

【研究論文】

項目反応理論による
日本語リテラシーテストの検討（1）
—プレイスメントテストの改善—

高 橋 あすみ
松 浦 年 男

研究論文

項目反応理論による日本語リテラシーテストの検討(1)

——プレイスメントテストの改善——

高橋 あすみ 松浦 年男

Asumi TAKAHASHI Toshio MATSURA

目次

I 問題

1. 初年次生のリテラシーテスト
2. 項目反応理論
3. 本研究の目的

II 方法

1. プレイスメントテストの概要
2. 受検手続き
3. 対象者
4. 分析

III 結果

1. 項目分析の結果
2. 局所独立性と一次元性の確認
3. 項目母数の推定
4. テスト情報量
5. 学科・年度によるリテラシーの比較

IV 考察

1. プレイスメントテストの妥当性
2. 得点と潜在特性値 θ の比較
3. 改訂版の提案

I 問題

1. 初年次生のリテラシーテスト

大学初年次を対象とした基礎的リテラシ

[Abstract]

Japanese Language Literacy Test Using Item Response Theory (1): Improvement of the Placement Test

Aim: First-year students at Hokusei Gakuen University take a placement test for measuring their Japanese literacy. This study examined the Japanese language literacy test's validity by applying the item response theory to propose a revised version of the test. **Methods:** The placement test comprises four areas with 70 multiple-choice questions. All 6,285 first-year university students completed the test during 2014–2020. We validated the nature of each item (i.e., correct response rate, non-response rate, and point-biserial correlation coefficient), local independence, and one-dimensionality. A 2-Parameter logistic model was then applied to the remaining items. **Results:** Each item's discriminative power was estimated from 0.15 to 1.65. Eight items' discriminative power was under 0.4. The difficulty parameter was estimated from -3.84 to 4.38 . Test information was 8.14 when their latent traits were -0.80 . **Discussion:** Further, 50 items remained in the excluded low-discriminative power items were proposed to be reconstructed as the revised test. Excluded items, such as word meanings and honorific expressions, were found to be unfamiliar to first-year university students. Because the amount of test information in the revised test deviates from the trend in the target population's latent traits θ , continuing to reevaluate the test questions and options is crucial.

ーに関する研究は多く行われてきた(松浦, 2015; 山路・因・藤木, 2013; 日本数学会, 2012)。しかし, これらの調査が対象としているリテラシーテストそのものの妥当性につ

キーワード: 初年次教育, 日本語リテラシー, テスト開発, 項目反応理論

Key words: initial education, Japanese language literacy, test development, Item Response Theory

いて、検討が十分に進んでいるとは言い難い。本学で行われている基礎的リテラシーに関する調査についても同様である。

北星学園大学(以下、本学)では、2013年度から初年次生を対象とした「入学時基礎力調査」(以下、プレイスメントテスト)、及び「初年次到達度テスト」(以下、到達度テスト)を実施している。二つのテストは、2012年に採択された文部科学省大学間連携共同教育推進事業の課題「学士力養成のための共通基盤システムを活用した主体的学びの促進」の一環として開発された(中嶋, 2022)。テスト開発の詳細な開発経緯や事業内容は、中嶋(2022)や金子(2021)に譲るが、プレイスメントテストと到達度テストは現在まで実施実績を積み重ねている。初年次生を対象とした基礎的リテラシーテストは、学生個人のリテラシー向上だけでなく、初年次教育の評価や改善のために重要であり、大学生の初年次教育の有効性を高めるためにまずはテストの妥当性や信頼性を高めていくことが求められる。

これまで、実施されている基礎的リテラシーテストの改善を図った研究には、石原(2014)や仲道・秋山(2021)がある。石原(2014)は英語プレイスメントテストの点数を古典的テスト理論に基づいて検討し、テスト改善の提言を行っている。ここでは、テストの得点分布、信頼性、古典的テスト理論に基づく項目分析(項目難易度、点双列相関係数、項目弁別力指数、実質選択肢数とその適切度)、妥当性を検証した上で、テストが概ね適切であることを確認している。その一方で、易しすぎる項目や、選択肢の適切度が低い項目などの改善を指摘した。仲道・秋山(2021)は、日本語プレイスメントテストの試験時間と出題順が回答率と正答率にどのように影響しているかを検討するために、試験時間と出題順の条件が異なる4群でテストを実施した。その結果、元々プレイスメン

トテストで実施していた、高校生にも解き慣れた順番(漢字→語彙→文法・敬語→短文読解)での出題で、75問に対して試験時間30分が適切であることを実証した。

本学の日本語リテラシーテストにおいても、終盤の問題が無回答になる傾向が見受けられるなど改善の余地がある。例えば、現在の日本語プレイスメントテストは70問を20分で解く構成であり、仲道・秋山(2021)の知見と照らし合わせると、問題数に対して試験時間が短い可能性が考えられる。結果として、本来測定しようとしているリテラシーではない部分、意欲や集中力などがテスト得点に反映されている可能性がある。

さらに、プレイスメントテストと到達度テストは問題数や配点が異なるため、得点をそのまま比較できる設計にはなっていない。そこで金子(2021)は日本語プレイスメントテストと到達度テストの結果を得点率によって比較している。その結果、1年間で読解力の低下が見られたことを確認したが、これには二つのテストの読解に関する問題数の違いや、制限時間といった要因、すなわち学生ではなくテスト側の要因があると考察している。元々、二つのテストは初年次教育の成果を評価するために開発されたものではない(金子, 2021)とはいえ、現行のテストで初年次生の日本語リテラシーの変化を測定できなければ、学生自身が自分のリテラシーを確かめたり、初年次教育の成果を明らかにしたりすることは厳密には困難である。

日本語リテラシーテストの抱える以上のような課題の改善を図ることで、学生の受検負担の軽減と、より質の高いフィードバックの提供につながるるとともに、得られた知見を初年次生の日本語教育や学習支援に活かすことが期待される。

2. 項目反応理論

異なる問題から構成されていても、問題が

測定している能力を比較できるテストを開発する方法に、項目反応理論 (Item Response Theory: 以下, IRT) がある。石原 (2014) も、プレイメントテストの改善の検討方法に古典的テスト理論ではなく IRT を用いる可能性について言及している。IRT とは、調査対象者に依存する情報である潜在特性値 θ と、項目に依存する情報である項目母数 (識別力 a , 困難度 b) とを切り分けて、テストを検討することができる理論である。日本語プレイメントテストで同じ得点の学生が2名いる場合に、より難易度の高い問題に正答した学生の方が、日本語のリテラシーは高いことが推察される。テスト得点にはこういった問題の性質は加味されないが、IRT を適用すれば、問題から独立した日本語リテラシーを潜在特性値 θ として算出することができるのである。併せて、各項目の項目母数が明らかになるため、各項目の特徴からテスト問題を検討する (例えば中嶋・佐々木, 2017) ことにも役立つ。

加えて、IRT における等化の手続きを経ることによって、問題や項目数の異なるプレイメントテストと到達度テストから、受検者の潜在特性値 θ を同一尺度上で比較することができるようになる。等化の方法には、共通項目を含む二つのテストを別々の集団が受検するテスト計画を組む共通項目法や、共通の受検者が二つのテストを受検する共通被験者法がある (高橋, 2021)。本学では同一の対象者が二つのテストを受検しているが、受検時期が異なるために共通被験者法も適用できない。そのため等化をすぐに実現することは不可能であるが、IRT でそれぞれのテストを解析しておくことは今後のテスト開発に活かせらるだろう。

3. 本研究の目的

以上より本研究では、IRT を用いて日本語リテラシーを測定するテストの妥当性を検証

し、簡便で効果的な日本語リテラシーテストを提案することを目的とした。本稿では、初年次の4月に実施する日本語プレイメントテストを取り上げ、その改善を提案する。

II 方法

1. プレイメントテストの概要

プレイメントテストは入学初期の大学生の日本語リテラシーを測定するテストで、(1) 漢字, (2) 語彙, (3) 文法・敬語, (4) 短文読解の4分野から成る。Table 1に詳細を示した。本学での実施が開始された初年の2013年は、全100問の選択式問題を30分以内で回答する形式であったが、2014年以降は問題が削減され、前述のとおり全70問を20分で解く形式に変更されている。その後、2021年まで問題の内容や形式は毎年共通である。本研究では2014年から現在まで実施されている70問版を扱うこととした。

問題は1問につき正答1点、誤答0点で採点する。通常の採点では、合計点に10/7を掛けて100点満点に換算した得点 (小数点以下切り捨て) を算出するが、本研究ではIRTを適用するため、素点のまま分析に用いた。

Table1 プレイメントテストの構成

分野	小項目	問題番号
漢字	読みと書き	1 ~ 15
	四字熟語	16 ~ 20
語彙	ことわざ・成句	21 ~ 25
	語義	26 ~ 50
文法・敬語	受身・使役・可能, 敬語	51 ~ 60
短文読解	第1問	61 ~ 65
	第2問	66 ~ 70

2. 受検手続き

プレイメントテストは、本学の学習管理システム Moodle 上で提供されている。対象となる初年次生は毎年度4月中旬に、必修科目

Table2 プレイメントテスト受検者の人数内訳

学科／受検年度	14	15	16	17	18	19	20	合計
英文	140	125	137	130	125	121	120	898
心理・応用コミュニケーション	99	103	107	92	109	97	95	702
経済	175	163	173	170	165	160	168	1,174
経営情報	108	128	118	131	115	121	93	814
経済法	130	130	130	130	129	127	99	875
福祉計画	92	102	102	85	95	91	108	675
福祉臨床	91	92	88	97	88	80	102	638
福祉心理	73	61	70	70	74	81	80	509
合計	908	904	925	905	900	878	865	6,285

である情報系の初年度教育「情報入門」の授業内で担当教員から口頭で説明を受け、必ず受検するように指示される。指定された期間（約20日間）の中で随時受検し、Moodleで解答を送信することで受検は完了する。

解答データに関しては「教育研究目的で結果が学会等で発表されることがある」と説明時に案内され、テスト画面上にもその旨が記載されていた。解答の提出を以てこれらに同意を得たこととした。なお、本研究へのデータ利用については、北星学園大学研究倫理審査委員会の承認を得た（21-第23号）。

3. 対象者

2014～2020年の間に四年制大学8学科の初年次生6,456名がプレイメントテストに解答した。このうち未受検・未送信と記録されていた171名のデータを除外し、6,285名(有効数97.4%)の解答データを分析対象とした。対象者の内訳はTable 2のとおりである。

4. 分析

(1) 項目分析

以降は各問のことを項目と表記し、具体的な問いを指すときはQと番号で記す。まず、テストの信頼性に影響を与える項目を検討するために、各項目の正答率と未回答率を算出した。次に、対象者を合計得点の低い順から

並べて人数が最も均等な割合に分かれる5群に分け、各群の正答率を算出した上で、群の並び順で正答率の増加が見られるかどうかを確認した。さらに、点双列相関係数を算出し、合計点との整合性が低い0.20以下の項目を確認した。以上の項目分析によってIRTを適用する前に除外する項目を検討した。

(2) 局所独立性と一次元性の確認

IRTを適用するためには各項目の解答が他の項目から影響を受けず独立していること、すなわち局所独立性が成り立っている必要がある。プレイメントテストにおいては、短文読解の第1問(Q61～65)と第2問(Q66～70)が1段落の短い文章に適切な語句を当てはめていく問題であり、先の項目の解答が後の項目に影響する可能性がある。そこで、各5項目のYen's Q3統計量を算出した。Yen's Q3統計量が0.20以上の値にその項目間での局所独立性が疑わしいとされる(Chen & Thissen, 1997; 宇佐美・荘島・光永・登藤, 2019)ため、これを基準とした。

また、IRTを適用するためのもう一つの条件として、全項目の一次元性が成立している必要がある。そのため、上記の項目分析と局所独立性の検定を経て残った項目に対し、スクリープロット基準と因子寄与率、信頼性 α 係数を確認した。

(3) IRT の適用

最終的に抽出された項目について 2 母数ロジスティックモデル (2-Parameter Logistic Model: 以下, 2PL) を適用した。2PL のモデル式は下記のとおり表される。

$$P(u_j=1|\theta) = \frac{1}{1+\exp\{Da_j(\theta-b_j)\}}$$

この式において, θ は対象者の潜在特性値, a_j は項目 j の識別力, b_j は項目 j の困難度を表す。本研究では $D=1.0$ として, 各項目の識別力 a と困難度 b を推定した。その後, 推定した項目母数を用いてプレイメントテストの信頼性を意味するテスト情報量と, 対象者の潜在特性値 θ を算出した。

(4) 対象者の所属学科・受検年度による差

対象者の日本語リテラシーが, 所属学科, 及び受検年度によって差があるかを検討するために, 抽出された項目数での合計点と, 項目から算出された潜在特性値 θ それぞれについて, 一元配置分散分析を実施した。有意差が確認された場合には多重比較検定 (Holm 法) を行った。

(5) 分析ソフト

基本的な集計と分析は Excel で行い, IRT には EasyEstimation(熊谷, 2009)を用いた。また Yen's Q3 統計量は R の irtoys パッケージを用いた。分散分析には HAD (清水, 2016) を用いた。

III 結果

1. 項目分析の結果

各項目の正答率と未回答率を Table 3 に示した。70項目全体の平均得点は $M = 36.17$ ($SD = 9.68$, $0 \sim 65$ 点), 平均正答率は 51.7% ($SD = 0.14$, $0 \sim 92.9\%$) であった。

各項目のうち正答率が85%以上と高かったのは6項目(Q1, 9, 17, 19, 20, 61), 15%以下と低かったのは3項目(Q40, 58, 59)であった。

70項目全体の平均未回答率は2.1%で, Q1では未回答率は0.1%であったが, Q4の1.2%からは漸次増加し, Q67では10.2%, 最終のQ70では12.6%であった。

合計得点にしたがい対象者数が最も均等な割合になる5群に分けたところ, L群は0~27点 ($n = 1,209$, 19.2%), LM群は28~33点 ($n = 1,267$, 20.2%), M群は34~38点 ($n = 1,266$, 20.1%), MH群は39~44点 ($n = 1,259$, 20.0%), H群は45点以上 ($n = 1,284$, 20.4%) となった。正答率がL群からH群にかけて増加していれば妥当性の高い項目と考えられるが, 6項目(Q3, 14, 24, 50, 58, 59)で群ごとの直線的な増加が見られなかった。

次に, 70項目の合計点と各項目の点双列相関係数を算出した (Table 3 参照)。0.2以下の低い値を示したのは12項目(Q3, 4, 9, 14, 17, 23, 29, 50, 55, 58, 59, 60)であった。

以上の項目分析の中で, 除外基準の一つでも該当した項目は18項目あり, そのうち二つ以上の基準に該当したのは7項目であった。したがって, 7項目(Q3, 9, 14, 17, 50, 58, 59)を除外して以降の分析を行うこととした。

2. 局所独立性と一次元性の確認

短文読解問題の局所独立性を確認した結果, 第1問の5項目については, Q61×Q64 (Yen's Q3 = .28) の間に, 第2問の5項目については Q66×Q70 (Yen's Q3 = .22), Q67×Q68 (Yen's Q3 = .28), Q67×Q69 (Yen's Q3 = .24), Q68×Q69 (Yen's Q3 = .33) の間にそれぞれ局所独立性の成り立たない可能性が示された。第1問のQ61とQ64は, 異なる文章に当てはまる語句を別々に答える問題で, Q61が誤答や未回答でもQ64への影響

Table 3-1 プレイメントテストの項目分析 (Q1 ~ 35)

Q	全体の 正答率	群ごとの正答率					未回答率	点双列相関係数
		L	LM	M	MH	H		
1	85.8%	65.3%	82.1%	89.3%	93.2%	97.8%	0.1%	0.33
2	66.1%	47.7%	59.4%	63.9%	72.6%	85.9%	0.2%	0.28
3	61.5%	56.1%	62.2%	61.8%	61.6%	65.6%	0.2%	0.06
4	50.5%	36.2%	45.9%	50.6%	56.1%	63.0%	0.5%	0.19
5	30.4%	19.0%	24.2%	27.3%	32.3%	48.4%	0.6%	0.22
6	78.3%	48.8%	70.7%	82.5%	91.5%	96.4%	0.3%	0.41
7	48.8%	31.4%	39.6%	47.6%	56.3%	68.1%	0.3%	0.26
8	65.6%	47.4%	62.0%	67.9%	73.5%	76.3%	0.4%	0.22
9	90.4%	81.4%	89.2%	91.2%	93.6%	96.4%	0.4%	0.17
10	66.9%	32.2%	54.8%	68.7%	83.2%	93.8%	0.7%	0.46
11	39.2%	19.6%	25.0%	34.7%	45.8%	69.6%	0.4%	0.36
12	54.6%	28.5%	44.8%	52.8%	65.2%	80.4%	0.5%	0.37
13	80.9%	51.2%	75.5%	85.9%	93.6%	97.0%	0.3%	0.40
14	18.1%	13.8%	13.6%	14.5%	18.0%	30.1%	0.9%	0.16
15	60.8%	43.4%	55.8%	61.7%	66.4%	75.5%	0.5%	0.23
16	70.8%	49.8%	65.2%	72.0%	78.5%	87.3%	0.5%	0.28
17	90.6%	81.8%	89.5%	89.5%	95.2%	96.8%	0.6%	0.18
18	68.0%	43.9%	61.6%	69.0%	77.3%	86.9%	0.6%	0.31
19	88.5%	70.2%	85.2%	91.2%	96.3%	98.5%	0.5%	0.32
20	91.8%	77.9%	91.0%	93.4%	96.5%	99.4%	0.5%	0.27
21	36.9%	20.7%	27.4%	34.8%	41.0%	59.9%	0.6%	0.28
22	36.6%	16.5%	23.3%	33.5%	43.8%	64.4%	0.3%	0.36
23	23.6%	17.5%	20.0%	20.1%	26.5%	33.3%	0.4%	0.14
24	20.3%	12.6%	11.8%	15.3%	21.5%	39.7%	0.5%	0.25
25	59.0%	38.7%	52.2%	58.3%	66.7%	78.2%	0.4%	0.28
26	58.8%	32.8%	49.7%	58.1%	69.8%	82.2%	0.6%	0.35
27	63.5%	42.4%	53.6%	62.4%	71.2%	86.8%	0.5%	0.33
28	42.7%	16.2%	24.2%	39.2%	56.1%	76.2%	0.6%	0.44
29	34.7%	29.4%	32.9%	36.2%	37.0%	37.9%	0.6%	0.07
30	24.1%	14.6%	16.9%	21.6%	25.7%	41.1%	0.6%	0.23
31	31.2%	14.5%	20.3%	25.7%	34.4%	59.9%	0.7%	0.34
32	37.7%	19.5%	27.1%	33.1%	41.8%	65.7%	0.8%	0.33
33	65.6%	39.2%	54.8%	64.7%	77.2%	90.7%	0.6%	0.38
34	32.2%	14.5%	22.3%	29.1%	36.9%	57.4%	0.7%	0.33
35	21.6%	9.2%	11.8%	15.0%	23.0%	48.0%	0.9%	0.34

注) 太字：項目分析の基準に該当した項目

Table 3-2 プレイメントテストの項目分析（Q36～70）

Q	全体の 正答率	群ごとの正答率					未回答率	点双列相関係数
		L	LM	M	MH	H		
36	58.8%	27.4%	45.5%	57.7%	72.4%	89.4%	0.7%	0.44
37	33.7%	15.4%	21.7%	27.8%	39.4%	63.0%	0.8%	0.35
38	31.3%	16.1%	17.0%	24.6%	35.0%	62.9%	0.7%	0.36
39	61.2%	35.4%	51.9%	60.4%	74.5%	82.6%	0.7%	0.34
40	11.4%	5.0%	7.5%	9.2%	12.5%	22.3%	0.8%	0.20
41	68.3%	34.2%	57.1%	74.7%	83.0%	90.7%	0.7%	0.42
42	39.1%	19.0%	23.8%	33.9%	46.7%	70.6%	0.9%	0.38
43	28.8%	13.4%	19.3%	25.2%	33.8%	51.2%	0.9%	0.29
44	53.4%	23.2%	39.5%	51.8%	66.6%	84.3%	1.0%	0.43
45	61.1%	31.8%	52.8%	61.8%	72.8%	84.9%	0.9%	0.37
46	66.5%	25.9%	52.9%	71.5%	84.8%	95.3%	1.2%	0.51
47	70.2%	35.8%	60.3%	75.6%	83.7%	93.8%	1.4%	0.43
48	63.3%	35.5%	52.6%	67.0%	74.2%	85.9%	1.4%	0.35
49	67.5%	40.3%	65.0%	71.2%	77.5%	82.1%	1.7%	0.29
50	29.5%	22.9%	26.8%	26.2%	29.1%	42.1%	2.0%	0.15
51	43.0%	19.9%	25.7%	37.8%	54.4%	76.0%	2.0%	0.40
52	45.9%	25.6%	33.1%	43.3%	54.2%	72.1%	2.1%	0.32
53	61.8%	33.6%	49.4%	63.4%	73.6%	87.5%	2.4%	0.38
54	50.0%	27.9%	41.7%	51.8%	58.1%	69.3%	2.7%	0.27
55	76.9%	58.6%	75.0%	81.3%	84.0%	84.7%	2.5%	0.18
56	51.8%	33.5%	44.3%	51.9%	60.0%	68.2%	2.9%	0.23
57	33.6%	15.4%	24.1%	32.8%	40.6%	54.4%	3.3%	0.27
58	5.4%	8.2%	6.8%	5.0%	3.3%	3.7%	3.2%	-0.09
59	6.7%	9.3%	5.4%	5.8%	6.4%	7.0%	3.6%	-0.03
60	34.3%	22.1%	29.7%	32.2%	37.1%	49.9%	4.0%	0.19
61	87.2%	65.0%	85.6%	92.0%	94.8%	97.7%	4.3%	0.30
62	20.6%	8.9%	11.4%	17.0%	24.7%	40.3%	5.0%	0.27
63	42.3%	24.2%	32.7%	42.4%	51.2%	60.3%	6.3%	0.24
64	80.4%	49.0%	77.0%	87.8%	90.8%	96.0%	6.7%	0.37
65	47.5%	25.1%	38.4%	48.3%	55.5%	69.1%	7.7%	0.28
66	35.9%	8.3%	18.9%	33.7%	48.0%	68.9%	9.7%	0.44
67	68.4%	37.4%	61.4%	73.5%	80.2%	88.0%	10.2%	0.36
68	75.3%	46.5%	68.2%	80.1%	86.9%	93.5%	11.3%	0.38
69	71.8%	44.3%	64.3%	76.6%	82.8%	89.6%	12.0%	0.34
70	37.0%	15.6%	27.3%	37.5%	45.2%	58.3%	12.6%	0.30

注）太字：項目分析の基準に該当した項目

力は低いと考えられた。一方で、第 2 問で語句を当てはめる箇所は段落の最後の 2 文に集中しており、Q3 統計量からも Q66～70 のほとんどの項目間で局所独立性が成立しているとは言い難かったことから、第 2 問 (Q66～70) は以降の分析から除外した。

残りの 58 項目について、一次元性を確認するためにスクリープロット基準を確認した (Figure 1)。第 1 因子の寄与率は 10.85, 第 2 因子は 2.47, 第 3 因子は 1.88 であり, 第 1 因子で寄与率が十分高く, 第 2 因子以降の寄与率の変化はゆるやかであることが確認された。クロンバックの信頼性 α 係数は 0.85 であった。以上から, 58 項目の一次元性は十分であると判断し, 残った 58 項目で 2PL を適用することとした。

3. 項目母数の推定

各項目の識別力 a と困難度 b を推定した結果を Table 4 に示す。識別力 a は平均 0.79 で全体的な識別力は十分であったが, 0.15～1.65 の範囲をとり, 識別力が低いと考えられる 0.4 以下を示した項目も 8 項目 (Q4, 8, 23, 29, 55, 56, 60, 63) あった。困難度 b は平均 0 で $-3.83 \sim 4.38$ と幅広い範囲をとり, バランスがとれていた。

4. テスト情報量

以上の結果から識別力の低かった 8 項目を除外し, 残った 50 項目に 2PL を適用して改めて項目母数を推定し, テスト情報量を算出した。テスト情報量は Figure 2 のとおり山

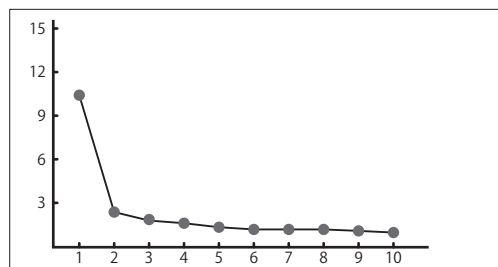


Figure 1 58 項目のスクリープロット基準

型で, 潜在特性値 $\theta = -0.80$ の場合に 8.14 で最大となった。

次に, 対象者の潜在特性値 θ (事後期待値) を算出した。 θ の平均は $M = 0.00$ ($SD = 0.93$) で, $-2.48 \sim 3.15$ の幅をとった。 Figure 3 にヒストグラムを示した。

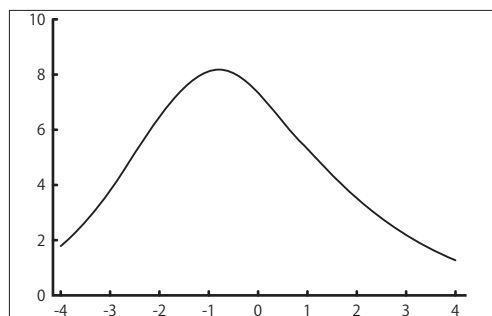


Figure 2 テスト情報量

(注) 横軸が潜在特性値 θ 、縦軸がテスト情報量を表す。

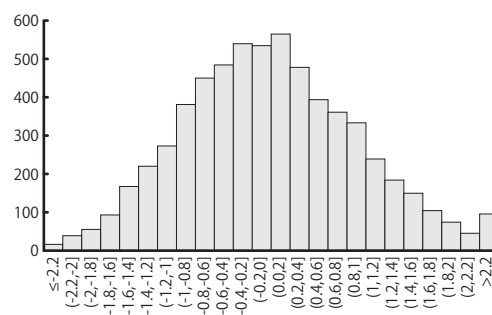


Figure 3 潜在特性値 θ のヒストグラム

(注) 横軸が潜在特性値 θ 、縦軸が対象者の人数を表す。

5. 学科・年度によるリテラシーの比較

算出した潜在特性値 θ と 50 項目での合計得点について, 所属学科と受検年度による差を一元配置分散分析によって比較した。結果を Table 5 に示す。なお, 潜在特性値 θ と合計得点の相関係数は $r = .98$ であった。

まず学科について, 潜在特性値 θ の有意な差が確認された ($F(7, 6275) = 77.33, p < .001, \eta_p^2 = .08$)。多重比較の結果, θ が 0.3 以上だった英文学科 ($\theta = 0.40$) と福祉心理学

Table4 プレイスメントテスト58項目の項目母数

Q	a	SE	b	SE	Q	a	SE	b	SE
1	1.17	0.05	-1.90	0.06	34	0.74	0.03	1.12	0.06
2	0.62	0.03	-1.16	0.07	35	0.90	0.04	1.63	0.06
4	0.31	0.03	-0.05	0.08	36	1.12	0.04	-0.46	0.03
5	0.43	0.03	1.96	0.14	37	0.81	0.03	0.93	0.05
6	1.37	0.05	-1.25	0.04	38	0.86	0.03	1.01	0.05
7	0.50	0.03	0.09	0.05	39	0.78	0.03	-0.70	0.04
8	0.38	0.03	-1.82	0.15	40	0.61	0.04	3.62	0.23
10	1.32	0.04	-0.75	0.03	41	1.18	0.04	-0.86	0.03
11	0.81	0.03	0.60	0.04	42	0.89	0.03	0.54	0.04
12	0.83	0.03	-0.27	0.03	43	0.63	0.03	1.53	0.08
13	1.41	0.05	-1.39	0.04	44	1.07	0.04	-0.20	0.03
15	0.43	0.03	-1.03	0.09	45	0.88	0.03	-0.63	0.04
16	0.63	0.03	-1.54	0.08	46	1.70	0.05	-0.66	0.02
18	0.72	0.03	-1.21	0.06	47	1.22	0.04	-0.96	0.03
19	1.28	0.06	-2.04	0.07	48	0.83	0.03	-0.81	0.04
20	1.21	0.06	-2.50	0.10	49	0.60	0.03	-1.40	0.08
21	0.58	0.03	0.98	0.07	51	0.93	0.03	0.29	0.03
22	0.82	0.03	0.75	0.04	52	0.69	0.03	0.20	0.04
23	0.24	0.03	4.91	0.62	53	0.87	0.03	-0.74	0.04
24	0.65	0.03	2.30	0.12	54	0.51	0.03	-0.13	0.05
25	0.57	0.03	-0.72	0.06	55	0.35	0.03	-3.88	0.37
26	0.77	0.03	-0.56	0.04	56	0.40	0.03	-0.35	0.07
27	0.76	0.03	-0.87	0.05	57	0.56	0.03	1.19	0.08
28	1.15	0.04	0.30	0.03	60	0.32	0.03	1.90	0.18
29	0.15	9.98	4.23	0.79	61	1.17	0.06	-2.41	0.10
30	0.54	0.03	2.27	0.13	62	0.66	0.03	2.12	0.11
31	0.82	0.03	1.08	0.05	63	0.40	0.03	0.49	0.07
32	0.75	0.03	0.73	0.05	64	1.23	0.05	-1.88	0.06
33	0.97	0.04	-0.82	0.04	65	0.51	0.03	-0.12	0.05

注) a は識別力、b は困難度、SE は標準誤差を表す

科 ($\theta = 0.36$) は、心理・応用コミュニケーション学科 ($\theta = 0.27$) をはじめとする他の学科よりも有意に高かった。福祉臨床学科 ($\theta = 0.02$) は、他の7学科すべてに対して有意な差があった。 θ が負を示した残り4学科の間においては、他学科との有意な差は認められなかった。

合計得点についても、学科による有意な差が認められた ($F(7, 6284) = 75.23, p < .001, \eta_p^2 = .07$)。得点が僅差の英文学科 vs 福祉心理学科、福祉心理学科 vs 心理・応用コミュニケーション学科では有意な差は認められなかった。また、経済学科、経営情報学科、経

済法学科の3学科間でも有意差は見られなかった。特性値 θ で全学科と有意差が見られた福祉計画学科は、vs 経済学科、vs 経営情報学科との間の得点差は有意ではなかった。以上に挙げた学科間を除く群間で有意な差が認められた。

受検年度によっても潜在特性値 θ には有意な差が見られた ($F(6, 6276) = 10.28, p < .001, \eta_p^2 = .01$)。多重比較の結果、2020年度受検者 ($\theta = 0.21$) のみ他の年度と比較して有意に高かった。

得点においても有意な差が認められ ($F(6, 6284) = 10.19, p < .001, \eta_p^2 = .01$)、2020

Table5 学科・受検年度別日本語リテラシー

	θ	SD	得点	SD
英文	0.40	0.91	29.97	7.72
心コミ	0.27	0.96	28.77	8.18
経済	-0.21	0.89	24.73	7.82
経営情報	-0.26	0.83	24.36	7.25
経済法	-0.14	0.88	25.27	7.71
福祉計画	-0.26	0.88	24.07	7.63
福祉臨床	0.02	0.89	26.67	7.64
福祉心理	0.36	0.92	29.41	7.92
14年度	0.03	0.94	26.82	8.04
15年度	-0.03	0.93	26.22	8.03
16年度	-0.10	0.93	25.71	8.11
17年度	-0.06	0.92	25.92	8.10
18年度	-0.03	0.90	26.28	7.89
19年度	-0.03	0.97	26.09	8.32
20年度	0.21	0.89	28.27	7.53

(注) 学科は略称で表記

年度の得点が他全ての受検年度より高いのは同様であった。また、2014年受検者の得点の方が2016年受検者よりも有意に高かった。

IV 考察

1. プレイメントテストの妥当性

本研究は、本学で実施している日本語プレイメントテストの妥当性をIRTによって検証し、テスト改善の提案をすることが目的であった。

プレイメントテストには、全70問に対して未回答率が60問あたりから増加するという問題点があった。項目分析(正答率、点双列相関係数)と局所独立性の検定、IRTによる識別力を検証した結果、日本語リテラシーを測定する妥当性の高い項目は50問に選定された。

まず漢字分野について「読みと書き」では15項目のうち5項目(Q3, 4, 8, 9, 14)が除外基準に該当し、「四字熟語」では5項目

のうち1項目(Q17)が該当した。語彙分野では「ことわざ・成句」は5項目中1項目(Q23)、「語義」は25項目中2項目(Q29, 50)が該当した。これらの分野は元の問題数に対して除外された項目数が少なかった。それに比して、文法・敬語分野では10項目のうち5項目(Q55, 56, 58, 59, 60)が除外基準に該当しており、敬語の問題は妥当性が低かった。

「漢字」および「語義」の該当項目について、「(平成版)日本語の語彙特性」(天野・近藤, 1999)を用いて「単語読み」の親密度¹を調べたところ、Q29, 50はいずれもなじみ度が2.5と1.5で、NTT語彙量で換算すると60,000語以上のものに相当し、少なくとも大学1年生に課すテストとしては難易度が高すぎるものと考えられる。ただし、例えばQ14も親密度は1.9と低い、今回の調査結果では十分な妥当性があるとされた。これは設問の形式や選択肢の設定によって変わるところなので、さらなる検討を要する。また、令和版(NTT研究所, 2021)の語彙データを用いてより現代を生きる大学生の親密度に沿った分析を行う必要もあるだろう。

文法・敬語分野において除外基準に該当した問題は、「ご苦勞様／お疲れ様」の上司に対する言い方(Q55)、「課長が出勤する」の部下からの言い方(Q56)、「お気をつけて」の客に対するあいさつ表現(Q58)、謙遜する「とんでもない」の言い方(Q59)、学生から先生に対する「話す」の言い方(Q60)であった。これらの問いは高校までにそれほど多くの時間を割いて習う範囲とは言いがたく、また、日常生活でも比較的なじみのない種類の事項であったため正答率が低かったと考えられる。テスト問題全体との整合性が低く、他の問題が測定している日本語リテラシーとは異なる性質を持つ問いであると判断されたとも解釈できる。漢字や語彙、読解の問題が文語表現を問うものであるのに対し、敬

語は会話表現を問う問題であった。敬語表現は、ロールプレイなどを取り入れたアクティブ・ラーニング形式での教育実践が提案されている(杉本, 2019; 福岡, 2019)ことから、漢字や語彙などの文語表現やアカデミックライティングを目的とした日本語教育とは別に、教育や評価を行う必要性が示唆される。

短文読解分野においては、第1問の1項目(Q63)の識別力が低く、第2問が局所独立性の観点から除外されたため、文法・敬語分野と同様に問題数は半減した。ただし、抽出された4項目の識別力や困難度はバランスよく、日本語リテラシーを調べるのに良質な項目が選定できたと考えられる。

2. 得点と潜在特性値 θ の比較

50項目に選定された日本語プレイスメントテストから受検生の潜在特性値 θ を算出し、学科別、受検年度別に比較したところ、得点による比較に比べて潜在特性値 θ による比較では、有意差が示された群間が少なくなった。これは、テスト得点に含まれていた項目側の要因によって余分な有意差が検出されてしまっていたためであると考えられる。得点だけの差でみれば多くの学科間で日本語リテラシーに差があるように見え、各学科に対応した工夫を初年次教育に取り入れる必要があるように思われるが、潜在特性値 θ が正の値をとった学科、0付近の学科、負の値をとった学科の3種類に分けて考えれば十分、内容の工夫を検討できるであろう。

3. 改訂版の提案

本研究の分析から、プレイスメントテストは50問に選定できる可能性が見いだされた。一般に学力テストは易しい問題から難しい問題に移行するように出題されるため、各問を困難度の低い順番から並べ替えることによって、受検生にとって解答しやすい構成になる。そこで短文読解分野を除く各分野の問題を、

困難度の降順に並べ替えたものを改訂版としてTable 6に示した。本研究で提案した改訂版を実際に運用するには、50問を20分で解く形式で実際にテストを実施し、終盤の問題の未回答率という課題が解消されるかどうかを確かめることが望ましい。

また、テスト情報量からは潜在特性値 $\theta = -0.80$ のとき信頼性が最も高く、日本語リテラシーが少し低い対象者の日本語リテラシーを測るのに適したテストであることが示された。今回の対象者のヒストグラムをみるに全体の潜在特性値の傾向は $\theta = -0.80$ より高いことが明らかであり、本学の初年次生にとって改訂版のテストは易しめであると考えられる。プレイスメントテストの役割を考えると、クラス分けやリメディアル教育の必要な学生の識別に適している点では有用である。しかしながら、2020年度受検者の潜在特性値 θ が他年度より有意に高かったことから、これ以降、学生の日本語リテラシーが高まり、プレイスメントテストとのずれが大きくなっていく可能性も考えられる。そのような場合は、50項目の中で困難度の低かった項目の選択肢や、困難度が高くなるような問題の追加を検討し、本学の初年次生に適したテストに改善することも考えていく必要がある。

【謝辞】

本研究は、2021年度北星学園大学特定研究費の助成を受けて実施したものである。本研究データの取得に協力をいただいた中嶋輝明先生に御礼申し上げます。

【注】

(1) 単語の親密度は1~7の数値で表され、高いほどなじみ深い単語であることを示す。

Table6 プレイメントテスト50項目改訂版

Q	小項目	困難度 <i>b</i>	Q	小項目	困難度 <i>b</i>	Q	小項目	困難度 <i>b</i>
1	読みと書き	-1.91	24	ことわざ・成句	2.27	38	語義	1.03
13	読みと書き	-1.38	49	語義	-1.45	34	語義	1.08
6	読みと書き	-1.28	47	語義	-0.96	31	語義	1.09
2	読みと書き	-1.18	41	語義	-0.86	43	語義	1.53
15	読みと書き	-1.11	27	語義	-0.85	35	語義	1.63
10	読みと書き	-0.74	33	語義	-0.83	30	語義	2.23
12	読みと書き	-0.28	48	語義	-0.83	40	語義	3.48
7	読みと書き	0.08	39	語義	-0.70	53	文法・敬語	-0.73
11	読みと書き	0.59	46	語義	-0.66	54	文法・敬語	-0.12
5	読みと書き	1.99	45	語義	-0.63	52	文法・敬語	0.19
20	四字熟語	-2.46	26	語義	-0.55	51	文法・敬語	0.30
19	四字熟語	-2.03	36	語義	-0.43	57	文法・敬語	1.24
16	四字熟語	-1.55	44	語義	-0.20	61	短文読解	-2.46
18	四字熟語	-1.20	28	語義	0.30	62	短文読解	2.13
25	ことわざ・成句	-0.70	42	語義	0.55	64	短文読解	-1.89
22	ことわざ・成句	0.73	32	語義	0.73	65	短文読解	-0.14
21	ことわざ・成句	0.98	37	語義	0.93			

〔参考文献〕

- 天野成昭・近藤久(1999) 日本語の語彙特性
第1巻 語彙親密度. 三省堂
- Chen W.H., & Thissen D. (1997) Local
Dependence Indexes for Item Pairs Using
Item Response Theory. *Journal of Educational
and Behavioral Statistics*, 22, 265-289
<https://doi.org/10.2307/1165285>
- 福岡昌子(2019) アクティブ・ラーニングによる
敬語指導の実践研究-場面と人間関係を意識
した授業を目指して-. 三重大学高等教育研
究, 26, 27-37
- 石原知英(2014) 古典的テスト理論を用いた2012
年度新入生英語プレイズメントテストの分
析と改善への提言. 言語と文化: 愛知大学語
学教育研究室紀要, 57(30), 1-10
- 金子大輔(2021) 大学の初年次学生に対するプレ
イズメントテストと到達度テストの結果—
—初年次教育の学修成果の可視化の試み—
—. 北星学園大学経済学部北星論集, 60(2),
34-42
- 熊谷龍一(2009) 初学者向けの項目反応理論分析
プログラム EasyEstimation シリーズの開発.
日本テスト学会誌, 5, 107-118
- 松浦年男(2015) 大学初年次の学生に対する日本
語語彙力調査の試行. 北星学園大学文学部北
星論集, 52(2), 53-61
- 中嶋康博・佐々木良勝(2017) 項目反応理論によ
る数学基礎力確認テストの解析. 久留米工業
大学研究報告, 39, 63-71
- 中嶋輝明(2022) 大学間連携共同教育推進事業の
成果概要—「主体的学び」プログラムを中心
に—. 北星学園大学文学部北星論集, 59(2),
111-121
- 仲道雅輝・秋山英治(2021) 日本語プレイズメン
トテストにおける試験時間と出題順が解答
率と正答率に及ぼす影響. リメディアル教
育研究, 15, 21-33
- 日本数学会(2012) 「大学生数学基本調査」に
基づく数学教育への提言. [https://www.
mathsoc.jp/comm/kyoiku/chousa2011/index.
html](https://www.mathsoc.jp/comm/kyoiku/chousa2011/index.html) (2022年5月8日参照)
- NTT 研究所(2021) 「NTT 語彙データベース」
<https://www.nttprint.com/lexicon-db/> (2022
年5月8日参照)
- 清水裕士(2016) フリーの統計分析ソフト HAD-
機能の紹介と統計学習・教育, 研究実践にお
ける利用方法の提案-. メディア・情報・コ
ミュニケーション研究, 1, 59-73
- 杉本亜由美(2019) 敬語教育における効果的なア
クティブラーニングの研究. 人間生活文化研
究, 29, 287-292
- 高橋信(2021) IRT 項目反応理論入門—統計学の
基礎から学ぶ良質なテストの作り方—. オー
ム社
- 宇佐美慧・荘島宏二郎・光永悠彦・登藤直弥
(2019) 項目反応理論 (IRT) の考え方と実践
——測定の高いテストや尺度を作成す
るための技術——. 教育心理学年報, 58, 321-
329
- 山路奈保子・因京子・藤本裕行(2013) 日本人
大学生の書き言葉習得——初年次と3年次にお
ける調査結果の比較から——. 専門日本語教
育研究, 15, 47-52

