験	イメージ例	個別試験
印象	関心・意欲	.267**	.046	.058
	知識・能力	.179**	-.049	.020
	新奇性	-.065*	-.031	-.075*
資質・能力	言語理解力	-.023	.031	.061*
	言語運用力	-.037	.012	.061*
	深く学ぶ力	.078**	.006	.040
	情報活用力	.015	-.028	.028

** p<.01 * p<.05

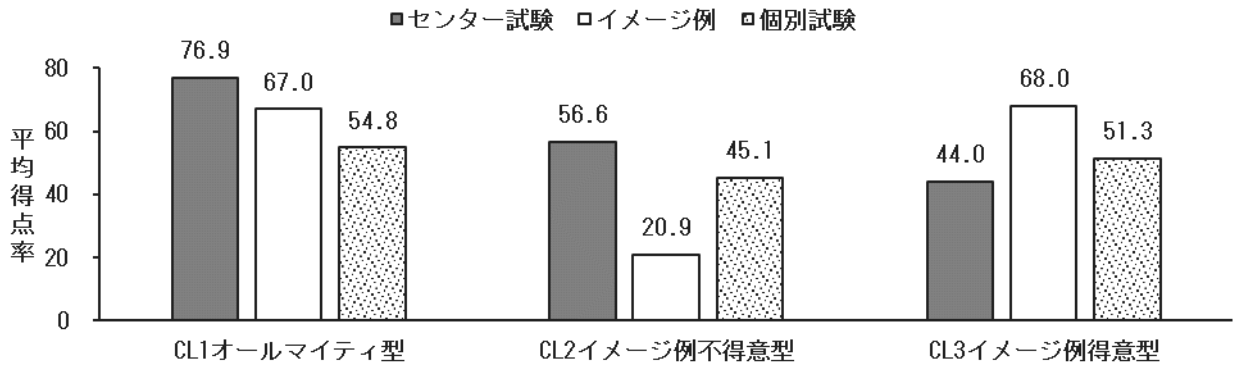


図5 各クラスターの問題別の得点率

各問題の得点率について、クラスターを要因にして1要因の分散分析を行った。センター試験ではクラスター間で有意差が認められ、効果量は大きかった($F(2, 1110)=403.58, p<.01, \eta_p^2=.421$)。多重比較によれば、 $CL3 < CL2 < CL1$ であった(図5)。イメージ例でも、クラスター間で有意差が認められ、効果量も大きかった($F(2, 1110)=1332.31, p<.01, \eta_p^2=.706$)。多重比較によれば、 $CL2 < CL1 = CL3$ であった。個別試験ではクラスター間で有意差は認められたが、効果量は中程度であった($F(2, 1110)=44.22, p<.01, \eta_p^2=.074$)。多重比較によれば、 $CL2 < CL3 < CL1$ であった。

以上から、CL1は、3つの問題の成績がすべて良好である「オールマイティ型」、CL2は、イメージ例の成績が著しく低い「イメージ例不得意型」、それとは対照的に、CL3は、イメージ例の成績が良好な「イメージ例得意型」と命名した。

各問題に対する印象、解答に必要な資質・能力の各尺度得点について3群で比較した結果、有意差が認められ、効果量が一定水準以上であったものは、センター試験に対する印象の「関心・意欲」のみであった($F(2, 1094)=13.53, p<.01, \eta_p^2=.024$)。多重比較によれば、イメージ例不得意型=イメージ例得意型<オールマイティ型であった。この結果は、表9において、センター試験に対する「関心・意欲」と得点率の間に正の有意な相関が見られたことと符合する。

4 まとめ

本研究の結果から示唆されることをまとめると次の通りとなる。

①問題に関わらず難易度の認知、解答に要する時間、得点率に対しては、性別や文系・理系別はほとんど影響しない。

②宮本・倉元・田中(2017)によれば、イメージ例は、他の2問よりも新奇性が高い印象であったが、そうした新奇性は、解答に必要な時間を延長させたり、問題を難しいと感じさせたりするわけではない。

③選択式と記述式では、解答に費やす情報処理プロセスが異なり、それらは、所要時間に反映される可能性がある。

④得点率の相関から、イメージ例が測定する資質・能力は、センター試験、個別試験が測定する資質・能力とは異質である。

⑤問題に対する印象や解答に必要な能力・資質の認知と、実際の成績との間には関連がほとんどない。

⑥各問題の得手不得手から、高校生には、「オールマイティ型」「イメージ例不得意型」「イメージ例得意型」の3タイプが存在する。

以上から、今回取り上げた3つの問題が測定する資質・能力には違いがあり、とくに、新しい記述式であるイメージ例は、これまでの記述式とは異なる資質・能力を測定していると考えられる。また、そうした新しい資質・能力とこれまでの資質・能力を、バランスよく獲得している者と、偏りがある者がいるようである。

今後の課題としては、各問題が測定する能力と個人差をもたらす要因を明確にすることである。また、素材文の内容、すなわち、評論問題か、小説問題かによる影響についても検討する必要がある。

注

1 試験時間は、次のようにして設定した。①センター試験：出典が大問4題で構成され、試験時間80分であり、また、各大問の配点と同じであることを勘案して、試験時間を大問数で割り、大問1問の所

要時間を 20 分と見積もり、そのまま割り当てた。
 ②個別試験：上記①と同様に、出典が大問 4 題、120 分なので、30 分とした。③イメージ例：小問 2 題で構成されており、個別試験の小問 5 題よりも少ないことから 20 分とした。④語彙問題：1 項目の所要時間を 30 秒と見積もり、20 項目構成であることから 10 分とした。以上から、総計 80 分とした。

2 山際・服部(2016)によれば、偏イータ 2 乗(η_p^2)は次のように解釈される。

- .01～.06 小さい効果
- .06～.14 中程度の効果
- .14～ 大きい効果

付記

本研究は、大学入学者選抜改革推進委託事業「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」(北海道大学(代表大学)、東北大学、九州大学、長崎大学、大学入試センター)の平成 28 年度の取り組みに基づくものである。なお、調査協力校のうち 1 校については、予備調査として JSPS 科研費 JP16H02051 の助成を受けて実施した。

本研究の実施にあたっては、伊藤博美氏にご協力いただいた。心より感謝申し上げます。また、本調査にご協力いただいた高校の皆様にご礼申し上げます。

参考文献

高大接続システム改革会議 (2016). 高大接続システム改革会議「最終報告」 文部科学省 2016年 3月 11日
 <http://www.mext.go.jp/component/b_menu/shingi/toushin/_icsFiles/afieldfile/2016/06/02/1369232_01_2.pdf> (2018年 8月 31日)

宮本友弘・倉元直樹・田中光晴 (2017). 「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」『日本テスト学会第 15回大会発表論文抄録集』, 130 - 133.

文部科学省 (2016). 高大接続改革の進捗状況について 文部科学省平成28年 8月 31日
 <http://www.mext.go.jp/b_menu/houdou/28/08/1376777.htm> (2018年 8月 31日)

平直樹・前川眞一・小野博・林部英雄・内田照 (1998). 「日本語基礎能力テストの項目プールの作成」『大学入試センター研究紀要』, 28, 1-12.

田中光晴・宮本友弘・倉元直樹 (2018). 「新共通テスト(イメージ例)が測定する資質・能力の分析—高

校生対象のモニター調査から—」『大学入試研究ジャーナル』, 28, 1 - 6.

山際勇一郎・服部環(2016). 『文系のための SPSS データ解析』ナカニシヤ出版

付表 1 各問題に対する印象評定項目 (因子別)

[関心・意欲] 2 嫌いな—好きな / 8 解きたくない—解きたい / 7 むりそう—できそう / 1 つまらない—おもしろい / 6 意地悪な—素直な
[知識・能力] 12 無意味な—有意義な / 11 実力がわからない—実力がわかる / 10 役に立たない—役に立つ / 3 ふざけた—まじめな / 9 下品な—上品な
[新奇性] 5 R 典型的な—奇抜な / 4 古い—新しい

注)数字は質問紙上の番号。R は逆転項目。

付表 2 各問題の解答に必要な資質・能力 (因子別)

[言語理解力] 1 言葉の働きや役割に関する理解 / 2 言葉の特徴や決まりに関する理解 / 3 言葉の使い方に関する理解 / 4 言語文化に関する理解 / 5 一般常識や社会規範 / 13 歴史の中で創造され継承されてきた言語文化の担い手としての自覚
[言語運用力] 18 表現力 / 7 自分の文の構成や表現形式を振り返り評価する力 / 17 要約力 / 19 コミュニケーション力 / 10 言葉を通じて伝える力
[深く学ぶ力] 14 粘り強く、柔軟に考え考察を深めていく態度 / 21 論理的思考力 / 16 読解力 / 11 考えを形成し深める力 / 8 言葉によって感じたり想像したりする力 / 12 自分の感情をコントロールして学びに向かう態度
[情報活用能力] 22 図やグラフを描いたり読んだりする力 / 20 統計的思考力 / 6 情報を様々な見方でよく調べ、まとめる力

注)数字は質問紙上の番号。2 項目は複数の因子に高い負荷量を示したので除外。