

英語の基礎学力を担保する総合試験の研究（総括）

——令和3年度「志」特別選抜の総合試験に類似したサンプル問題を代用して——

吉田 健三（神戸大学）

本稿は、「志」特別選抜（神戸大学の総合型選抜）・第1次選抜における英語力評価を主とする総合試験の評価を統計的手法によって客観的、間接的に実証することを主な目的とした。類似したサンプル問題を代用し、テストの妥当性、信頼性、基礎学力の担保の有無を検証した。2021年、2022年の複数年のデータの多面的な分析を通して、構成概念妥当性、信頼性、意図した学力の差異を弁別する識別力、それぞれの高さが示唆された。研究協力者を対象としたアンケート調査からは、基礎学力の担保や早期実施に対する配慮に関して肯定的な回答が得られた。

キーワード：総合型選抜、テストの妥当性・信頼性、基礎学力の担保

1 はじめに

本研究は、2018年度にスタートさせた「志」特別入試（2020年度より「志」特別選抜）における英語の基礎学力を担保する総合試験（以下「志」）の妥当性、信頼性を検証し、学内に残る一部の懸念を払拭し、「志」特別選抜をさらに発展させ、大学入試改革の進展を主な目的としている。さらに、本稿での分析や考察は、過去の入学者選抜試験の分析を基に新たな問題を作成する他大学における出題検討にも資することをねらいとする。「志」で使用した英文や日本語は、著作物使用許諾が未完了のため、それと類似したサンプル問題を代用する必要があり、両者の類似性は、設問構成の対照に加えて、統計的に示す手法を用い、後者の弁別妥当性、信頼性、基礎学力の担保に関する検証を通じて、前者を間接的に検証した。

中間報告（吉田、2023）では、2021年度の分析結果を報告した。学内で募った研究協力者に、「志」に類似したサンプル問題を解答させ、解答後に設問に関するアンケート調査に回答させてデータを取得し、分析を行ったが、特に以下の点については、さらなる検討が必要であった。(1)「志」とサンプル問題との類似性の統計的な確証、(2)サンプル問題の解答データの因子分析。本稿は、分析精度を高めるため2022年のデータを追加した上で、これらの課題を再考察し、新たに項目分析を含め、総合的かつ多面的に分析を行い、総括した最終報告である。なお、本稿での2021年度のデータは、中間報告で既出のものである。

2 サンプル問題と「志」の類似性の分析

サンプル問題は、出題者（筆者）の経験値に基づいて、令和3年度「志」（文系型・理系型）に類似した

2種類（A・B）を作成した。設問構成の類似性については、既に明らかにした（吉田、2023; 258-260）が、ここでは、英文の特徴に関する類似性を統計的に検証した。

2.1 方法

異なった英文の類似性を検証するには、標準的なリーダビリティ指標以外の指標のデータも活用する必要があると判断した。「従来の標準的なリーダビリティ指標は、単語の長さや文の長さに依存して難易度を評価するが、Coh-Metrixは、文や語句の結束性、文体や論理性、言語および談話の特性など、より多面的な特性を精緻に評価する（筆者訳）」（Graesser, A.C., et al, 2004; 193）ことが可能であり、本研究ではCoh-Metrixで算出できる12の指標を活用した。水本（2013）の手法を参考に、下記(1)～(5)の手順で検証した。大問項目名付きの樹形図の作成が容易なためエクセル統計（BellCurve for Excel）を用いた。

(1) 12指標（①Total Words ②Words per Sentence ③Flesch-Kincaid Grade Level ④Narrativity ⑤Syntactic Simplicity ⑥Word Concreteness ⑦Referential Cohesion ⑧Deep Cohesion ⑨Verb Cohesion ⑩Connectivity ⑪Temporality ⑫Type-Token Ratio）のデータを算出する。算出プログラムとしてCohmetrix3 desktop program（The University of Memphisより提供）を用い、各大問の英文単位で算出した。水本（2013）では、Average Word Level (SVL) / Average Word Level (JACET) を含めているが、算出プログラム Word Level Checker にアクセスできない（2023年11月23日現在）ため本稿では除き、上記の12指標を算出した。

12 指標の内容は次の通りである (水本, 2013; 144)。①総語数 ②一文に含まれる平均語数 ③一文に含まれる平均語数と一語の平均音節数を基に計算 ④物語や会話調で読み手が親しみのある英文かどうかを示す ⑤文体が簡潔でわかりやすいかを示す ⑥語が具体的にイメージしやすいかを示す ⑦文や句の結束性が英文全体でどの程度あるかを示す ⑧因果・論理関係を示す接続語がどの程度含まれているかを示す ⑨同じ動詞がどの程度繰り返し使われているかを示す ⑩反対, 追加, 比較などの関係がどの程度明確にわかるかを示す ⑪時制などが文中でどの程度同じように使われているのかを示す ⑫総語数に対する異なり語の数で, 1 に近くなるほど異なり語が多く難しい。

(2) 12 指標のデータについて主成分分析を行う。

入力データは分散 1 に基準化した。

(3) 第 1 主成分, 第 2 主成分のいずれの主成分負荷量も 0.3 以下の指標を除外し, 6 指標 (Words per Sentence, Flesch-Kincaid Grade Level, Word Concreteness, Referential Cohesion, Deep Cohesion, Temporality) を選定する。

1 回目の主成分分析の結果, 累積寄与率 70% 以上および固有値 1 以上の主成分が 3 つ算出されたが, 第 2 主成分と第 3 主成分において Temporality の主成分負荷量が 0.3 以上で重複した。「志」とサンプル問題との類似性について, それぞれの英文の特徴を明確に示すため, 第 3 主成分を除いた。

(4) 6 指標のデータについて主成分分析を行う。

上記(3)の第 1 主成分, 第 2 主成分のいずれの主成分負荷量も 0.3 以下の指標を除外し, 再度, 主成分分析を行った。

(5) 固有値 1 以上の 2 つの主成分 (累積寄与率 77.48%) の主成分得点についてクラスタ分析を行う。

2.2 結果

6 指標の主成分分析の結果を図 1 に示した。第 1 主成分 (以下「第 1」) は, 一文や語彙の難易度 (語の具体性, 一文に含まれる平均語数, 一語の平均音節数, 文や句の結束性) を表し, 第 2 主成分 (以下「第 2」) は, 文章の一貫性・整合性 (因果・論理関係を示す接続語, 同じ時制などの使用程度) と推察できる。

クラスタ分析により, 3 つのクラスタが示された。サ A と志文の大問 2, 大問 3 以外は各設問の英文において高い類似性が認められた (図 2)。各クラスタ主成分の平均値 (図 3) により, クラスタ 1 の第 1, 第 2 とともに平均的である。クラスタ 2 の第 1 は難しく, 第 2 はやや強い。クラスタ 3 の第 1 は易しく, 第 2 は

平均的である (図 3, 表 1)。以上の結果より, 「志」とサンプル問題の英文の特徴については, 類似性が概ね高いと考えられる。

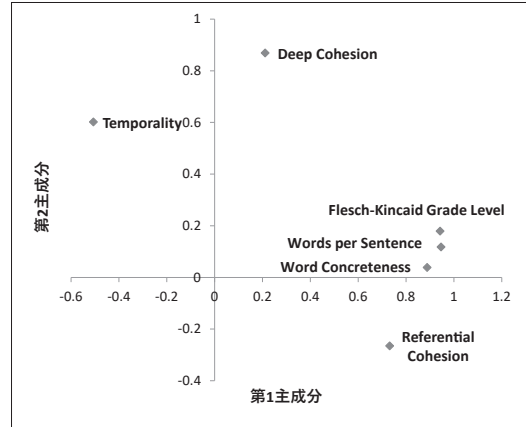


図 1 6 指標の主成分分析の結果 (累積寄与率 77.48%)

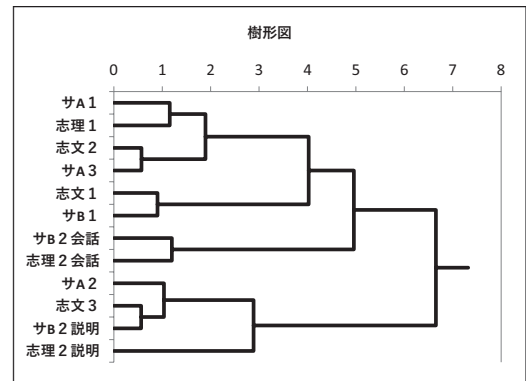


図 2 クラスタ分析の結果 (ワード法・ユークリッド距離)

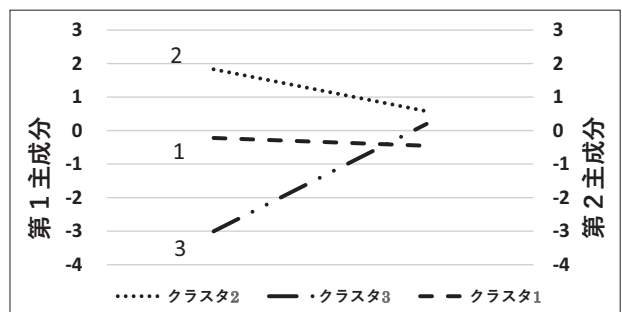


図 3 各クラスタの平均値のプロット

表 1 各クラスタの第 1・第 2 主成分と設問

クラスタ	第 1 主成分	第 2 主成分	サンプル問題と「志」			
	一文や語彙の難易度	文章の一貫性・整合性	サ A	志文	サ B	志理
クラスタ 1	平均的	平均的	大問 1 大問 3	大問 1 大問 2	大問 1	大問 1
クラスタ 2	難しい	やや強い	大問 2	大問 3	大問 2 説明	大問 2 説明
クラスタ 3	易しい	平均的			大問 2 会話	大問 2 会話

3 サンプル問題の分析

3.1 研究協力者

学内での公募による2021年、2022年の研究協力者(1年生)は、それぞれ97名(文系50名、理系47名)、89名(文系53名、理系36名)であった(表2)。

表2 研究協力者の学部別内訳(人)

2021			
文系		理系	
文	7	工	9
国際人間	19	国際人間	0
法	8	理	8
経済	2	医	17
経営	13	農	11
海洋政策	1	海洋政策	2
計	50	計	47

2022			
文系		理系	
文	5	工	8
国際人間	18	国際人間	3
法	4	理	4
経済	7	医	5
経営	19	農	14
海洋政策	0	海洋政策	2
計	53	計	36

注) 国際人間、海洋政策には文系・理系あり

3.2 各テストの記述統計量

2021年と2022年は同じサンプル問題を活用した。分析結果を表3、表4に示す(2021年のデータは、吉田(2023)で既出)。

サンプル問題A・B、共通テストR・L・総合、TOEIC L&R IP(以下TOEIC)、TOEFL ITP(以下TOEFL)の平均値、標準偏差、得点分布は表3、表4、図4、図5の通りである。共通テスト、TOEIC、TOEFLの数値は研究協力者の自己申告による。TOEIC、TOEFLは入学時に大学が実施するプレイスメントテストで、TOEIC(990点満点)、TOEFL(310~677)のスコアである。所属する学部によって受験するテストが指定されている。TOEIC、TOEFLには合計以外に各技能別スコアがあるが、不明なものが一部あるため、総合スコアのみを用いた。各数値は、統計ソフトSPSSを用いて算出した(以下、同じ)。

表3 各テストの記述統計量(2021)

テストの種類	度数	平均値	標準偏差
サンプルA(100点満点換算)	97	50.1	12.1
サンプルB(100点満点換算)	97	44.3	13.5
共通テストR(100点満点)	97	87.5	8.7*
共通テストL(100点満点)	97	75.5	8.9
共通テスト総合(200点満点)	97	163.1	13.7
TOEIC L&R IP(990点満点)	41	574.0	103.9
TOEFL ITP(310~677)	53	499.0	46.4

* 共通テストRの標準偏差について、数値8.6(吉田, 2023; 260)を訂正

表4 各テストの記述統計量(2022)

テストの種類	度数	平均値	標準偏差
サンプルA(100点満点換算)	89	51.9	11.2
サンプルB(100点満点換算)	89	47.4	13.1
共通テストR(100点満点)	87*	88.7	6.8
共通テストL(100点満点)	87*	83.2	10.8
共通テスト総合(200点満点)	87*	172.0	14.9
TOEIC L&R IP(990点満点)	45	609.7	119.1
TOEFL ITP(310~677)	44	513.8	36.6

* 度数87は、共通テストを課さない入学者選抜の合格者2名を除く

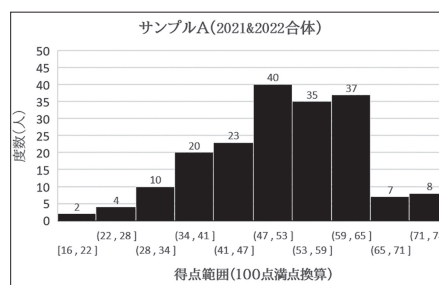


図4 サンプル問題Aの得点分布

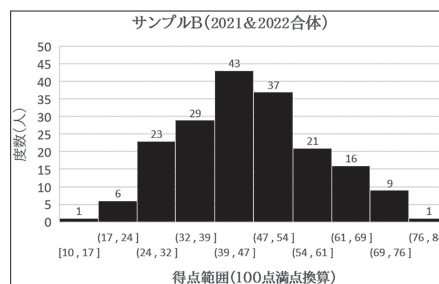


図5 サンプル問題Bの得点分布

3.3 構成概念の分析

テストが測りたい対象を出題のねらい通りに測れているかを検証するため、各テスト間の相関を調べた(表5~表7)。英語民間試験は、受験するテストが異なっても同一価値を評価するテストであるため2021年と2022年のデータを合体して算出した。共通テストは、同一価値を評価するテストではないため、2021年と2022年のデータごとに算出した。

3.3.1 弁別的証拠の確認

異なった構成概念を測定しているテスト間の関係を見るために、相関分析を行った。相関が弱い場合は、妥当性が高いことが示唆される(野口ほか, 2014; 44)。

表5 サンプル問題と英語民間試験の相関
(2021&2022 合体)

	A	B	TOEIC	TOEFL
A	1	.562**	.253*	.282**
B	.562**	1	.333**	.341**
TOEIC	.253*	.333**	1	...
TOEFL	.282**	.341**	...	1

** 相関係数は 1% 水準で有意 (両側)

* 相関係数は 5% 水準で有意 (両側)

表6 サンプル問題と共通テストの相関 (2021)

	A	B	共通R	共通L	共通総合
A	1	.605**	.201*	.195	.254*
B	.605**	1	.190	.168	.226*
共通R	.201*	.190	1	.213*	.770**
共通L	.195	.168	.213*	1	.786**
共通総合	.254*	.226*	.770**	.786**	1

** 相関係数は 1% 水準で有意 (両側)

* 相関係数は 5% 水準で有意 (両側)

表7 サンプル問題と共通テストの相関 (2022)

	A	B	共通R	共通L	共通総合
A	1	.502**	.335**	.329**	.391**
B	.502**	1	.216*	.423**	.404**
共通R	.335**	.216*	1	.405**	.750**
共通L	.329**	.423**	.405**	1	.909**
共通総合	.391**	.404**	.750**	.909**	1

** 相関係数は 1% 水準で有意 (両側)

* 相関係数は 5% 水準で有意 (両側)

表5～表7の結果より、各テスト間で次の(1)～(3)の関係がみられた。(1) サンプル問題と英語民間試験の間には、弱い正の相関が認められた ($r=.253 / .333, .282 / .341$)。サンプル問題と共通テストの間において、2021年と2022年で共通した相関関係は次のとおりである。(2) サンプル問題A、共通テストRの間には、弱い正の相関が認められた ($r=.201 / .335$)。(3) サンプル問題A、Bと共通テスト総合の間には、弱い正の相関、あるいは正の相関が認められた ($r=.254 / .226, .391 / .404$)。共通テストの得点率は75～88%と高く、切断効果の影響も想定されるが、構成概念が異なるテスト間の相関は全般的に弱いことが認められたと考えられる。

3.3.2 A・B問題の出題構成の類似性の確認

類似の構成概念を測定しているテスト(サンプル問題Aとサンプル問題B)間の相関関係は、表5に示されている。サンプル問題A、サンプル問題Bの間には、正の相関が認められた ($r=.562$)。問題のトピックが異なっても、問題の構成や問いのあり方が得点に影響したことを示唆している。

3.4 因子分析による構成概念の確認

サンプル問題A、Bの共通因子を調べた。Aは18項目の設問、Bは17項目の設問を用いて因子分析を行った。因子の抽出に主因子法を用いて固有値を1以上の基準を設け、さらに因子の解釈の可能性を考慮して、A、Bそれぞれ4因子とした。プロマックス回転を行った結果の因子パターンと共通性をAは表8、表9、Bは表10、表11に示した。因子負荷量がどの因子に対しても0.35に満たなかった項目を除いた項目について共通特性を考察し、因子名を決めた(松尾ほか、2002: 77, 109を参考にした)。

Aについては、[大問3-問1オ、カ:文脈の理解力を基に英語で記述する文章作成能力、大問3-問1ア、ウ、エ:情報を検索する能力と検索した箇所を日本語に変換する能力、大問1-問2、大問1-問3:文脈の理解力を基に日本語で記述する文章作成能力、大問3-問1キ、大問3-問4:情報を検索する能力と検索した内容を日本語で記述する文章作成能力]を測っている、と推察される。

表8 サンプル問題Aのパターン行列

	1	2	3	4
大問1-問1	-0.014	0.201	0.101	-0.102
大問1-問2	0.051	0.037	0.699	-0.113
大問1-問3	-0.034	0.259	0.425	0.145
大問1-問4	0.100	-0.193	0.284	0.127
大問2-問1ア	0.031	-0.028	0.310	0.008
大問2-問1イ	-0.095	0.036	0.244	-0.048
大問2-問1ウ	0.062	0.074	0.062	0.022
大問2-問2	-0.046	0.042	-0.038	0.143
大問2-問3	0.177	0.118	0.143	-0.111
大問3-問1ア	0.050	0.432	-0.036	0.034
大問3-問1イ	0.242	0.171	0.002	0.151
大問3-問1ウ	-0.012	0.602	0.095	-0.036
大問3-問1エ	0.045	0.798	-0.140	0.050
大問3-問1キ*	-0.054	0.289	-0.027	0.409
大問3-問2オ*	0.698	0.085	-0.011	-0.073
大問3-問2カ	0.832	-0.050	-0.051	-0.017
大問3-問3	0.264	-0.028	0.101	0.111
大問3-問4	0.050	-0.158	0.005	0.717

*大問3-問1キ、大問3-問2オの設問名は誤記ではない

表9 サンプル問題Aの共通性

大問	大問1				大問2				
	問1	問2	問3	問4	問1ア	問1イ	問1ウ	問2	問3
初期	0.083	0.233	0.250	0.141	0.117	0.107	0.052	0.086	0.128
因子抽出後	0.047	0.507	0.326	0.134	0.101	0.058	0.022	0.025	0.082
大問	大問3								
	問1ア	問1イ	問1ウ	問1エ	問1キ	問2オ	問2カ	問3	問4
初期	0.215	0.191	0.354	0.430	0.244	0.409	0.408	0.143	0.181
因子抽出後	0.210	0.176	0.366	0.678	0.319	0.493	0.645	0.122	0.473

表 10 サンプル問題 B のパターン行列

	1	2	3	4
大問 1- 問 1	0.247	0.020	0.052	-0.023
大問 1- 問 2 ア	0.551	-0.079	-0.047	-0.088
大問 1- 問 2 イ	0.634	-0.092	0.080	-0.055
大問 1- 問 2 ウ	0.286	-0.018	-0.041	0.107
大問 1- 問 3 エ	0.562	0.104	-0.020	-0.053
大問 1- 問 3 オ	0.395	0.169	-0.093	0.151
大問 1- 問 4	0.270	0.026	0.236	0.254
大問 1- 問 5	0.150	-0.142	0.024	0.200
大問 2- 問 1 ア	-0.004	0.134	-0.296	0.477
大問 2- 問 1 イ	0.106	-0.033	0.044	0.413
大問 2- 問 1 ウ	-0.139	-0.137	0.204	0.503
大問 2- 問 1 エ	-0.146	0.396	-0.004	0.299
大問 2- 問 2 オ	-0.044	0.015	0.645	-0.020
大問 2- 問 2 カ	0.129	0.076	0.635	0.197
大問 2- 問 2 キ	0.099	0.858	-0.023	-0.082
大問 2- 問 2 ク	-0.150	0.411	0.364	-0.143
大問 2- 問 3	0.025	0.448	0.065	-0.074

表 11 サンプル問題 B の共通性

大問	大問 1								
	問 1	問 2 ア	問 2 イ	問 2 ウ	問 3 エ	問 3 オ	問 4	問 5	
初期	0.091	0.230	0.290	0.133	0.302	0.334	0.237	0.113	
因子抽出後	0.056	0.294	0.378	0.125	0.296	0.267	0.255	0.098	
大問	大問 2								
	問 1 ア	問 1 イ	問 1 ウ	問 1 エ	問 2 オ	問 2 カ	問 2 キ	問 2 ク	問 3
初期	0.183	0.172	0.141	0.221	0.314	0.346	0.390	0.392	0.229
因子抽出後	0.304	0.219	0.237	0.283	0.433	0.531	0.691	0.478	0.217

Bについては、[大問 1- 問 2 ア, イ, エ, オ：文脈の理解力を基に日本語で記述する文章作成能力, 大問 2- 問 1 エ, 問 2 キ, ク, 問 3：文脈の理解力を基に要点をまとめる文章作成能力（大問 2- 問 2 キ, ク, 問 3 は日本語で記述し, 大問 2- 問 1 エは英語で記述する文章作成能力）, 大問 2- 問 2 オ, カ, ク：情報を検索する能力と検索した内容を日本語で記述する文章作成能力, 大問 2- 問 1 ア, イ, ウ：情報を検索する能力と検索した内容を英語で記述する文章作成能力]を測っている, と推察される。

サンプル問題 A は「志」総合問題 II（文系型）, B は「志」総合問題 I（理系型）に類似した出題を意識し, 英語・日本語で記述されたそれぞれの文章を読解する力, 思考する力, および英語や日本語での確に表現する力を問うことを意図したが, 全般的に出題の意図はテストの結果に反映していると考えられる。

3.5 項目分析による識別力の確認

各設問の得点率と合計点との相関を分析することによって, 設問が受験者の能力を弁別する識別力があるかを検証した。相関係数が 0.2 以上が望ましいとされ

る（山下, 2022; 243）。サンプル問題 A で一部相関が 0.2 より小さい問題がみられた（大問 2- 問 1 イ, ウ, 問 2）が, A, B 全般において, 受験者の能力を弁別する識別力があることが確認されたと考えられる。

表 12 サンプル問題 A 各設問の得点率と合計点の相関

大問 1		大問 2		大問 3			
問 1	.267**	ア	.245**	ア	.373**		
問 2	.507**	問 1	イ	.107	問 1	イ	.382**
問 3	.507**		ウ	.154*		ウ	.393**
問 4	.338**	問 2		.114		エ	.410**
		問 3		.563**		キ	.320**
					問 2	オ	.548**
						カ	.504**
					問 3		.367**
					問 4		.307**

* 相関係数は 5% 水準で有意（両側）
** 相関係数は 1% 水準で有意（両側）

表 13 サンプル問題 B 各設問の得点率と合計点の相関

大問 1		大問 2			
問 1	.268**	ア	.418**		
	ア	.244**	問 1	イ	.442**
問 2	イ	.355**		ウ	.306**
	ウ	.375**		エ	.425**
問 3	エ	.463**		オ	.274**
	オ	.483**	問 2	カ	.486**
問 4		.513**		キ	.510**
問 5		.210**		ク	.263**
			問 3		.364**

* 相関係数は 5% 水準で有意（両側）
** 相関係数は 1% 水準で有意（両側）

3.6 項目分析による測定能力の確認

各設問の得点率の分布（合計点を 20 点ごとの範囲で 5 区分）を分析することによって, 各設問の意図（測りたい能力）が解答結果に反映されているかを検証した。得点率と合計点の相関係数が大きい設問とは, 合計点の高い受験者は得点率が高く, 合計点の低い受験者は得点率が低いことを示している（山下, 2022; 242）。各設問の得点率の分布から, 意図した学力の差異を弁別する識別力があるかどうかを判別できる。

3.6.1 得点率と合計点の相関係数が大きい設問

共通テストの得点率が高い受験者間でも設問の識別力が大きくなる理由としては, さまざまな情報を検索し, 論理的な思考や分析を通して, 英語あるいは日本語で文章作成するという多面的な能力が問われているからであると考えられる。以下に具体例を示す。

【設問】 A 大問 2- 問 3：英文を読み取り, 英文の内容

に反する自分の意見を、85語～105語の英語で適切に表現する

【意図】英文の読解能力・論理的思考力・英語の文章作成能力を測る

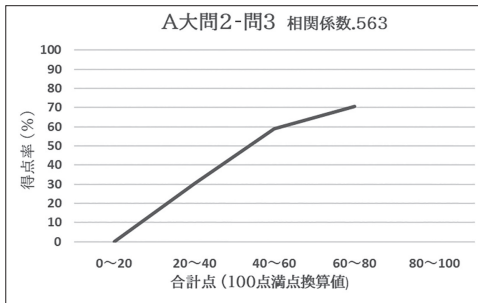


図6 A大問2-問3の得点率と合計点

【設問】A大問3-問2オ：日本語とその内容に関する日本語の会話文の意味を読み取り，文脈に応じた適切な英語で空所を埋める表現を記述する

【意図】日本語の読解能力・論理的思考力・英語の文章作成能力を測る

【該当箇所の会話文】

マスオ: その条例のことを英語のプレゼンテーションでも説明しようと考えているんだ。こんな説明ではどうかな? (英語で) [オ] <45語以内>

サオリ: 「2020年3月に亀岡市がプラスチック製買入物袋の提供禁止を宣言した」という表現で始めているのね。禁止のねらいと、日本で初めての宣言で、結論に達するまでに多くの障壁があったという3点を簡潔にまとめていると思う。

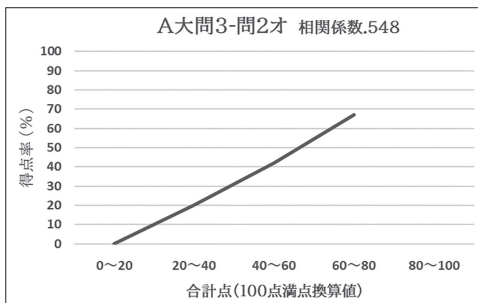


図7 A大問3-問2オの得点率と合計点

3.6.2 得点率と合計点の相関係数が小さい設問

得点率と合計点との相関がない，あるいは弱い理由として(1)～(3)の場合が考えられる。

(1) 正解することが易しく，合計点が低い受験者も得点率が高く，全体的に差異が小さい。

【設問】A大問2-問2：英文の理解にかかわらず，日本語で指示された内容を自分で考えて，日本語で説明する。次に続く問3で英文を書くための予備的な設問となっている

【意図】論理的思考力・日本語の文章作成能力を測る

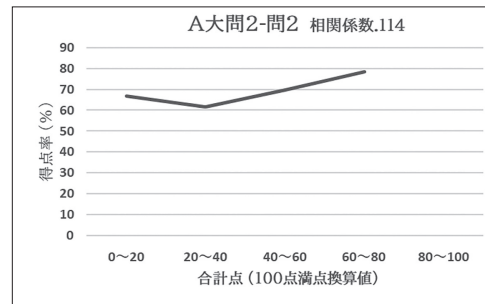


図8 A大問2-問2の得点率と合計点

(2) ある程度までのポイントを正解できるが，高い得点率を得ることができない。

【設問】B大問1-問2ア：英文の理解に基づいて，英文の要点(多くの人々が予測した道路閉鎖後の影響と，実際の閉鎖後の結果)を日本語で説明する

【意図】英文の読解能力・論理的思考力・日本語の文章作成能力を測る

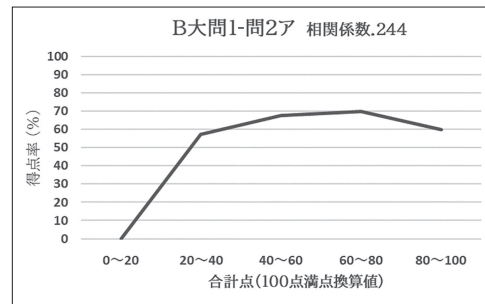


図9 B大問1-問2アの得点率と合計点

【該当箇所の英文】

On Earth Day this year, New York City's Transportation Commissioner decided to close 42nd Street, which as every New Yorker knows is always crowded. "Many predicted it would be a terrible tragedy," said the Commissioner, Lucius J. Riccio. "You didn't need to be a rocket scientist or have a sophisticated computer queuing model* to see that this could have been a major problem."

But to everyone's surprise, Earth Day generated no historic traffic jam. Traffic flow actually improved when 42nd Street was closed.

注 queuing model 待ち行列モデル

(出典: *What if They Closed 42d Street and Nobody Noticed?* by Gina Kolata, The New York Times, published on Dec. 25, 1990 を一部改変)

(3) 正解することが難しく，合計点が高い受験者も得点率が低く全体的に差異が小さい。

【設問】B大問1-問5：英文全体の理解に基づき，英文中の空所に入る適切な内容を3語の英語句で答える

【意図】英文の読解能力・論理的思考力・英語の文章作成能力を測る

【補足】得点範囲 80～100 の得点率が突出しているために傾斜が大きく見えるが度数は 1 であり，得点率は全体的に低い

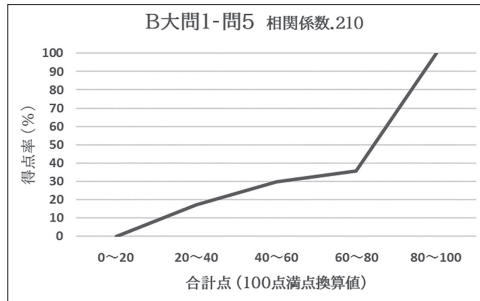


図 10 B 大問 1- 問 5 の得点率と合計点

3.7 信頼性の確認

サンプル問題 A, B, A・B 合体の信頼性を検証するために，Cronbach の α 係数の算出を行った（表 14）。信頼性は，同じテスト内での内的一貫性のことで，その測定手段で試行（trial）を繰り返した場合，どの程度同じ結果を示すかを表す。 α 係数が 0.7 以上が望ましいとされる（山下，2022; 247-248）。

サンプル問題 A は .587，サンプル問題 B は .621 であった。項目数が少ない場合は数値が小さくなる場合がある（野口ほか，2014; 16）と言われ，A・B 合体では .728 を示した。

表 14 サンプル問題 A, B, A・B 合体の信頼性統計量

	Cronbach の α 係数	項目の数
A	.587	18
B	.621	17
A・B 合体	.728	35

4 アンケート調査回答の分析

サンプル問題解答直後に，各設問の難易度や様式に関して 4 肢選択あるいは 5 肢選択で回答するよう依頼した。質問項目は以下の通りで，平均値は，各設問の選択肢の番号を数値として算出した平均である。

【アンケート調査の質問】

- ア．問題の内容は，興味深かった
 - 1. そう思う 2. ややそう思う 3. あまりそう思わない 4. そう思わない
- イ．問題の難易度について
 - 1. 低すぎる 2. やや低い 3. 適切 4. やや高い 5. 高すぎる
- ウ．問題の量について
 - 1. 少なすぎる 2. やや少ない 3. 適切 4. やや多い 5. 多すぎる

- エ．設問が問おうとしている意図は，理解できた
 - 1. そう思う 2. ややそう思う 3. あまりそう思わない 4. そう思わない
- オ．明示的あるいは暗示的なヒントを与えるなど，早期実施の試験として配慮されている
 - 1. そう思う 2. ややそう思う 3. あまりそう思わない 4. そう思わない

内容に対する興味(ア)は，総じて高かった（回答 1, 2 の合計：78.0%～86.6%）。難易度(イ)について，サンプル問題 A では「適切」が最も多かった（46.8%, 65.1%, 51.1%）。サンプル問題 B では，「やや高い」が多かった（47.8%, 50.5%）。量(ウ)については，A 大問 1, 大問 2 では「適切」が多かった（46.8%, 65.1%）が，A 大問 3, B 大問 1, 大問 2 で「やや多い」が多かった（54.8%, 47.3%, 43.0%）。意図(エ)について「そう思う」「ややそう思う」の合計値が高く（A：79.0%, 81.2%, 86.5% / B：74.2%, 66.7%），設問の意図は理解されたと考えられる。ヒント(オ)については，「そう思う」「ややそう思う」の合計が高かった（A：63.8%, 73.7%, 75.6% / B：62.4%, 55.6%）。サンプル問題 B 大問 2 に関して，B の合計点上位 50% 以上では 60.9%，上位 25% 以上では 71.1% で，得点が高いほどヒントが理解できていた。設問に織り込まれた明示的，暗示的なヒントは理解されたと考えられる。アンケートの回答全般で，文系と理系において特筆すべき差は認められなかった。

表 15 サンプル問題各大問に関する回答（平均値）

	A 大問 1	A 大問 2	A 大問 3	B 大問 1	B 大問 2
ア	文理	1.9	1.7	1.7	1.9
	文	1.8	1.7	1.8	1.8
	理	1.9	1.8	1.7	2.1
イ	文理	3.5	3.1	3.6	3.7
	文	3.4	3.0	3.5	3.7
	理	3.6	3.2	3.6	3.8
ウ	文理	3.4	3.1	3.7	4.1
	文	3.3	3.0	3.7	4.1
	理	3.4	3.2	3.7	4.1
エ	文理	2.0	1.8	2.1	2.2
	文	1.9	1.9	2.0	2.2
	理	2.1	1.8	2.2	2.3
オ	文理	2.3	1.9	2.2	2.4
	文	2.2	1.8	2.2	2.3
	理	2.3	2.0	2.3	2.5

5 考察

令和 3 年度「志」総合試験に類似したサンプル問題を用いて 2021 年，2022 年に実施した模擬試験のデータを分析した結果，次の点が認められた。

- (1) サンプル問題と「志」との類似性については，

各大問の英文の特徴を数値化した12指標を、主成分分析に基づき6指標に選定し、多面的に分析した(2.1)。6指標の主成分分析およびクラスタ分析の結果、各クラスタに含まれる大問と、クラスタの第1主成分、第2主成分の特徴を示し、サンプル問題と「志」の英文の類似性が統計的に概ね認められた(2.2)。

(2) テスト間の相関分析、および因子分析により出題構成の類似性が高いことが確認された(3.3, 3.4)。

(3) 各設問の得点率と合計点との相関分析によって、サンプル問題Aの3題において相関が0.2より小さかった(その内1題は、英語文章作成能力を測る次の設問の予備問題で大きな差異は想定していなかった)が、全般的に各設問が受験者の能力を弁別する識別力が高いことが示唆された(3.5)。

(4) 各設問得点率と合計点分布の分析によって、測りたい能力が、設問で測られている場合と不十分な場合の要因を具体的に例示できた(3.6)。

(5) 信頼性の分析では、Cronbachの α 係数を算出し、A、Bでは期待される数値には至らなかったが、項目数が少ない場合は数値が小さくなる傾向がある(野口ほか, 2014)ことを考慮すれば、A・B合体では.728を示し、信頼性が高いことが示唆された(3.7)。

(6) アンケート調査の回答分析では、内容に対する興味(ア)は、総じて高かった。難易度(イ)については、Aでは「適切」、Bでは「やや高い」が多く、テストの平均点の差と同じ傾向がみられた。量(ウ)については、A大問1、大問2では「適切」、A大問3、B大問1、大問2で「やや多い」が多く、合計得点と関連していると推察される。意図(エ)については、「そう思う」「ややそう思う」の合計値より、設問を通してどのような能力を測ろうとしているかの意図への理解が示された。ヒント(オ)については、得点が高い受験者ほどヒントが理解できていた。設問の理解力が反映していると推察される。以上により、基礎学力の担保や早期実施に対する配慮に関して総合的に肯定的な回答が得られたと考えられる(4)。アンケートの回答全般で、文系と理系において特筆すべき大きな差は認められなかった。テーマに関して文系的、理系的な差異を設けたが、文系・理系に共通して大学で学修する際に最低限必要な英語の基礎学力の評価を意図した結果であると判断できる。

6 まとめ

本稿では、大学入学共通テストを課さない入学者選抜に対する学内に残る一部の懸念を払拭し、本学の大学入試改革を推進することに加え、他大学における入

学者選抜試験の出題検討に資することをねらいとした。「志」を本研究に活用できない制約があるためサンプル問題を代用したが、両者の類似性を統計的手法で示すことにより、「志」の妥当性および信頼性の高さ、設問に対する受験者の肯定的な評価を客観的、間接的に検証できた。

謝辞

本研究の分析にあたり山下仁司大阪大学教授(スチューデント・ライフサイクルサポートセンター)より指導助言をいただきました。ここに厚くお礼申し上げます。また、本研究は、科研費(21K02658)の助成を受けたものである。

参考文献

- Graesser, A.C., McNamara, D., Louwerse, M.M., and Cai, Z. (2004). "Coh-Matrix: Analysis of text on cohesion and language." *Behavior Research Methods, Instruments, & Computers*, 36(2), 193-202.
<https://link.springer.com/article/10.3758/BF03195564> (2023年11月11日).
- 野口裕之・大隈敦子(2014).『テストングの基礎理論』研究社.
- 松尾太加志・中村知靖(2002).『誰も教えてくれなかった因子分析』北大路書房.
- 水本篤(2013).「英文解析プログラムから得られる各種指標を使ったテキスト難易度の推定:教材作成への適用可能性」『外国語教育メディア学会(LET)関西支部メソドロジー研究部会2012年度報告論集』142-150.
<https://core.ac.uk/download/pdf/228724976.pdf> (2023年11月11日).
- 山下仁司(2022).「入試専門家が知っておくべきテスト理論・テスト分析方法」大阪大学高等教育・入試研究開発センター編『未来志向の大学入試デザイン論』大阪大学出版, 234-257.
- 吉田健三(2023).「英語の基礎学力を担保する総合試験の研究—令和3年度「志」特別選抜の総合試験に類似したサンプル問題を活用して—」『大学入試研究ジャーナル』33, 256-263.
<https://www.dnc.ac.jp/albums/abm.php?d=474&f=abm0003501.pdf&n=38> (2023年11月11日).