

項目応答理論を用いた 「OIDAI英語力診断テスト」の研究開発

松 宮 新 吾

Validation and Implementation of the OIDAI Diagnostic Test of English Through Item Response Theory

Shingo MATSUMIYA

要 旨

本研究では、2022 年 4 月に追手門学院大学 1 年生を対象として実施した「OIDAI 英語力診断テスト 2022」の結果を、古典的テスト理論による項目分析と、項目応答理論の枠組みに基づく Three Parameter Logistic Model（スリー・パラメータ・ロジスティックモデル）による分析とを平行して実施し、それぞれの分析結果の比較検証を行うプロセスを通じて、診断テストの質的改善を図ることを主たる目的として考察を行った。

その結果、Three Parameter Logistic Model から導き出された項目特性曲線による個々のテスト項目の特性と、項目分析結果とを照合することにより、各テスト項目の改善や、テスト項目群、また、テスト全体の構成やデザインを行うために必要なエビデンスを得ることができた。

（キーワード）

英語力診断テスト（English Diagnostic Test）、古典的テスト理論（Classical Test Theory）、項目応答理論（Item Response Theory）、スリー・パラメータ・ロジスティックモデル（Three Parameter Logistic Model）、項目特性曲線（Item Characteristic Curve）

1. はじめに

大学生のリスニングとリーディングによる英語運用能力を測定するためのツールとして、「OIDAI 英語力診断テスト」を自主開発し、2021 年から追手門学院大学生を対象に診断テストを実施している。本研究は、2022 年に追手門学院大学に入学した 1 年生を対象に実施した「OIDAI 英語力診断テスト」の結果をベースに、テスト開発におけるテスト問題の質的向上を実現するために、項目応答理論に基づきテスト項目の質的分析を行うものである。

2. 古典的テスト理論に基づく「OIDAI 英語力診断テスト」の分析結果概要

2021 年 4 月に新入生 1,685 名が受験した「OIDAI 英語力診断テスト 2021」の結果を、古典的テスト理論（Classical Test Theory）による項目分析（Item Analysis）を行った松宮（2022）¹ は、自主開発した英語力診断テストの質的分析結果を以下のとおり報告している。

（1）項目難易度について

診断テストやプレースメントテストに求められる項目難易度については、文法・語法問題と読解問題の項目難易度が高くなっている（通過率が低くなっている）ものの、ほぼ妥当であること（.3 ～ .6）が判明した。

（2）弁別力について

弁別力については、上位下位項目弁別力指数（成績上位群 27%、下位群 27% による項目難易度差）と点双列相関係数による項目弁別力指数を算出し検証を行った。その結果、一般的に、.3 以上が求められる点双列相関係数による項目弁別力指数においては、23 項目が基準を下回っていることが判明した。

（3）信頼性について

診断テストの信頼性（内部一貫性）を検証するために、クロンバックの α 係数を算出した。その結果、各テスト項目カテゴリの信頼係数は、.62 から .79 を示し、診断テストの内部一貫性は満足できるレベルであることが確認された。

（4）妥当性について

開発した英語力診断テストの妥当性を、TOEIC L&R のテストスコアを用いた併存的妥当性

1 松宮新吾（2022）. 「項目分析と古典的テスト理論に基づく「OIDAI 英語力診断テスト」の研究開発」『国際教養学部紀要』、15：pp. 1-21、追手門学院大学国際教養学部.

(Concurrent Validity) により検証した結果、相関係数は、 $r = .714$ ($p < .000$) で、プラスのかなり強い相関があることが判明した。また、算出された相関係数(妥当性係数)から決定係数を導き出すと、.51 となり、開発した診断テストの妥当性は、十分に高いことが判明した。

上記で示したとおり、テスト開発に際しては、言語運用能力測定に用いる診断テストが本当にテストの備えるべき条件を満たし、精度の高い測定を行うことができるツールとなっているのかということを、項目分析から得られる情報に基づき検証してきた。

その結果、古典的テスト理論で求められる信頼性、難易度、弁別力、妥当性等の各指標については、一定レベルの基準を満たす診断テストを開発し、学生の英語運用能力を設定した4つの能力カテゴリ(語彙・表現力、文法・語法力、リスニング力、読解力)に基づき、満足できる精度で測定することができたものと評価した。

しかし、当該の研究で用いたテスト理論と項目分析の枠組みは、古典的テスト理論に依拠したものであるため、標本依存やテスト依存をはじめとする分析の限界と困難さを併存したものとなっていた。そこで、本研究においては、古典的テスト理論による項目分析の限界を補完する目的で、「項目応答理論」(Item Response Theory) による分析方法を用いて、自主開発している「OIDAI 英語力診断テスト」のテスト項目の特性をより詳細に把握し、各テスト項目、および、診断テスト全体の質的向上を図るために必要な情報を収集する。

3. 古典的テスト理論に基づく項目分析の限界

古典的テスト理論に基づく項目分析の限界は、すでに概観したように、テスト項目の困難度は「標本依存」(Sample-dependent) であること、また、被験者の能力の決定は「テスト依存」(Test-dependent) であることに集約することができる。

また、項目弁別力や弁別力指数も、被験者を構成する被験者集団の特性によって相対的に決定される値であるため、その項目の持つ絶対的な困難度や弁別力指数を求めることは、古典的テスト理論では困難となる。

すなわち、被験者を構成する受験集団の特性によって、項目困難度や項目弁別力指数(点双列相関係数)が相対的に決定され、テスト項目自体の持っている困難度と被験者の能力によってテスト項目の特性が相対的に決定されるという構造になっているのが、古典的テスト理論の特性である。そのため、古典的テスト理論で求めたデータは、そのデータを作成したものと同一のテスト項目が、データ収集のためにそのテストを受験した集団とまったく同じ集団にもう一度与えられた場合にだけ、意味のある情報を得ることができるという限界がある。

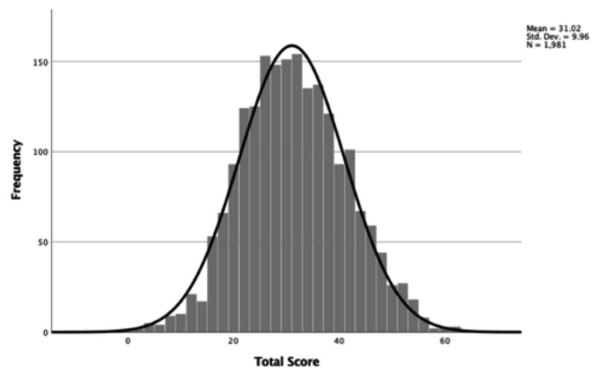
4. 「OIDAI 英語力診断テスト 2022」の実施概要

2022 年 4 月 1 日（金）から 4 月 5 日（火）までの 5 日間、新入生を対象とし、Web 上で「OIDAI 英語力診断テスト 2022」を実施した。

有効受験者数は 1,981 名で、4 月 1 日時点の入学予定者名簿に登録されているクラス編成対象者 2,084 名の 95.1% が診断テストの受験をした。その概要、および、基本統計量は表 1、図 1 に示すとおりである。

（表 1：「OIDAI 英語力診断テスト 2022」概要）

「OIDAI 英語力診断テスト 2022」基本統計量	
受験者数	1981
平均	31.02
標準偏差	9.96
分散	99.20
歪度	0.13
尖度	-0.14
最高点	63.00
最低点	0.00
総問題数	67
リスニング問題数	41
リーディング問題数	16
語彙・表現問題数	5
文法・構文問題数	5



（図 1：「OIDAI 英語力診断テスト 2022」度数分布表）

5. 古典的テスト理論に基づく「OIDAI 英語力診断テスト 2022」の項目分析結果

5.1 項目難易度の検証

診断テストやプレースメントテストに求められる項目難易度については、診断テスト全体としては、ほぼ妥当であること（.46）が判明した。しかし、テスト項目 67 項目を、測定する能力種別に分類した下位項目の難易度を算出した結果、語彙・表現（.24）と文法・構文（.29）に関わる能力を測定するための項目群の難易度が高くなりすぎていることが分かった。また、下位構成項目群として、リーディング問題（.39）の項目難易度も、高い傾向である（通過率が低くなっている）ことが判明した（表 2）。

5.2 弁別力の検証

上位下位項目弁別力指数においては、.5 に近い項目弁別力を示しているリーディング（読解）問題群を除いて、.4 を下回っており、一部難易度の調整が必要であることが判明した。特に、下

位構成項目群として算出した語彙・表現（.23）、文法・構文項目群（.28）においては、.3を下回っており、項目内容の質的検討が求められる。

また、一般的に .3 以上が求められる点双列相関係数による項目弁別力指数においても同様に、下位構成項目である語彙・表現（.22）、文法・構文項目群（.26）については、基準値を下回っていることが分かった。

なお、項目弁別力の算出には、上位下位項目弁別力指数（成績上位群 27%、下位群 27% による項目難易度差）と点双列相関係数による項目弁別力指数を用いた。

（表 2：「OIDAI 英語力診断テスト 2022」の項目分析概要）

項目分析	項目難易度	最適項目困難度	項目困難度適切度	項目弁別力	点双列相関係数	
テスト項目全体	0.463	0.605	0.569	0.370	0.320	**
リスニング	0.534	0.606	0.615	0.337	0.301	**
リーディング	0.386	0.603	0.566	0.494	0.398	**
語彙・表現	0.240	0.600	0.280	0.233	0.220	**
文法・構文	0.292	0.600	0.384	0.280	0.257	**
聴解	0.466	0.608	0.599	0.345	0.305	**
読解	0.389	0.605	0.569	0.500	0.401	**

5.3 信頼性の検証

信頼性の一つの指標となる内部一貫性を検証するために、クロンバックの α 係数を算出した。その結果、テスト項目全体としては、.89 であり、開発した診断テストの内部一貫性の程度は満足できるレベルであることが確認された。また、テストの下位構成項目として、リスニング（聴解）とリーディング（読解）能力を測定するための項目群を分離して信頼係数を算出した結果、リスニング項目群が .79、リーディング項目群が .77 となり、十分な信頼性を有していることが確認できた。

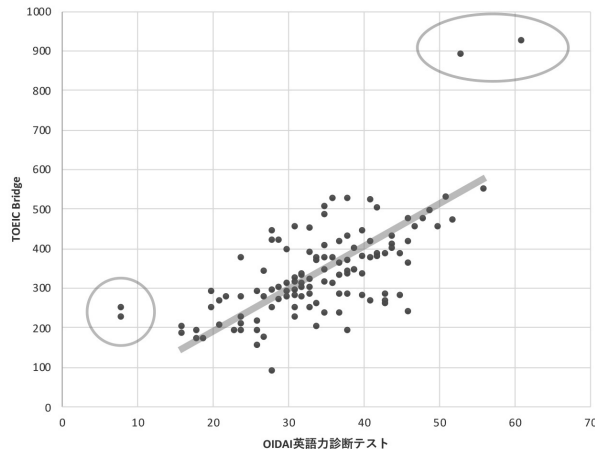
5.4 妥当性の検証

国際学部 of 1 年生が、診断テストと同時期に受験する TOEIC Bridge のテストスコアを、基準関連妥当性を算出するための外的基準として採用した。

その結果、診断テストと外的基準としての TOEIC Bridge のテストスコアとの相関係数は、 $r = .665$ ($p < .000$) で、プラスの中程度の相関があることが判明した。また、これによる決定係数は、.44 で、中程度の説明力を有していることが分かった。これらの分析結果から、診断テストの妥当性は一定レベルのものであると結論付けることができた。

参考までに、図 2 に OIDAI 英語力診断テストと同時期に受験した TOEIC Bridge テストのテストスコアの散布図を示す。散布図から、TOEIC Bridge のスコア上位得点者と下位得点者において、外れ値が存在していることが判明した。いずれの場合も、TOEIC Bridge においてより高得点側に

振れていることが確認できた。それ以外の被験者については、直線状に当てはまりの良い状態であることが判明した。このことから、一部、上位層・下位層に対する問題項目の妥当性を詳細に検討する必要があることが分かった。



(図 2：OIDAI 英語力診断テストと国際学部 1 年生の TOEIC Bridge テストスコアとの散布図)

5.5 古典的テスト理論に基づく考察

古典的テスト理論に基づく項目分析の結果を表 3 に示す。表 3 では、各質問項目の属性を L (リスニング問題：Listening)、R (リーディング問題：Reading)、V (語彙・表現問題：Vocabulary & Expressions)、G (文法・構文問題：Grammar & Structure) として示している。

また、項目特性においては、項目困難度 (Item Difficulty) と、項目弁別力 (Item Discrimination Power) と、点双列相関係数 (Point Biserial Correlation Coefficient) による項目弁別力指数 (Item Discrimination Power Index) について、それぞれ妥当とされている基準値から大きく逸脱している「外れ値」を示す項目について、その評価と特性をアルファベットによる記号表記で示している (表 4-1、4-2、4-3)。

項目分析の結果から、項目難易度のみに問題のある項目が 3 項目、項目難易度と弁別力に問題のある項目が 14 項目、弁別力に問題のある項目が 3 項目、延べ 20 項目において問題項目の質や適切さを検証・改善する必要があることが判明した。

その結果、写真描写や短い対話に対する応答問題においては通過率が高く、リスニング能力の低い被験者の正答率が高くなっており、リスニング能力の弁別を行う上で十分な指標とならない問題が配列されていることが判明した。一方、ストーリー性を有し一定の長さがあるスニング問題においては、難易度が高くなっていることが判明した。また、文法・構文問題においても、難易度が高く、能力の高い群と低い群において正答率の差が小さく、弁別力が十分に高くなっていない項目があることが確認できた (表 3、表 4-3)。

(表 3：項目分析結果)

項目	項目属性	選択肢数	正答数	回答者数	項目困難度	最適項目困難度	項目困難度適切度	項目弁別力	点双列相関係数	項目特性
22A1	LP	4	1546	1981	0.78	0.63	0.69	0.34	.302**	
22A2	LP	5	1665	1981	0.84	0.60	0.52	0.23	.261**	D(E)/DP
22A3	LP	5	1358	1981	0.69	0.60	0.82	0.36	.317**	
22A4	LD	4	1413	1981	0.71	0.63	0.83	0.35	.317**	
22A5	LD	4	991	1981	0.50	0.63	0.75	0.50	.390**	
22A6	LD	4	778	1981	0.39	0.63	0.53	0.31	.287**	
22A7	LD	4	498	1981	0.25	0.63	0.25	0.13	.132**	D(D)/DP
22A8	LD	4	1263	1981	0.64	0.63	0.97	0.22	.186**	DP
22A9	LR	5	1547	1981	0.78	0.60	0.64	0.42	.417**	
22A10	LR	5	906	1981	0.46	0.60	0.72	0.65	.495**	
22A11	LR	5	1190	1981	0.60	0.60	1.00	0.43	.352**	
22A12	LR	5	1380	1981	0.70	0.60	0.80	0.42	.374**	
22A13	LR	5	1569	1981	0.79	0.60	0.62	0.26	.257**	D(E)/DP
22A14	LR	5	1338	1981	0.68	0.60	0.84	0.42	.351**	
22A15	LR	5	1164	1981	0.59	0.60	0.98	0.50	.389**	
22A16	LR	5	1628	1981	0.82	0.60	0.56	0.21	.251**	D(E)/DP
22A17	LR	5	1763	1981	0.89	0.60	0.42	0.27	.349**	D(E)/DP
22A18	LR	5	737	1981	0.37	0.60	0.54	0.27	.226**	
22A19	LR	5	1312	1981	0.66	0.60	0.88	0.25	.221**	
22A20	LR	5	1637	1981	0.83	0.60	0.54	0.22	.218**	D(E)/DP
22A21	LR	5	1423	1981	0.72	0.60	0.76	0.33	.304**	
22A22	LR	5	1596	1981	0.81	0.60	0.58	0.37	.381**	D(E)
22A23	LR	5	1578	1981	0.80	0.60	0.60	0.42	.424**	D(E)
22A24	LR	5	1088	1981	0.55	0.60	0.90	0.30	.239**	
22A25	LR	5	684	1981	0.35	0.60	0.50	0.47	.400**	
22A26	LR	4	754	1981	0.38	0.63	0.51	0.21	.174**	DP
22A27	LR	4	532	1981	0.27	0.63	0.29	0.24	.220**	
22A28	LR	4	653	1981	0.33	0.63	0.41	0.29	.257**	
22A29	LR	4	375	1981	0.19	0.63	0.13	####	-0.033	D(D)/DP
22A30	LS	5	530	1981	0.27	0.60	0.34	0.34	.318**	D(D)/DP
22A31	LS	5	439	1981	0.22	0.60	0.24	0.31	.308**	D(D)/DP
22A32	LS	5	434	1981	0.22	0.60	0.24	0.05	.076**	D(D)/DP
22A33	LS	5	893	1981	0.45	0.60	0.70	0.30	.255**	
22A34	LS	1	1078	1981	0.54	0.60	0.88	0.52	.410**	
22A35	LS	1	971	1981	0.49	0.60	0.78	0.60	.469**	
22A36	LS	1	782	1981	0.39	0.60	0.58	0.16	.159**	DP
22A37	LS	1	981	1981	0.50	0.60	0.80	0.54	.433**	
22A38	LS	1	420	1981	0.21	0.60	0.22	0.28	.302**	D(D)
22A39	LS	1	901	1981	0.45	0.60	0.70	0.49	.396**	
22A40	LS	1	954	1981	0.48	0.60	0.76	0.48	.392**	
22A41	LS	1	575	1981	0.29	0.60	0.38	0.42	.381**	
22A42	V	1	515	1981	0.26	0.60	0.32	0.20	.207**	
22A43	V	1	301	1981	0.15	0.60	0.10	0.15	.182**	D(D)/DP
22A44	G	1	1014	1981	0.51	0.60	0.82	0.47	.373**	
22A45	V	1	620	1981	0.31	0.60	0.42	0.31	.271**	
22A46	G	1	521	1981	0.26	0.60	0.32	0.33	.319**	
22A47	G	1	501	1981	0.25	0.60	0.30	0.32	.309**	
22A48	G	1	423	1981	0.21	0.60	0.22	0.13	.134**	D(D)/DP
22A49	G	1	458	1981	0.23	0.60	0.26	0.17	.150**	D(D)/DP
22A50	R	1	485	1981	0.24	0.60	0.28	0.24	.241**	D(D)/DP
22A51	R	1	694	1981	0.35	0.60	0.50	0.50	.420**	
22A52	R	1	784	1981	0.40	0.60	0.60	0.53	.439**	
22A53	R	1	604	1981	0.30	0.60	0.40	0.35	.315**	
22A54	R	1	825	1981	0.42	0.60	0.64	0.64	.495**	
22A55	R	1	758	1981	0.38	0.60	0.56	0.50	.417**	
22A56	R	1	697	1981	0.35	0.60	0.50	0.53	.435**	
22A57	R	1	700	1981	0.35	0.60	0.50	0.24	.210**	
22A58	R	1	798	1981	0.40	0.60	0.60	0.54	.422**	
22A59	V	1	823	1981	0.42	0.58	0.67	0.68	.539**	
22A60	R	1	1089	1981	0.55	0.63	0.85	0.55	.440**	
22A61	R	1	784	1981	0.40	0.63	0.55	0.55	.437**	
22A62	R	1	802	1981	0.40	0.60	0.60	0.56	.445**	
22A63	V	1	587	1981	0.30	0.60	0.40	0.23	.217**	
22A64	R	1	595	1981	0.30	0.60	0.40	0.46	.394**	
22A65	R	1	924	1981	0.47	0.60	0.74	0.65	.500**	
22A66	R	1	823	1981	0.42	0.60	0.64	0.46	.359**	
22A67	R	1	998	1981	0.50	0.63	0.75	0.57	.443**	

** p<.01

(表 4-1：質問項目属性一覧)

記号	項目属性
LP	リスニング問題(ピクチャー表現問題)
LD	リスニング問題(対話問題)
LR	リスニング問題(聞き・読み問題)
LS	リスニング問題(ストーリー問題)
V	語彙・表現問題
G	文法・構文問題
R	リーディング問題(読解問題)

(表 4-2：項目特性)

記号	項目特性
D(D)	項目難易度が.24以下で、極端に難しい項目
D(E)	項目難易度が.8以上で、極端に易しい項目
DP	項目弁別力指数が.2以下で弁別力が極端に弱い項目

(表 4-3：項目特性と項目番号)

記号	項目番号
D(E)/DP	A2, A13, A16, A17, A20
D(D)/DP	A7, A29, A30, A31, A32, A43, A48, A49, A50
D(E)	A22, A23
D(D)	A28
DP	A8, A26, A36

このように、古典的テスト理論を用いた分析においては、得点分布から、全体的なスコアの散らばり具合を確認し、天井効果やフロアー効果の有無を検討することや、内部一貫性を確認するために、クロンバックの α 係数や相関係数を計算し、テスト全体の信頼性を確認したり、正答率から項目弁別力や点双列相関係数を算出し、不当に難しい、あるいは、易しい問題を特定し、今後のテスト項目の質的改善を行うために必要な情報を得ることが可能である。また、TOEIC Bridge テスト等の外部テストスコアとの相関係数を算出し、開発した診断テストの併存的妥当性を検討することが可能となる。

しかし、いずれの場合も標本依存型、場依存型の情報しか得ることができず、被験者集団や受験するテスト項目が異なると、それぞれ別個の異なった情報が計算され提示されることになる。そこで、本研究では、項目応答理論を用いた分析を古典的テスト理論に基づく項目分析と併用することにより、「OIDAI 英語力診断テスト」の各テスト項目の特性と課題をより明確に「あぶり出し」、診断テストとしての質的向上を図る有意な情報を得るための考察を行うこととする。

6. 項目応答理論について

項目応答理論では、被験者の能力推定値は、その被験者が受験する特定のテスト項目の特性とは切り離して、独立して求めることができるとされている。また、項目応答理論においては、被験者の能力、テスト項目の困難度、正答率などの間の関係を示すロジスティックモデルによる項目特性曲線 (Item Characteristic Curve) を用いて、テスト項目の特性と被験者の能力との関係を検証することができるとされている。すなわち、項目応答理論においては、古典的テスト理論の「正答数に基づく得点」(Number-right Score) を能力として見なすのではなく、観測されたテスト項目に対する応答結果と、現実には直接観測することができない仮説の構成体である「潜在特性」(Latent Trait) を示す尺度とをリンクさせるロジスティックモデルを構築し、被験者の能力を推定しようとするものである。

これにより、項目応答理論においては、被験者の能力測定は、どのようなテスト項目を用いても一定 (不変) なものとして測定することが可能であるとされている。また、テスト項目の困難

度や弁別力、信頼性等は、どのような集団に対しても一定（不変）なものとして算出することができる。とされている。

この一定性（不変性）が意味していることは、大友（1996）²によると、「テスト項目を特徴づけているパラメータは『被験者の持っている能力の分布に依存して決定されるものではない』ということである。」また、同時に「被験者の能力を特徴づけるパラメータは、『テスト項目に依存して決定されるものではない』ということである。」と説明されている。

さらに、テスト項目の精度を被験者毎に定義することも可能となる。すなわち、項目応答理論の精度は、「項目情報関数」（Item/Test Information Function）で示される情報に基づき、被験者毎に測定精度が算出されることになる。

6.1 パラメータの推定

本研究においては、ロジスティックモデルの精度の指標となる Akaike Information Criterion (AIC) と Bayesian Information Criterion (BIC) が最小の値を示すモデルを採用し分析を行うこととする。また、開発した質問項目は4択問題を中心とした多肢選択式の問題であるため、「当て推量 (Guessing)」の可能性を考慮した、Three Parameter Logistic Model を採用する。すなわち、本研究で採用する Three Parameter Logistic Model では、「項目困難度パラメータ」(Item Difficulty Parameter) と「能力パラメータ」(Ability Parameter) と、「項目弁別力パラメータ」(Item Discrimination Parameter) と、さらに、「当て推量パラメータ」(Guessing Parameter) を用いて、「OIDAI 英語力診断テスト」のテスト項目の分析と検討を行う。

6.2 項目応答理論の前提に基づくテスト項目の作成

項目応答理論の前提となる概念として、「局所独立の仮定」(Assumption of Local Independence) と、「一次元性」(Unidimensionality) がある。「OIDAI 英語力診断テスト」の開発においては、これら2つの概念を反映したテスト項目の作成を行った。

まず、各項目間の正答確率は統計的には互いに独立しているという「局所独立の仮定」に従い、被験者が一つのテスト項目に正答する確率が、他の項目の正答確率に影響を及ぼさないようテスト項目を設定した。具体的には、3つのテスト項目により構成されているリスニングテストで、項目3に正解するためには、項目1、2が正解できなければならないような設問がされている場合には、互いに独立しているとはいえないのである。

2 大友賢二（1996）.『項目応答理論入門』大修館書店, p. 141.

(例 1)「局所独立の仮定」を満たさないテスト項目群の事例

項目 1:「〇〇の開発が始まったのは何年か?」

項目 2:「〇〇の開発が終了したのは何年前か?」

項目 3:「〇〇の開発には何年かかったか?」※項目 1、2 の正解が項目 3 の正解を導くため、「局所独立の仮定」が成り立たない。

次に、テスト項目が特定の能力分野を測定するものとして妥当であることを保証する「一次元性」を担保するために、複数の能力分野にまたがる項目や、関係性の弱い項目を同一項目群に設定しないよう質問項目を配列した。このことは、各項目群における信頼係数や相関係数により判定することもできるが、実際には各項目を因子分析し、その因子負荷量により判定するものである。

(例 2)「一次元性」を満たさないテスト項目群の事例

項目 1: 文中の語○の示す意味を選択せよ。※語彙・表現の意味理解を評価する項目

項目 2: 下線部 A、B、C、D で語順に誤りがあるものを選択せよ。※連語関係（文法）の理解を評価する項目

項目 3: 文中の空所に入る語句を選択せよ。※文脈と語彙・表現の理解を評価する項目

このように項目応答理論の前提に基づき各質問項目を設定し分析することにより、被験者の同一の潜在特性や能力を高い精度で測定することができるようになる。また、同一のテスト項目を異なる被験者集団で実施したとしても、その項目の持っている特性の値に変わりはないという結果を得ることができる。これにより、被験者の能力推定値を、テスト項目の特性とは切り離して独立して求めることができるテスト開発が可能となる。

7. 結果と考察

本節では、2022 年 4 月に実施した「OIDAI 英語力診断テスト」を、Three Parameter Logistic Model を用いて分析した結果に基づき考察を行う。

まず、Rasch Model（ワン・パラメータ・ロジスティックモデル）と Three Parameter Logistic Model によるモデルの精度（当てはまり度合い）を、Global Fit Statistics として算出される数値から判定した。その結果、本研究では、AIC が最小となる Three Parameter Logistic Model を採用することとした（表 5）。

(表 5 : Global Fit Statistics の数値)

Three Parameter Item Response Model	
Akaike Information Criterion (AIC)	151178.97
Bayesian Information Criterion (BIC)	152302.83
Rasch Model (One Parameter Item Response Model)	
Akaike Information Criterion (AIC)	153263.80
Bayesian Information Criterion (BIC)	153644.01

7.1 当て推量パラメータ (Guessing Parameter)

当て推量パラメータ (GP) は、能力推定値が低い被験者が多肢選択式のテスト項目に回答する際に生じる当て推量の確率を推定し、説明するために設けられたパラメータである。ここで、算出された当て推量パラメータが、.0 から .4 の範囲 ($.0 < GP < .4$) に収まっていれば、各項目における被験者の推量による回答、および、能力推定に及ぼす影響は問題がないと判断した。「OIDAI 英語力診断テスト 2022」の各項目の当て推量パラメータは .39 以内であるため、当て推量による影響は限定的であると判定した (表 6、図 3)。

しかし、質問項目 1 ($A1 = .39$)、項目 2 ($A2 = .36$) と項目 36 ($A36 = .38$) の数値は、他の項目と比較して高く、項目特性曲線をさらに詳細に検証する必要が認められた。

7.2 項目困難度 (Item Difficulty Parameter) パラメータ

Three Parameter Logistic Model による項目困難度 (IDF) パラメータの値は、プラス無限大からマイナス無限大の値をとるが、一般的に困難度パラメータは、-3 から +3 までの幅 ($-3 < IDF < +3$) で示される。すなわち、-3 が最も易しく、+3 が最も難しい項目であると判定される。したがって、難しくも易しくもない普通の項目は 0 となる。

項目困難度パラメータの係数表 (表 7)、および、グラフ (図 4) から判断して、項目 A7、A29、A32、A43、A49 の 5 項目が想定される数値から逸脱していることが判明した。特に項目 32 の標準誤差は 52.726 と他の係数と比較して極端に大きくなっていることが明らかになった。このことから抽出した 5 項目の項目困難度については、項目特性曲線をさらに詳細に分析することが求められる。

7.3 項目弁別力 (Item Discrimination Parameter) パラメータ

Three Parameter Logistic Model による項目分別力 (IDS) パラメータの値は、一般的に .3 よりも大きな値をとることが被験者の能力差を識別する弁別力の良さを示す基準 ($.3 < IDS$) とされている。しかし、この値が 1.0 を超える場合には、項目難易度が難しすぎるために、過度に弁別をしていることが示唆されるものである。また、項目弁別力パラメータの値が -1.0 より小さい場合

には、項目難易度が過度に易しく、項目の弁別力が適切でないことが示されている（表 8、図 5）。

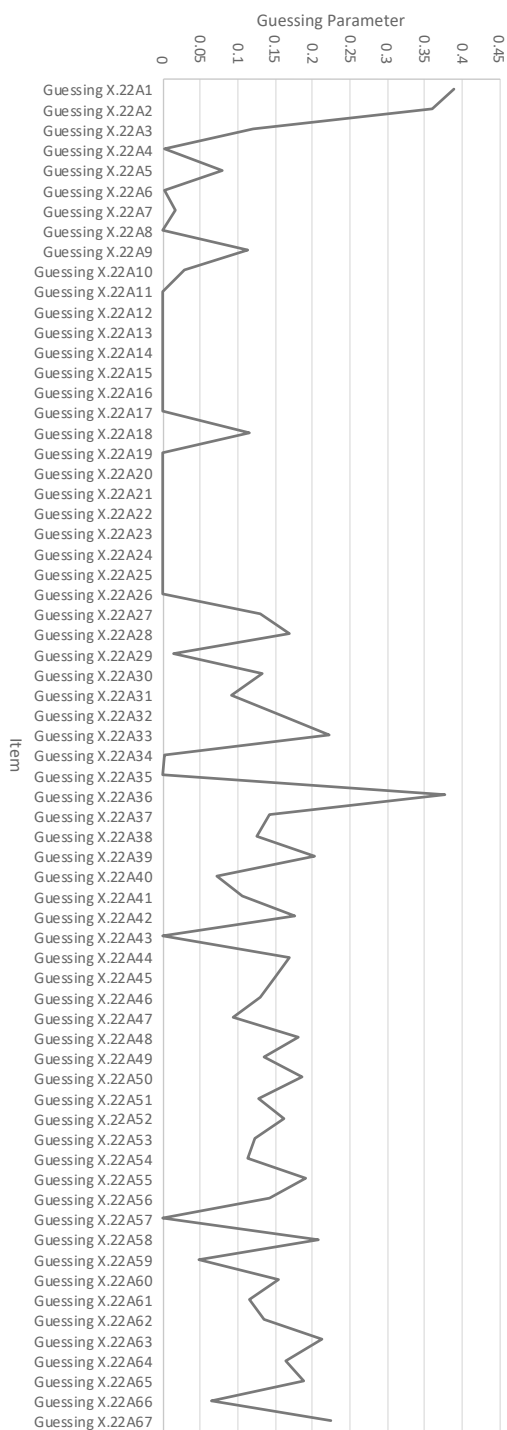
分析の結果、35 項目の項目弁別パラメータが 1.0 を超えており、過半数の項目難易度が高く、過度に弁別を行っている可能性があることが確認された。このことから「OIDAI 英語力診断テスト 2022」のテスト全体の難易度を調整する必要があることが明らかとなった。

7.4 項目特性曲線 (Item Characteristic Curve)

本節では、Three Parameter Logistic Model を使って作成される項目特性曲線を用いて、各パラメータが示す数値の意味を視覚的に確認し、それぞれの項目の特性をより分析的に把握する。そのために用いる特性曲線として、各項目の特性を表す Item Characteristic Curve（項目特性曲線）と、各項目で測定することができる能力の幅（バンド）と情報量を示す Item Information Curve（項目情報曲線）と、被験者全体の能力指標の推定分布を表す Kernel Density Estimation for Ability Estimates（カーネル密度推定）、さらに、各項目に対する被験者の正答率と得点をプロットした得点正答率分布図を参照する。なお、本論文への掲載は、項目応答理論に基づき算出された各パラメータの分析結果から、特色のある 3 つのパターンを有する項目群のみを取り上げ記述する。

(表 6：当て推量パラメータの係数表)

Item	Statistic Value	Standard Error	Z
A1	0.390	0.116	3.364
A2	0.360	0.308	1.168
A3	0.121	0.250	0.485
A4	0.001		
A5	0.079	0.090	0.882
A6	0.002	0.047	0.039
A7	0.017	0.129	0.135
A8	0.000	0.004	0.015
A9	0.112	0.137	0.822
A10	0.027	0.043	0.641
A11	0.000	0.000	0.002
A12	0.000	0.002	0.005
A13	0.000	0.000	0.002
A14	0.000	0.000	0.001
A15	0.000	0.000	0.000
A16	0.000	0.001	0.003
A17	0.000	0.001	0.002
A18	0.115	0.203	0.569
A19	0.000	0.004	0.014
A20	0.000	0.003	0.010
A21	0.000	0.000	0.001
A22	0.000	0.000	0.003
A23	0.000	0.001	0.002
A24	0.000	0.002	0.008
A25	0.001		
A26	0.000	0.009	0.049
A27	0.131	0.099	1.331
A28	0.168	0.038	4.402
A29	0.015	0.035	0.421
A30	0.133	0.027	5.005
A31	0.093	0.022	4.245
A32	0.156	0.184	0.850
A33	0.223	0.100	2.227
A34	0.001		
A35	0.000	0.001	0.001
A36	0.378	0.014	27.333
A37	0.144	0.059	2.434
A38	0.126	0.016	7.975
A39	0.204	0.034	6.034
A40	0.073	0.108	0.682
A41	0.105	0.027	3.968
A42	0.176	0.040	4.456
A43	0.000	0.020	0.024
A44	0.170	0.057	2.960
A45	0.149	0.041	3.635
A46	0.131	0.021	6.096
A47	0.094	0.028	3.302
A48	0.182	0.016	11.385
A49	0.134	0.076	1.771
A50	0.187	0.014	12.965
A51	0.129	0.025	5.052
A52	0.162	0.024	6.791
A53	0.124	0.038	3.239
A54	0.114	0.024	4.779
A55	0.191	0.022	8.734
A56	0.142	0.021	6.876
A57	0.000	0.005	0.023
A58	0.207	0.023	9.130
A59	0.048	0.022	2.154
A60	0.154	0.056	2.751
A61	0.115	0.033	3.488
A62	0.134	0.027	5.016
A63	0.212	0.028	7.529
A64	0.164	0.016	10.098
A65	0.187	0.026	7.195
A66	0.064	0.075	0.859
A67	0.225	0.029	7.723

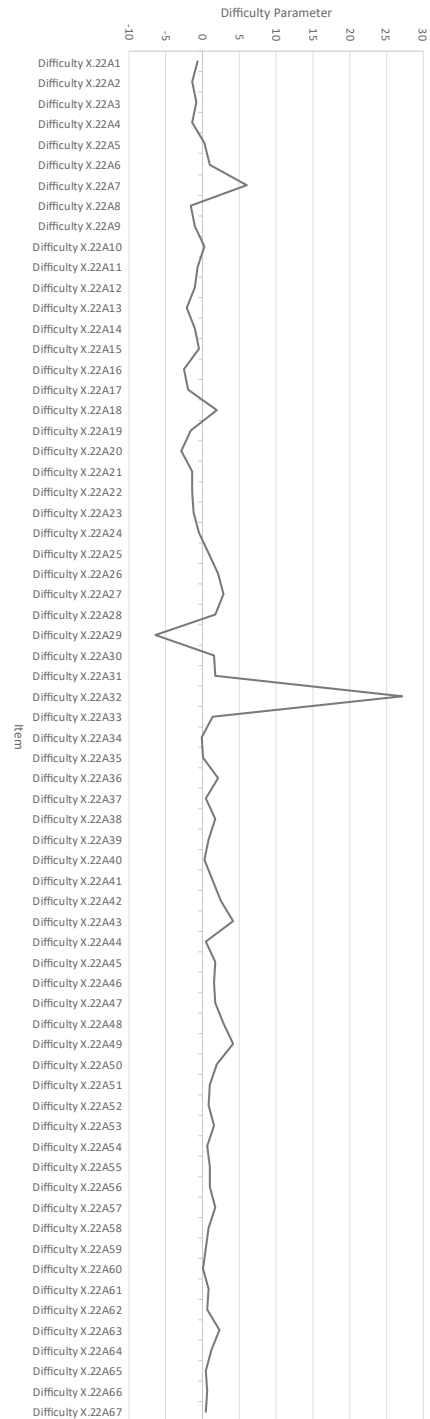


(図 3：項目別当て推量パラメータ)

項目応答理論を用いた「OIDAI 英語力診断テスト」の研究開発

(表 7：項目困難度パラメータの係数表)

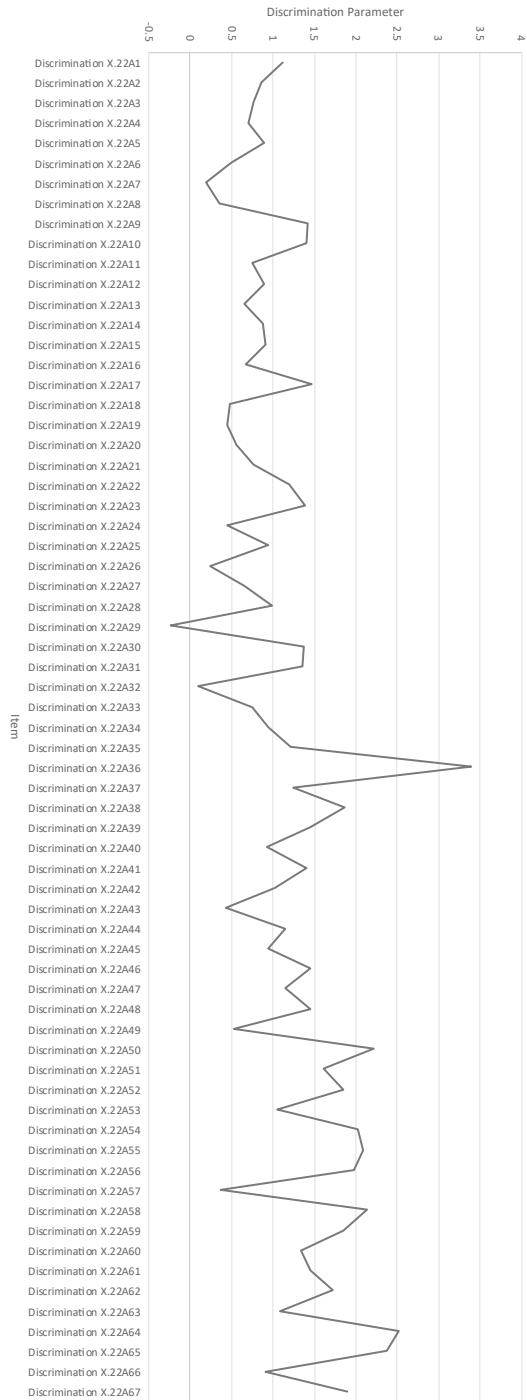
Item	Statistic Value	Standard Error	Z
A1	-0.634	0.399	-1.587
A2	-1.474	1.120	-1.316
A3	-0.853	0.801	-1.064
A4	-1.435		
A5	0.232	0.261	0.887
A6	0.939	0.230	4.078
A7	6.048	1.891	3.198
A8	-1.680	0.277	-6.075
A9	-1.059	0.264	-4.009
A10	0.241	0.099	2.449
A11	-0.608	0.080	-7.614
A12	-1.082	0.089	-12.164
A13	-2.219	0.208	-10.664
A14	-0.969	0.084	-11.465
A15	-0.450	0.064	-6.998
A16	-2.463	0.230	-10.695
A17	-1.920	0.107	-17.908
A18	1.960	0.745	2.632
A19	-1.554	0.202	-7.678
A20	-2.968	0.332	-8.937
A21	-1.372	0.117	-11.705
A22	-1.493	0.092	-16.317
A23	-1.307	0.074	-17.560
A24	-0.464	0.119	-3.915
A25	0.814	0.069	11.782
A26	2.030	0.456	4.452
A27	2.759	0.439	6.281
A28	1.713	0.158	10.821
A29	-6.454	1.706	-3.784
A30	1.628	0.112	14.565
A31	1.732	0.116	14.977
A32	27.133	52.726	0.515
A33	1.314	0.332	3.958
A34	-0.213		
A35	0.054	0.049	1.113
A36	2.137	0.144	14.869
A37	0.389	0.151	2.569
A38	1.759	0.102	17.161
A39	0.749	0.097	7.758
A40	0.314	0.303	1.039
A41	1.291	0.088	14.604
A42	2.495	0.338	7.381
A43	4.118	0.651	6.322
A44	0.404	0.161	2.508
A45	1.790	0.167	10.700
A46	1.597	0.107	14.937
A47	1.672	0.124	13.465
A48	2.807	0.389	7.218
A49	4.133	1.261	3.277
A50	1.860	0.113	16.467
A51	0.967	0.075	12.953
A52	0.809	0.068	11.950
A53	1.553	0.131	11.861
A54	0.549	0.060	9.097
A55	0.935	0.066	14.214
A56	0.924	0.063	14.555
A57	1.721	0.272	6.335
A58	0.888	0.066	13.399
A59	0.409	0.057	7.175
A60	0.142	0.140	1.017
A61	0.740	0.086	8.619
A62	0.699	0.070	9.915
A63	2.320	0.261	8.880
A64	1.195	0.061	19.623
A65	0.514	0.062	8.274
A66	0.664	0.207	3.216
A67	0.497	0.075	6.602



(図 4：項目困難度パラメータ)

(表 8：項目弁別力パラメータ)

Item	Statistic Value	Standard Error	Z
A1	1.115	0.197	5.655
A2	0.859	0.222	3.877
A3	0.768	0.176	4.363
A4	0.699	0.040	17.275
A5	0.899	0.148	6.076
A6	0.498	0.067	7.432
A7	0.194	0.096	2.016
A8	0.345	0.052	6.696
A9	1.421	0.178	7.993
A10	1.397	0.149	9.396
A11	0.748	0.059	12.619
A12	0.890	0.066	13.381
A13	0.656	0.065	10.079
A14	0.872	0.065	13.416
A15	0.909	0.064	14.220
A16	0.678	0.069	9.892
A17	1.459	0.114	12.828
A18	0.481	0.287	1.675
A19	0.453	0.054	8.408
A20	0.559	0.066	8.417
A21	0.765	0.063	12.141
A22	1.198	0.086	13.982
A23	1.392	0.094	14.825
A24	0.442	0.052	8.544
A25	0.942	0.061	15.320
A26	0.244	0.050	4.843
A27	0.658	0.401	1.641
A28	0.984	0.212	4.647
A29	-0.241	0.070	-3.426
A30	1.368	0.262	5.212
A31	1.351	0.229	5.894
A32	0.093	0.253	0.367
A33	0.753	0.262	2.875
A34	0.938	0.058	16.132
A35	1.204	0.074	16.250
A36	3.398	1.674	2.030
A37	1.248	0.187	6.677
A38	1.868	0.326	5.737
A39	1.448	0.190	7.629
A40	0.922	0.183	5.034
A41	1.399	0.208	6.734
A42	1.026	0.413	2.484
A43	0.434	0.086	5.053
A44	1.141	0.165	6.930
A45	0.935	0.209	4.464
A46	1.443	0.231	6.237
A47	1.142	0.199	5.742
A48	1.454	0.494	2.942
A49	0.526	0.361	1.456
A50	2.216	0.464	4.779
A51	1.617	0.204	7.910
A52	1.854	0.218	8.518
A53	1.046	0.206	5.076
A54	2.023	0.203	9.982
A55	2.088	0.273	7.652
A56	1.977	0.229	8.643
A57	0.363	0.053	6.872
A58	2.141	0.284	7.545
A59	1.856	0.162	11.428
A60	1.347	0.170	7.901
A61	1.454	0.183	7.935
A62	1.719	0.197	8.737
A63	1.092	0.304	3.588
A64	2.522	0.371	6.790
A65	2.372	0.287	8.269
A66	0.912	0.165	5.514
A67	1.898	0.216	8.794

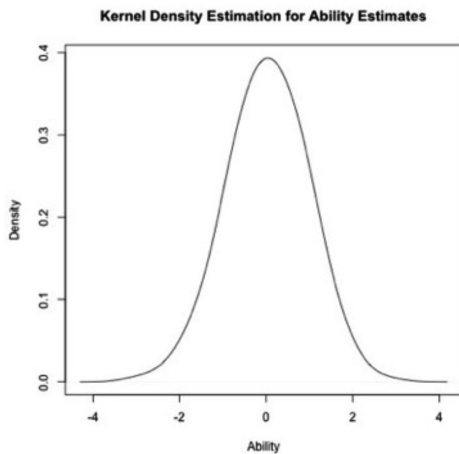


(図 5：項目弁別力パラメータ)

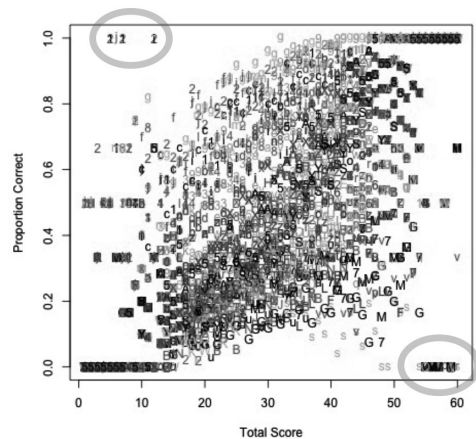
7.4.1 診断テスト全体の評価

被験者全体の能力指標の推定分布を表す Kernel Density Estimation for Ability Estimates（カーネル密度推定）（図 6）から、推定能力が -1 から +1 の範囲に位置している被験者の出現数が最大となっており、その分布が正規分布グラフに近い滑らかなものとなっていることが分かる。このことから、67 項目により構成されている「OIDAI 英語力診断テスト 2022」は、多様で幅の広い被験者の英語運用能力を推定することができていると考えられる。

次に、各項目に対する被験者の正答率と得点をプロットした得点正答率分布図（図 7）で、赤丸で示した領域にプロットされた項目群に着目すると、高得点群にもかかわらず、正答率が低くなっている項目群と、その逆に、低得点群の被験者の正答率が著しく高くなっている項目群が存在することが確認できた。これらの項目は統計上、「外れ値」として認められるもので、項目内容を精査したり、問題項目から削除するなどの検討が必要となる。また、正答率が 1.0 と .0 の項目が、高得点群と低得点群の両端に出現していることが確認できた。これらの項目については、問題の難易度を調整することが求められる。



（図 6：Kernel Density Estimation for Ability Estimates）



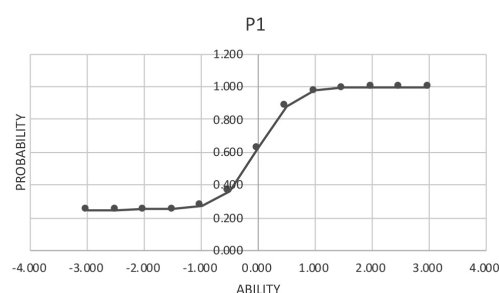
（図 7：得点正答率分布図）

7.4.2 項目特性曲線について

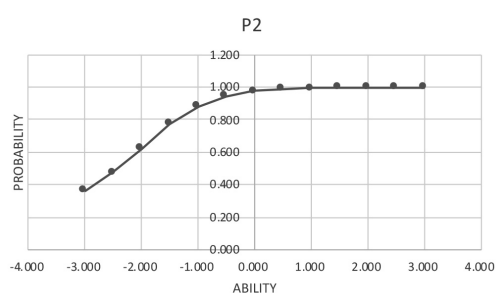
本研究で用いる、Three Parameter Logistic Model における Item Characteristic Curve（項目特性曲線）は、項目困難度パラメータ、項目弁別力パラメータと、当て推量パラメータの 3 つのパラメータを用い、被験者の推定能力（-3.0 から +3.0）と正答確率（.0 から 1.0）との関係性を示す曲線を設定し、テスト項目の特性と被験者の能力との関係を分析し検証するものである。そこで、本研究においては、特性曲線の形状を分析し、改善すべき問題の検出や難易度の判定を行うこととする。具体的には、項目特性曲線は、一般的に推定能力 .0 の近辺で正答率が 50% を越えるような、滑らかな S 字カーブ（P1）が望ましいとされている（図 8）。一方、正答率が 50% を超える

位置が左にずれている図 9 の特性曲線（P2）は、項目難易度が低いことを示している。また、正答率が 50% を超える位置が右にずれている図 10 の項目 P3 の難易度は、高いことが示されている。さらに、図 11 の特性曲線は、水平な直線（P4）であり、項目弁別力が .0 に近いものであることが分かる。

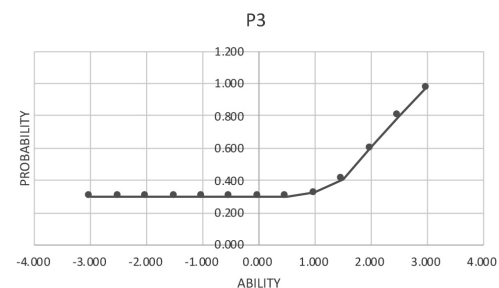
このように、項目特性曲線の形状が、S 字の形から外れて、直線に近い形になったり、逆 S 字の曲線を描いているような項目は、被験者の能力を適切に測定したり推定したりすることが困難な項目であると判断することができる。そこで、このような項目の問題文や選択肢を検討し、指示文や問題文、選択肢の曖昧さや、ダブルミーニングを含む不適切な表現等が含まれていないかを確認し、項目の改訂や削除を行う。このように、項目応答理論に基づく分析により、視覚的に改善すべき項目を特定したり、項目を取捨選択したりすることが可能となる。



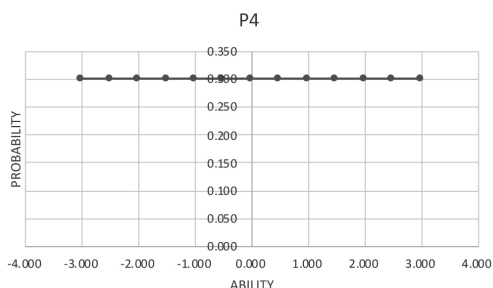
（図 8：理想的な項目特性曲線）



（図 9：項目難易度が「低い」項目特性曲線）



（図 10：項目難易度が「高い」項目特性曲線）



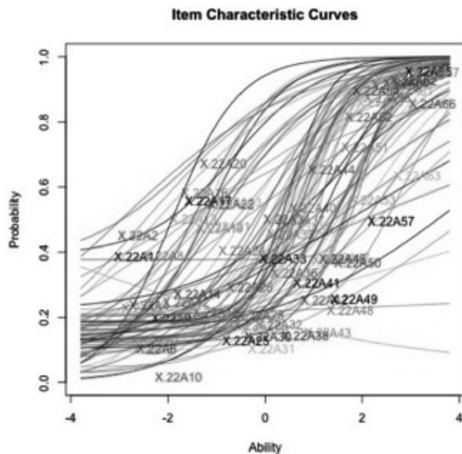
（図 11：項目弁別力が弱い項目特性曲線）

「OIDAI 英語力診断テスト 2022」の全項目の Item Characteristic Curve（項目特性曲線）を図 12 に示す。その結果、項目特性曲線の形状から判断して、被験者の能力を適切に測定することができない、また、弁別力が弱い項目が複数存在することが判明した。

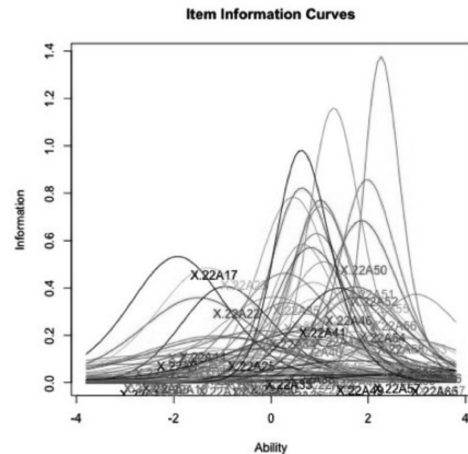
7.4.3 項目情報曲線について

項目情報曲線（Item Information Curve）は、各項目で測定することができる能力の幅（バンド）

とその情報量を示すもので、情報量が最大である状態が、ある一定の能力水準の幅をカバーしていれば、その幅に入る能力を持っている被験者については、最も正確な測定ができるということを示している。すなわち、情報量が最大であるポイント付近を中心に、その周辺ではより正確な測定が可能であり、能力測定の精度を高めることができるということが示されている。図 13 に 67 項目の項目情報曲線を示す。



(図 12：診断テスト 67 項目の項目特性曲線)



(図 13：診断テスト 67 項目の項目情報曲線)

この結果、能力中位群以上の被験者群に対しては、より正確に、また、精緻に個人の能力を測定することができる項目が複数配列されていることが判明した。一方、能力下位群については、特定の能力幅において正確に能力を測定するために必要な情報量を提供することができる項目が十分ではないことが明らかとなった。

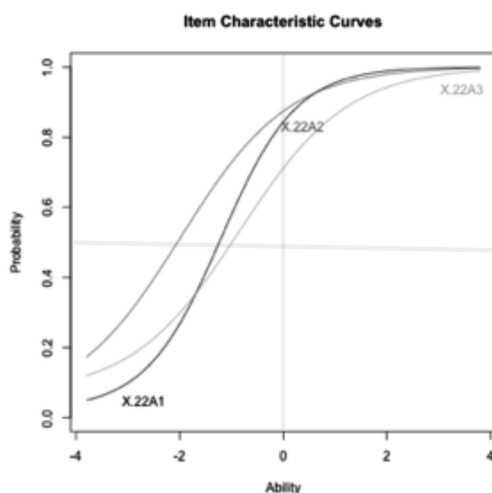
7.5 特色のある項目群別の分析結果と考察

本論文では、Three Parameter Logistic Model による項目特性曲線の分析結果から、特色のある 3 つの項目群について考察を行う。

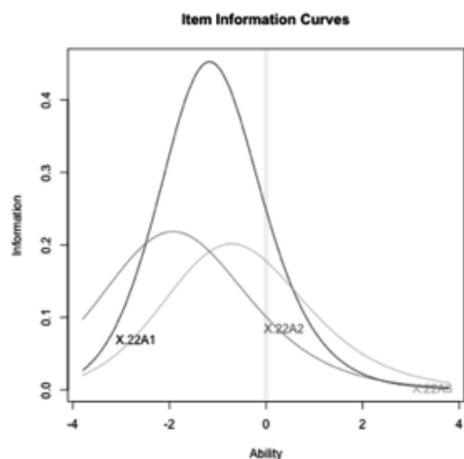
7.5.1 リスニング (写真描写) 問題項目群

項目 A1、A2、A3 は写真を描写する短文リスニングの問題である。写真を見ながら聞こえてくる 4 つの英文の中から写真の内容と合致する文を選択する問いである。問題の意図が明確であり、錯乱肢が設けられているものの、鮮明な視覚情報 (写真) により、被験者が経験した内容のある内容を反映している項目 A1、A2 の正答率が高くなっていることが確認できた。一方、写真内容を被験者の直接体験と結びつけることが難しい項目 A3 の正答率は A1、A2 と比較して低くなっていることが判明した。これらの項目特性は、図 14、15、16 に示す 3 つの特性曲線に明確に表

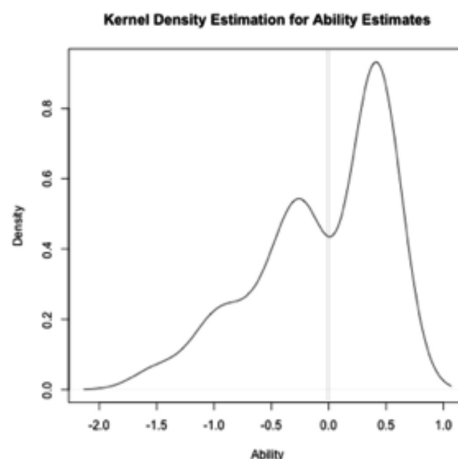
されている。なお、各特性曲線の横軸は推定能力値で、中心（0）から右へいくにつれ高い能力群であることを、左へいくにつれ低い能力群であることが示されている。カーネル密度推定（図 16）では、3つの項目群により測定される推定能力バンドが、-1.5 から +1.5 幅であり、+1.5 付近でその情報量が最も多くなっていることが判明した。以上の分析結果から、リスニングの写真描写問題は、能力中位・下位群（ $-1.5 < \theta < +1.5$ ）のリスニング能力測定に適しているものであることが確認できた。



（図 14：A1-3 項目特性曲線）



（図 15：A1-3 項目情報曲線）



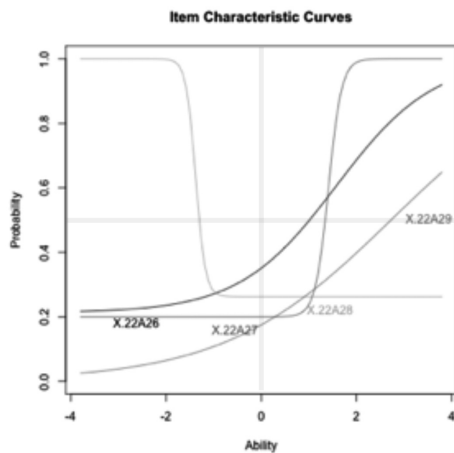
（図 16：A1-3 能力密度曲線）

7.5.2 リスニング&リーディング（聞き読み問題）

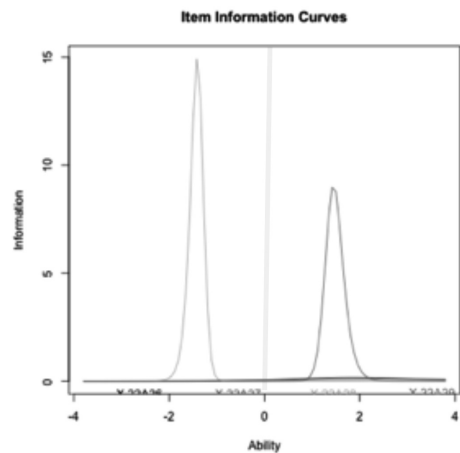
「OIDAI 英語力診断テスト 2022」の質問項目群の中で、最もユニークでチャレンジングな設問

形式として設定したのが、項目 26、27、28、29 から成る「聞き読み」問題である。これは、画面上でリーディングテキストとナレーション（音声）がシンクロされムービーファイルとして提供される問題であり、被験者がこれまで経験してきた一般的な英語問題形式とは、その問題内容であるリソースの提供方法が大きく異なるものである。すなわち、従来方式では、リスニング力やリーディング力等、英語の特定の技能に関わる能力を測定するために必要な、音声情報や文字情報等のリソースに基づく設問がなされているが、新たに設定した「聞き読み」項目群では、リスニングに関わる音声情報と、リーディングに関わるテキスト情報とを同時にハイブリッド型で提供するものである。ちょうど、英語字幕（キャプション）付きの映画の画像を省いたものを問題リソースとして採用したものである。

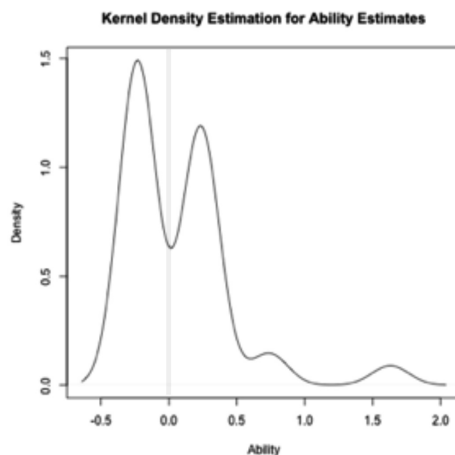
項目分析の結果、被験者の混乱の様子が特性曲線に明確に表れていた（図 17、18、19）。すなわち、



（図 17：A26-29 項目特性曲線）



（図 18：A26-29 項目情報曲線）



（図 19：A26-29 能力密度曲線）

理想的な質問項目の項目特性曲線は、なめらかな S 字カーブとなるものが、項目 27、28 の特性曲線は、漢字の「冂」型のように、一部は逆 S 字型として出現していることが確認できた。また、項目 26、29 は、J 字カーブを描き、項目難易度が高くなっていることが判明した。

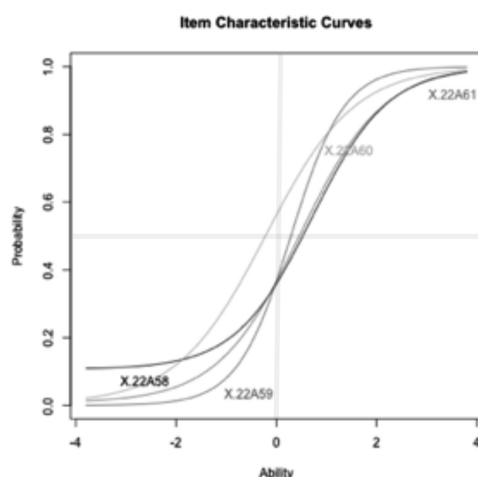
さらに、項目情報曲線 (Item Information Curve：図 18) とカーネル密度推定 (Kernel Density Estimation：図 19) から、各項目が測定することができる能力バンドを検証した結果、提供される情報量やカバーされる能力バンドが限定的であることから、英語運用能力の測定や推定には適していないことが明らかとなった。

7.5.3 リーディング (読解問題)

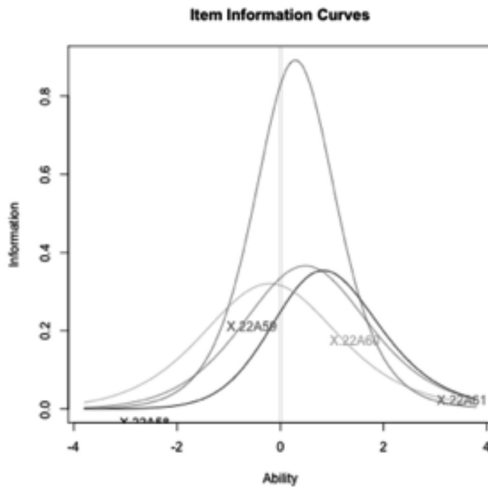
リーディング能力 (読解力) を測定する項目 58、59、60、61 の分析結果を考察する。この項目群は、被験者がこれまで最も多く経験してきたポピュラーなテスト形式であり、設問項目も本文のテキスト情報と関連させた安定的なものであった。分析結果は図 20、21、22 に示すとおりである。

各質問項目の項目特性曲線 (Item Characteristic Curve：図 20) は、能力推定値 0 を中心に滑らかな S 字のカーブを描き、各項目の難易度や弁別力が妥当であることが示されている。また、能力推定に必要な情報量を示す項目情報曲線 (Item Information Curve：図 21) も、どの能力バンドにおいても能力推定値 0 を中心に満遍なく滑らかなカーブを描き、十分な情報量が提供されていることが確認できた。次に、推定能力値の分布を示すカーネル密度推定 (Kernel Density Estimation：図 22) の分析結果から、推定能力値の分布がやや左寄り (マイナス側) ではあるものの、全能力バンドを十分にカバーしていることが確認できた。

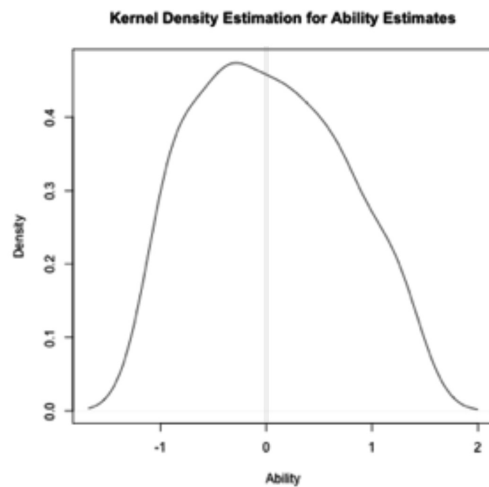
以上の分析結果から、読解力を測定する項目群は、被験者の能力を適切に測定・推定することができる質の高い項目群であることが判明した。



(図 20：A58-61 項目特性曲線)



(図 21：A58-61 項目情報曲線)



(図 22：A58-61 能力密度曲線)

8. 課題とまとめ

本研究では、項目応答理論の Three Parameter Logistic Model（スリー・パラメータ・ロジスティックモデル）により導き出される項目特性曲線を用いて、2,000 人規模の大学生の受験を想定し自主開発している「OIDAI 英語力診断テスト」のテスト項目分析を行った。

その結果、古典的テスト理論に基づく項目分析では十分に把握することができない個々のテスト項目の特性を浮かび上がらせ、各テスト項目の改善を行うための情報や、テスト項目を取捨選択する等、テスト項目群、あるいは、テスト全体の構成やデザインを行うために必要な情報を得ることができた。

今後は、項目応答理論を用い、各テスト項目と被験者の能力推定のパラメータに関する不変性（Invariance）を追究し、「OIDAI 英語力診断テスト」の妥当性や信頼性を高めることが求められる。また、項目応答理論を用い、「OIDAI 英語力診断テスト」の等化（水平的等化・垂直的等化）を実現することも今後の課題となる。

参考文献

- 阿部真理子, ブライアン・ウィスナー, 酒井英樹 (2008). 「古典的項目分析とラッシュモデリングを用いた英語熟達度テストの分析」『高崎経済大学論集』, 50, 3-4, 125-134.
- 張一平 (2009). 「2 パラメータと 3 パラメータ項目反応モデルにおける比較」『行動計量学』, 36 : 1, pp. 15-24.
- J.D. ブラウン. (1999). 『言語テストの基礎知識：正しい問題作成・評価のために』東京：大修館書店。

松 宮 新 吾

- 大友賢二（1996）.『項目応答理論入門：言語テスト・データの新しい分析法』東京：大修館書店．
- 松宮新吾（2022）.「項目分析と古典的テスト理論に基づく「OIDAI 英語力診断テスト」の研究開発」『国際教養学部紀要』, 15：pp. 1-21, 追手門学院大学国際教養学部.