

筑波大学博士（学術）学位請求論文

ビジネス日本語能力の評価・測定に関する研究

小野塚 若菜

2016年度

目次

第1章 研究の背景と目的	1
1.1 日本企業の外国人材採用の活発化	1
1.2 ビジネス日本語能力の評価・測定の必要性	2
1.3 ビジネス日本語テストによる測定の信頼性・妥当性の確保の重要性	4
1.4 研究の目的	5
1.5 本論文の構成	7
第2章 先行研究	10
2.1 ビジネス日本語に関する先行研究	10
2.1.1 ビジネス日本語のとらえ方	10
2.1.2 ビジネス日本語教育について	12
2.1.2.1 ビジネス日本語教育の対象	12
2.1.2.2 ビジネス日本語教育の内容	16
2.1.3 BJT ビジネス日本語能力テスト	18
2.1.3.1 開発の背景と目的	18
2.1.3.2 BJT の出題形式	19
2.1.3.3 等化得点とレベルガイドライン	20
2.1.3.4 運用実績と今後の展開	22
2.2 テストの信頼性・妥当性に関する先行研究	23
2.2.1 言語テスト研究について	23
2.2.2 測定と評価	23
2.2.3 テストの信頼性	25
2.2.3.1 信頼性係数	26
2.2.4 テストの妥当性	28
2.2.4.1 測定の妥当性の検証	30
2.2.5 信頼性と妥当性	37
2.2.6 項目反応理論 (Item Response Theory : IRT)	38
2.2.6.1 項目特性曲線	39
2.2.6.2 IRT に基づく等化	42

2.2.7 項目分析 (Item analysis)	44
2.2.7.1 BJT における項目分析	45
2.2.7.2 日本のテストにおける項目分析	57
第 3 章 構成概念としてのビジネス日本語能力	58
3.1 研究背景と研究目的	58
3.1.1 研究課題の設定	58
3.1.1.1 研究背景と目的	58
3.2 先行研究	62
3.2.1 外国語でのコミュニケーション能力の理論的背景	62
3.2.1.1 コミュニケーション能力の構成要素	62
3.2.2 ビジネスコミュニケーション能力	64
3.2.2.1 ビジネスコミュニケーション能力の構成要素	65
3.3 考察 構成概念としてのビジネス日本語能力	69
3.4 第 3 章のまとめと今後の課題	71
第 4 章 ビジネス日本語能力の評価・測定	73
4.1 研究目的と研究方法	73
4.1.1 研究課題の設定	73
4.1.1.1 先行研究と本章の意義	73
4.1.1.2 研究目的	77
4.1.2 研究方法	78
4.1.2.1 BJT が測定する能力の分析	78
4.1.2.2 ビジネス日本語 Cds の予備調査の概要	78
4.2 結果	79
4.2.1 BJT が測定するビジネス日本語能力の領域・技能	79
4.2.1.1 BJT のスケール	81
4.2.1.2 BJT の設問で要求される課題と情報処理の内容	82
4.2.1.3 BJT の部門別のタスク	84
4.2.1.4 BJT が測定する構成概念	87
4.2.2 ビジネス日本語 Cds の開発と検証	89

4.2.2.1	予備調査の結果.....	89
4.2.2.2	ビジネス日本語 Cds 項目の修正・削除	89
4.2.2.3	本調査の実施と分析結果	96
4.3	第 4 章のまとめと今後の課題	103
第 5 章	ビジネス日本語テストの測定の質と改善	105
5.1	研究目的と研究方法	105
5.1.1	研究課題の設定	105
5.1.1.1	研究の背景と目的	105
5.1.1.2	先行研究と本章の意義	106
5.1.2	研究方法	108
5.1.2.1	特異項目の抽出と修正内容の検討	108
5.1.2.2	特異項目の修正方針	109
5.1.2.3	調査の実施	111
5.2	結果	113
5.2.1	基本統計量	113
5.2.2	項目統計量の比較	114
5.2.3	項目特性曲線の比較	116
5.2.4	修正の実際例	122
5.3	第 5 章のまとめと今後の課題	132
第 6 章	ビジネス日本語テストにおける DIF の分析	136
6.1	研究目的と研究方法	136
6.1.1	研究課題の設定	136
6.1.1.1	研究背景と DIF 研究の意義	136
6.1.1.2	先行研究と本章の意義	138
6.1.2	研究目的	139
6.1.3	研究方法	139
6.1.3.1	テストデザインと受験者の概要	139
6.1.3.2	統計的 DIF 分析 (Study1)	141
6.1.3.3	主観的 DIF 分析 (Study2)	144

6.2 結果	147
6.2.1 統計的 DIF 分析 (Study1)	147
6.2.1.1 性別による DIF	147
6.2.1.2 居住地による DIF.....	148
6.2.2 主観的 DIF 分析 (Study2)	151
6.2.2.1 性別による DIF	151
6.2.2.2 居住地による DIF.....	152
6.2.3 統計的 DIF と主観的 DIF の比較 (Study3)	153
6.2.3.1 統計的 DIF と主観的 DIF が一致する項目	156
6.2.4 項目分析	158
6.3 第 6 章のまとめと今後の課題.....	198
第 7 章 本論文のまとめと今後の課題	202
7.1 本論文の研究課題と解答	202
7.2 本論文の意義.....	205
7.2.1 ビジネス日本語能力を評価するための指標.....	205
7.2.2 テスト開発・運用.....	207
7.3 本論文が与えるビジネス日本語能力の評価・測定と日本語教育への示唆	208
7.4 今後の課題と展望	211
付記	214
謝辞	215
参考・引用文献.....	216
資料	227
資料 1 第 4 章で使⽤したビジネス日本語 Cds.....	227
資料 2 第 5 章で使⽤した実験テストフォーム.....	229
資料 3 第 6 章で使⽤した DIF 分析⽤プログラムコード (PERSCALE)	254
資料 4 第 6 章で使⽤した実験テストフォーム.....	256
資料 5 第 6 章で使⽤した審査表	294

第1章 研究の背景と目的

本章では、本論文の背景と目的を述べ、本論文の構成を説明する。

1.1 日本企業の外国人材採用の活発化

経済のグローバル化の進展に伴い、日本企業の多くはさらなる成長戦略を求めて、経営や生産の基盤や販路拡大のための営業拠点を海外にも求めている。経済産業省の「海外事業活動基本調査」によると、現地法人企業数は、1997年度は13,166社だったのが、2014年度は19,887社となっており、その数は1.5倍以上に増えている。海外での売上高もアジア地域を中心に増加しているという。そのような背景において、経済産業省と文部科学省は、アジアの相互理解と経済連携の促進に向け、優秀な人材を育てるため、「アジア人財資金構想」（以下、アジア人財とする）を2007年度から2010年度に実施した。本事業では、日本の企業に就職する意欲の高い学業優秀な留学生を日本へ招聘し、ビジネス日本語や日本企業文化について学習の機会を提供するとともに、インターンシップの実施、就職支援を行い、日本企業への就職を見据えたサポートを行った。また2015年度からは新たに「外国人材活躍推進プログラム」を実施し、セミナーやイベント等を通じて、国内企業等での就職を希望する留学生をはじめとする外国人と、外国人の採用に興味・関心のある国内企業等を結びつける仕組みを強化している。

一方で、日本学生支援機構が平成28年1月に実施した私費留学生の生活実態調査によれば、卒業後の進路希望として、「日本において就職希望」と回答した者が3,838人(63.6%)で最も多く、「日本において進学希望」の3,042人(50.4%)を上回っている。平成23年度と平成25年度の同様の調査では「日本において就職希望」はそれぞれ3,234人(52.2%)、3,957人(65.0%)となっており、日本において就職を希望する外国人留学生もまた、今後ますます増加していくことが予想される。

外国人留学生の就職活動は、大企業志向、大都市志向が強く、経営資源の限られた地方の企業や中小企業は、人材面においても海外展開を開始・拡大するための十分な対応ができていない。そのため中小企業庁は、日本国内の外国人留学生を活用し、中小企業の海外展開を支援することを目的として、2013年4月より「地域中小企業の海外人材確保・定着支援事業」を実施している。当該事業では、中小企業団体、大学等、ハローワーク、留

学生支援団体等と連携して、日常的な関係づくりから、中小企業と留学生等のマッチング、人材育成・定着までを一体的に実施する取り組みに対して補助することにより、中小企業の海外展開に役立つ外国人材の確保を支援した。

このように、日本企業において、海外および日本国内の日本語教育機関で日本語や専門分野を身につけた日本語学習者が高度なグローバル人材として求められる時代が到来している。そのような中で、日本企業が外国人留学生に求める資質として、文系・理系ともに最も多いのが「コミュニケーション能力」であり、次に「日本語力」とあるという調査結果がある（株式会社 DISCO, 2015）。同調査では、外国人留学生に求める日本語力のレベルについては、内定時にビジネス上級レベル（「幅広いビジネス場面で日本語による適切なコミュニケーション能力がある」と定義）以上を求める企業は4割前後（文系 44.9%、理系 36.0%）だったが、入社後には文系 79.1%、理系 74.4%と大きく増えており、「かなり高い水準の日本語力を求める企業が多い」と報告している。

では、企業が外国人材に求める日本語能力とは具体的にどのようなものであろうか。本論文では、まず、日本語でビジネス活動をしようとする外国人材に求められる、構成概念（construct definition）としてのビジネス日本語能力を明らかにすることを目指す。

1.2 ビジネス日本語能力の評価・測定の必要性

前節で日本企業に就職を希望する外国人材に対する就職支援事業が実施されている現状を述べたが、株式会社 DISCO（2015）の調査によると、実際に、大卒以上の外国人留学生を高度人材として雇用している企業は、2015年時点で50.6%と半数を超え、2016年度の採用を見込んでいる企業は57.1%となっており、外国人材採用への意欲が高まっていることがわかる。しかしながら、採用人数は「1名」という企業が44.1%を占め、「2-3名」（34.4%）とあわせると、全体の8割近く（78.5%）が3名以下にとどまるという。日本企業への就職を希望する外国人留学生のうち、実際に採用に至っているのは3割程度とされている（法務省入国管理局, 2015）。また、外国人材を受け入れている企業でも人材を活用する仕組みが十分に整備されているとは言えない。企業活力研究所（2010）が行った「日本企業における高度外国人材の採用・活用に関する調査」では、企業における外国人材の活用が進まない要因のひとつとして、採用時の外国人材の能力の判定が難しいことを指摘している。このことは、外国人材をより積極的に採用するためには、能力を客観的に

評価・測定するツールが必要不可欠であることを示唆している。

外国人材に求められる能力のうち、日本語能力の評価・測定のツールとしては、**BJT** ビジネス日本語能力テスト (**Business Japanese Proficiency Test** : 以下、**BJT** とする) がある。**BJT** は、大規模日本語試験のひとつであり、日常のビジネス場面で日本語によるコミュニケーション能力がどの程度あるかを測るテストである (2.1.3 節, 4.2.1 節で詳述)。海外技術者研修協会 (2009) によると、先述したアジア人財では、**BJT** の個別テスト (個別機関等からの要請があった場合にその都度実施されるテスト) はビジネス日本語の習得度に関する外在化評価ツールであり、全コンソーシアム共通で利用できるものであることから、事業の円滑化と効率化を図るため、当該事業に参加している学生の、学習成果の評価ツールのひとつとして採用していた。そこで本論文は、**BJT** で出題されるテスト項目をより精緻に分析し、テストで測るビジネス日本語能力を明らかにすることを目指す。

ただし **BJT** は、受験者がビジネスコミュニケーション活動において、日本語を使用して具体的にどのようなことができるのか、あるいはどのような点が不足しているのかを、その得点から推し量ることはできない。大規模試験で、ある得点を得た受験者が現実場面でどれだけの運用能力があるかを自己評価で示すことができるものには **Can-do statements** (具体的な課題遂行能力を記述した文。以下、**Cds** とする) があるが、ビジネス日本語に特化された **Cds** についてはこれまで検証されてこなかった。そこで本論文では、**BJT** が測定対象としている能力を参考にビジネス日本語 **Cds** を開発し、その有効性を検証する。

1.3 ビジネス日本語テストによる測定の信頼性・妥当性の確保の重要性

外国人の日本語能力を測定する日本語テストは、近年、コンピュータを利用したものや、生活者の日本語支援のための能力評価を目的としたものなどが創設され、多様化が進んでいる。ビジネス日本語能力を測定するテストは、前節で述べた BJT 以外にも STBJ 標準ビジネス日本語テスト (Standard Test for Business Japanese) やビジネス J.TEST などがある¹が、現在もっとも社会的に広く活用されているのは BJT である。

一般に、日本語能力などの、ある特性を測定するテストにおいて重要とされるのがテストの信頼性 (reliability) と妥当性 (validity) である。信頼性とは、測定の安定性を示す概念である。つまり、あるテストにおける個人の測定結果の一貫性ということもできる。妥当性とは、テストで測ろうとしているものと実際に測っているものがどれほど合致しているか、換言すると、測定を意図した構成概念とは無関係の要因が含まれていないかを示す概念である。世界最大の非営利テスト開発機関である、米国の Educational Test Service (ETS) は、テスト開発理念として、制作・提供されるテストが、信頼性、妥当性、公平性 (Fairness) を満たしているかを、独自に調査研究を行い、統計分析の専門家によって分析し検証され続けているという (国際ビジネスコミュニケーション協会「ETS のブランドクオリティ」)。これらの作業を行うことによって、「信頼性と妥当性が確保されテストの測定の精度と質を高めている」としている。

テストの公平性について ETS は、TOEIC (Test of English for International Communication) プログラムにおいて、「特定の国・地域独自の言い回しや文化的背景への理解がなければ解答できないような問題は出題せず、世界共通語としての英語能力を公平に評価できるよう配慮」と述べている。ある特性を測定するテストにおいて、同じ能力水準を有するが異なる下位集団に属する受験者同士を比較したとき、その属する集団の違い、たとえば国籍や出身地、性別等によって特定のテスト項目の難易度に差異が認め

¹ STBJ は、一般社団法人応用日本語教育協会が実施しているテストで、日本企業の外国人従業員や日本企業への就職を希望する外国人日本語学習者を対象とした「日本語を使って仕事をする総合的な日本語能力」の測定を目的としている (一般社団法人応用日本語教育協会「STBJ とは」)。BJT と異なるのは、言語能力のみではなく、各場面でのマナーやデスクの使用法といった、日本の企業分野やビジネス事情についての知識や実践能力を測定するとしている点である。主に中国で実施されており、Web ページのサンプル問題を見る限りでは、出題は中国語で行われている。ビジネス J.TEST は、日本語検定協会・J.TEST 事務局が運営している。目的は、「日本語を使ったビジネスコミュニケーション能力及びその基礎となる知識を測定」することである (J.TEST 実用日本語検定「ビジネス J.TEST」)。また、一般的なビジネスマナーを出題対象としている点が BJT とは異なる。

られる場合、そのテスト項目について特異項目機能 (Differential Item Functioning : DIF) が生じていると言う (Angoff, 1993)。我が国におけるテスト・スタンダード (日本テスト学会 (編), 2007) の基本条項 6.2 では、「文化的な背景の違いなどにより、特定の受検者が著しく不利にならないような措置を講じるべきである」と記されており、DIF はテストの公平性にかかわる概念のひとつとして位置づけられるが、DIF の要因が測定を意図した構成概念とは無関係のものであるなら、それはテストが測定しようとする構成概念をどの程度適切に反映しているかという構成概念妥当性 (construct validity) を脅かしていると言える。

本論文では、ビジネス日本語テストにおける信頼性、妥当性の一側面に焦点を当て、外国人材のビジネス日本語能力を測定する BJT のテスト項目を資料として、その精度と質を高める方策を受験者の解答データから実証的に明らかにすることを目指す²。

1.4 研究の目的

ここまで述べてきたことを背景として、本論文では、外国人材のビジネス日本語能力を適切に評価・測定することを目的として研究を行った。特に、ビジネス日本語能力を客観的に測定する大規模試験の開発過程の諸側面に着目し、各側面の運用に有用な示唆を得ることを目指す。長期的に見てテストの開発過程は、本質的に各運用段階の繰り返しの作業であるが (Linn, 2006)、おおむね Downing (2006) が 12 のステップとして挙げているような開発過程を経て行われる。以下に、各ステップの概要を示す。

1. 全体計画：テストの目的の決定、確認
2. 内容の定義：測定すべき構成概念または内容領域の詳述
3. テストの仕様 (specification)：テストの形式、テストの長さ、採点手続き等の決定
4. テスト項目の作成：効果的なテスト項目の作成、編集
5. テストデザインと組み立て：テスト項目の選択と組み立て
6. テストの制作：印刷や CBT 化
7. テストの実施：公平な実施環境、安全管理の徹底

² 本論文で用いた BJT の項目データは、研究開始当初、日本貿易振興機構から使用許諾を得られているものであり、現在の BJT で再利用されているものではない。

8. テストの採点 (test scoring) : 得点の算出, 妥当性の議論, 項目分析
9. 合格点の設定 : 説明できる合格点 (分割点) の設定
10. テスト結果の報告 : 受験者への得点の通知
11. テスト項目の保管 (item banking) : 再利用のための保存, 安全管理
12. テストの報告 : テストの分析結果・評価の報告, 記録, 解釈資料の作成

本論文で扱う BJT も, 基本的に同様の開発過程を経て開発されてきた。そして, 他の大規模試験と同様, 各ステップでいくつかの課題を抱えている。本論文は, BJT の開発過程にある課題を明確化するとともに, その解決に寄与することを目指す。

テストの開発にあたっては, まずテスト全体の計画を立てるとともに, そのテストの仕様を確定する必要がある。仕様とは, 設計図に相当するもので, この段階でテストに含めるべき内容の大枠を決める。そのためには, 測定しようとする能力や特性に関して理論的な検討を十分にしておくことが重要であり, この段階で測定の対象を明確化していなければ, テストの妥当性にも影響する (野口ほか, 2014)。そこで本論文では, 第 3 章で, 「1. 全体計画」や「2. 内容の定義」, 「3. テストの仕様」に関わる, 構成概念としてのビジネス日本語能力について考察する。ビジネス日本語能力が指す内容は社会・経済情勢の流れの中で刻々と変化するものであるが, そういった中で, 現在, 外国人材にどのような日本語能力が求められているかを分析し, テストが測定すべき構成概念について時代的变化を踏まえて考察する。

テストで測定する構成概念としてのビジネス日本語能力の具体的な内容が明らかになれば, 次にテスト開発者は, そのテストではどのような能力を測定し, また, テストの得点はどのように解釈できるかを, テスト利用者に伝えなければならない。これは「12. テストの報告」における解釈資料の作成にも関わる。昨今の大規模言語試験においては, 野口ほか (2014) が述べているように, この解釈資料を Cds により示す傾向がある。しかし, テストを利用しない場合でも, Cds は外国人材のビジネス日本語能力を評価するための指標として単独で用いることも可能である。そこで第 4 章では, ビジネス日本語能力の具体的な内容を示すビジネス日本語 Cds を開発し, 外国人材のビジネス日本語能力を評価するシステムの構築を目指す。

「4. テスト項目の作成」から「11. テスト項目の保管」は、テストの実施ごとに繰り返して行う作業であり、良質なテストを運用する上で欠かせないステップであるが、これには各専門分野の技術を持つ開発者が必要となり、その技術を継続維持することが課題となる。そこで本論文では、第5章で、主にテストの信頼性に関連した、テスト項目の統計的性質を改善するための方策について明らかにする。そして続く第6章では、1.3節で述べた、テストの妥当性を脅かす要因のひとつであるDIFの分析を行う。この2つの研究を通して、ビジネス日本語テストにおいてテストの質に直結する観点である、高い信頼性、妥当性を確保する方策を実証的に明らかにし、実際の運用におけるテストの質の維持・向上に寄与することを目指す。

各章の目的は以下の通りである。

第3章 日本語でビジネス活動を行おうとする外国人材に求められているビジネス日本語能力について、先行研究ならびに各種調査報告をもとに考察し、時代的变化を踏まえながら、構成概念としてのビジネス日本語能力を明らかにすること。

第4章 テスト得点とビジネス日本語能力の具体的な対応関係を示し、外国人材のビジネス日本語能力の評価に資するためのビジネス日本語 Cds を開発すること。

第5章 ビジネス日本語テストのテスト項目において、測定の質を高めるための具体的な方策を明らかにすること。

第6章 ビジネス日本語テストにおいてDIFの分析を行い、テストの妥当性を脅かす要因のひとつであるDIFを、定量的・定性的に特定すること。

本論文が調査対象とする外国人材とは、主に、国内外の大学等に在籍し、卒業後に日本語を使ってビジネス活動をすることを目指している日本語学習者である。本論文の成果が、彼らの就職支援に寄与すること、そして優秀な外国人材の採用を希望する日本企業の要請に応えるものとなることが期待される。

1.5 本論文の構成

第1章に続く第2章では、ビジネス日本語に関する先行研究ならびにテストの信頼性・妥当性に関する先行研究について概観する。第3章では、構成概念としてのビジネス日本

語能力について先行研究ならびに各種調査報告から分析した結果を述べ、考察する。第 4 章では、ビジネス日本語能力の評価・測定を目的としたテストである **BJT** および **BJT** で測定されているビジネス日本語能力を反映する **Cds** について分析・検証した結果を述べる。第 5 章では、ビジネス日本語テストにおける測定の質と改善について、受験者の解答データに基づく実証的な分析結果を報告する。第 6 章では、ビジネス日本語テストにおける **DIF** 分析の結果を報告する。第 7 章では、第 3 章から第 6 章で得られた結論について総括する。次ページの図 1-1 に、本論文の構成を図示する。

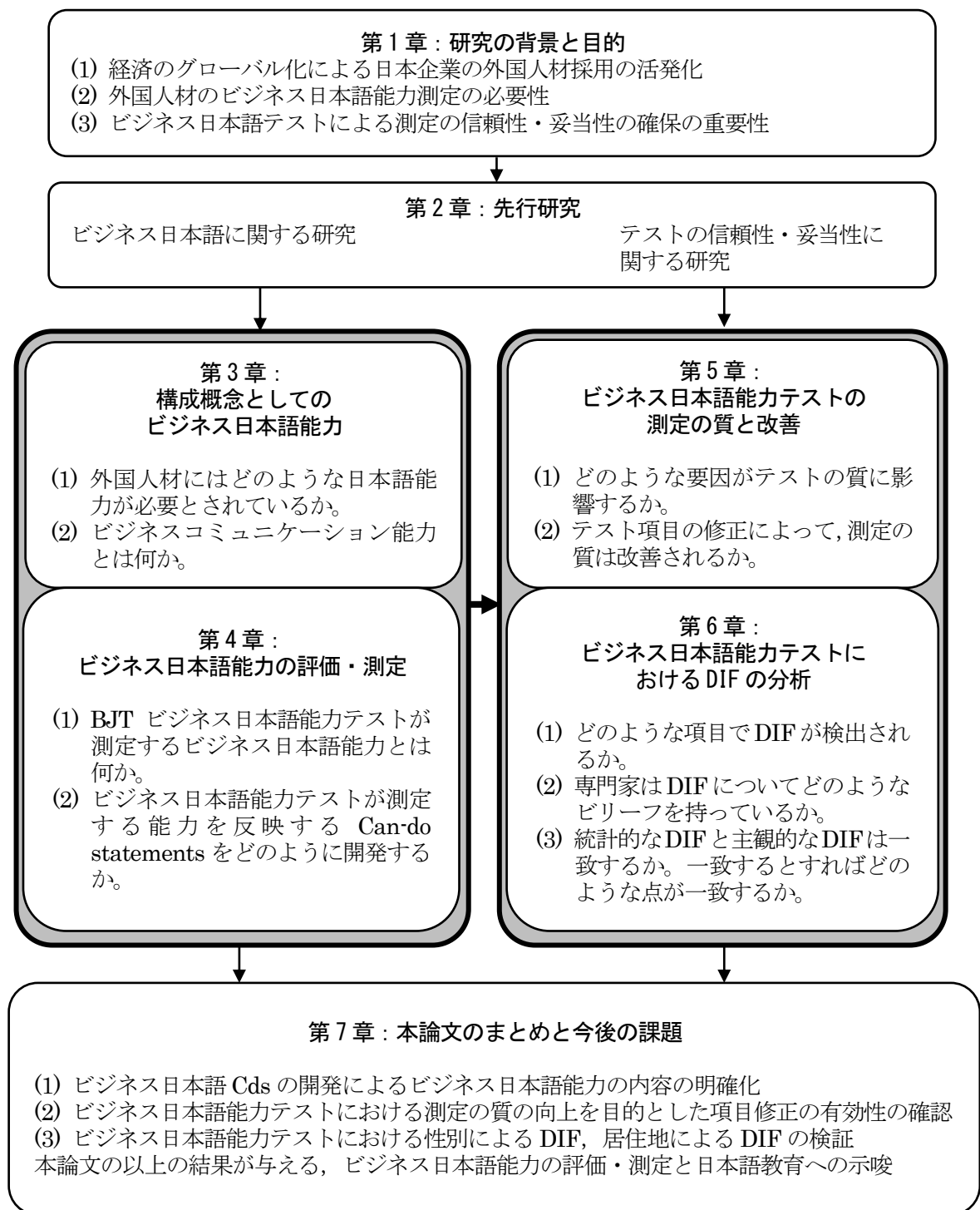


図 1-1 本論文の構成

第2章 先行研究

本論文は、ビジネス日本語に関する研究を踏まえてビジネス日本語について考察し、さらにビジネス日本語テストの信頼性・妥当性に関する研究を行うことから、本章ではそれぞれに分けて先行研究を概観する。

ビジネス日本語能力を測定するテストを開発・運用する上では、まずはビジネス日本語がどのようにとらえられてきたのか、あるいは外国人材に対するビジネス日本語教育ではどのような内容が選ばれてきたのか、先行研究を参照する必要がある。2.1 節ではさらに、BJT が開発されるに至った社会的背景やニーズについて、そして現在はどうに運用され、役割を果たしているのかについて、テスト実施団体の資料からの整理を試みる。

2.2 節では、主に第5章と第6章に関わる、テストの信頼性と妥当性についての先行研究を、テストを開発・運用する際の検証方法に焦点を当てて概観する。その上で、本研究の位置づけを明らかにする。

2.1 ビジネス日本語に関する先行研究

本論文の冒頭で述べたように、昨今、日本企業で外国人材が活用されるようになっていく。そしてこのことを背景に、大学や日本語学校等の日本語教育機関では、留学生の日本企業への就職支援がこれまで以上に行われるようになってきた。またそれに伴い、ビジネス日本語教育のニーズも高まり、ビジネス日本語クラスが設置されている教育機関も多くなってきている。本節では、まずビジネス日本語のとらえ方について先行研究の定義を概観する。また、近年日本語教育機関で行われているビジネス日本語教育の実践について述べる。

2.1.1 ビジネス日本語のとらえ方

ビジネス日本語について、松井（1993）は、「どのビジネス分野にも共通の部分」と「学習者が担当する業種、業務に特有な専門語の部分」の2つに大別されるとし、「共通の部分は仕事は何であれ、日本のビジネス社会で好ましい人間関係を保つための待遇表現、話の運び方であり、個別の部分は各業種、業務における特別な専門語彙と特有な言い回し、特殊な表現である」としている。池田（2001）はビジネス日本語は「商談」の日本語である

と主張する立場と、そうではなく日本人とのビジネスにおける仕事場以外のコミュニケーションの重要性を主張する立場があることを踏まえた上で、「仕事場で行われる職務に関係したコミュニケーションのための日本語」と定義している。李（2002）は、ビジネス日本語に関する先行研究を整理し、『ビジネス日本語』とは、日本語によりビジネス・コミュニケーションを行うもので、そこには日本的人間関係や日本的取引慣行なども絡み合っていると説明されている」（p.247）と述べている。また、「目的遂行能力が重視される点で、ビジネス日本語は、一般的に行われている日本語教育の目標言語としての日本語とは区別されている」としている。

小野寺ほか（2004）は、日本企業二社の新入社員教育マニュアルを比較分析し、その共通項目を取り上げて、言語面、非言語面に分け、図 2-1 のようにビジネス日本語の概念を定義している。この概念図によると、ビジネス日本語は言語面と非言語面とに分けられ、「会社の顔としての自覚」、「顧客とのコミュニケーションの大切さ」は、その両面で基礎となる項目であることを示している。言語面では、「言葉づかい（適切な敬語）」がすべての項目に関連しており、非言語面では、独立した項目もあるが、「言葉づかい（適切な敬語）」という言語面の項目が、「電話のエチケット、マナー」、「クレーム処理」といった項目や、「職場のルール遵守」を内包した「社内の円滑なコミュニケーション」と関連していることがわかる。

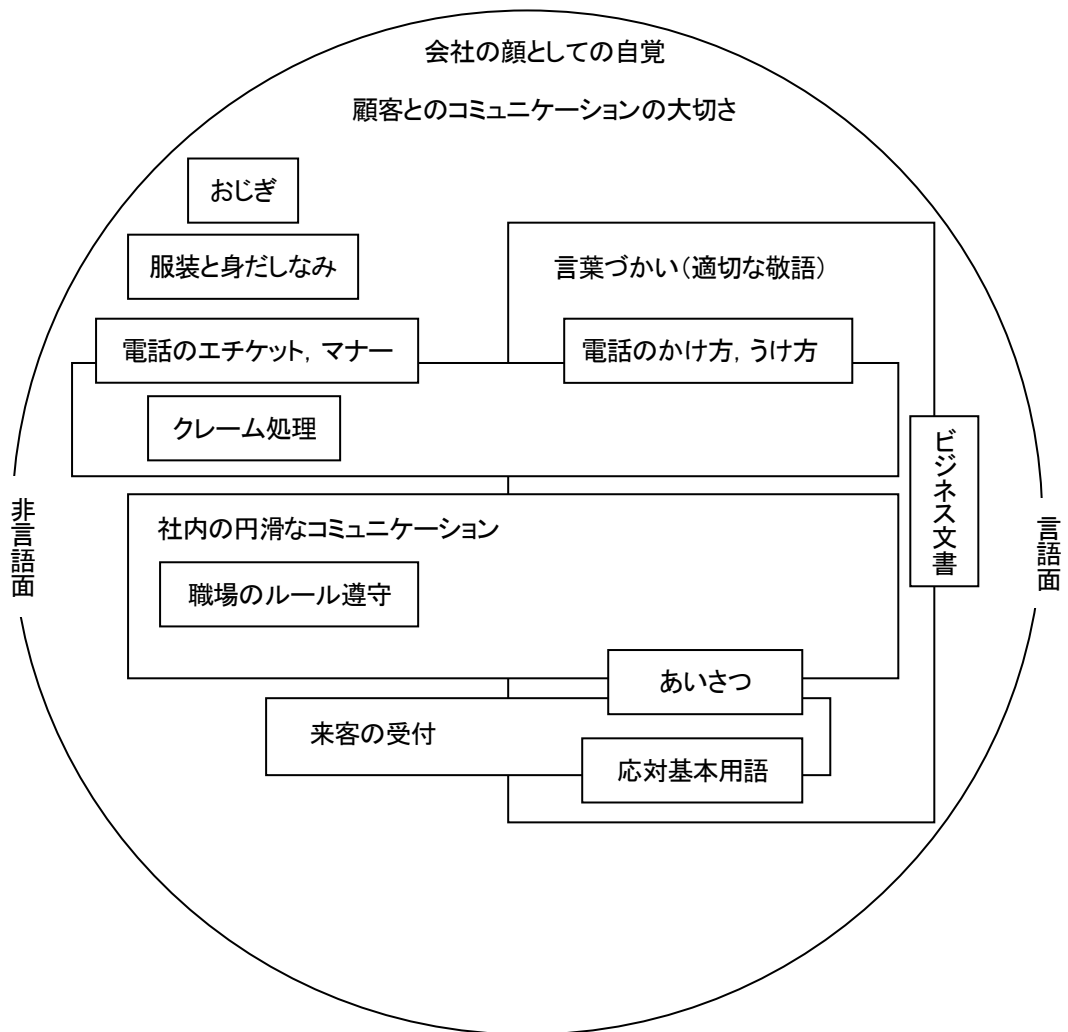


図 2-1 ビジネス日本語の範囲（小野寺ほか，2004：137 を基に作成）

このようにビジネス日本語のとらえ方はさまざまであるが，概ね共通しているのが「人間関係」「コミュニケーション」を重視している点であると言えよう。本論文においても，日本語教育においてビジネス日本語が指す内容には，ビジネス場面でコミュニケーションを行うための日本語であると認識し，調査，考察を行う。企業が外国人材に求めているビジネス日本語能力については，第3章で扱う。

2.1.2 ビジネス日本語教育について

2.1.2.1 ビジネス日本語教育の対象

外国人材のビジネス日本語教育へのニーズ調査として，池田(1995)がある。池田(1995)

は、ビジネス日本語の学習者に対し、ビジネス日本語の教授システムにおける目標の重要度に関するアンケート調査を行い、「言語的スキルよりもコミュニケーションストラテジーやビジネスにおける文化・習慣を重要だと感じている度合いが高い」という結果を得ている。また、外国人を雇用している日本企業に対して、雇用している外国人ビジネスパーソンの方の現状の能力についてもアンケート調査を行い、ビジネス日本語の学習者の結果と比較している（表 2-1）。表中の「不一致度」については、「重要度が高いのに現実においてそれができているかどうかを示す順位が低ければ＋、逆に重要度と比べてそれができている順位が高ければ－の値で示される。であるから、この不一致度は、『その目標の基準に達していない度合い』ということ」（p.51）であると説明している。

表 2-1 ビジネス日本語の学習者の「重要度の順位」と日本企業の「現状の能力の順位」との
不一致度（池田，1995：51）

順位	目標	不一致度
1	日本人ビジネスマンが用いるコミュニケーションストラテジーを理解し、使える	10
2	日本のビジネスにおける文化、習慣を理解できる	7
3	仕事（会議や商談など）中の会話が聞いて理解できる	2
3	仕事（会議や商談など）中の会話が書ける	2
4	日常の挨拶ができる	0
4	日常会話ができる	0
4	簡単な仕事上の文書（Fax、回覧板など）が読める	0
4	会議や商談に用いる資料が読める	0
4	会議や商談に用いる資料が書ける	0
5	ひらがな、かたかなが読める	-2
5	簡単な仕事上の会話が聞いて理解できる	-2
5	簡単な仕事上の会話が書ける	-2
6	簡単な漢字が読める	-3
6	仕事上の簡単な文書（Fax、回覧板など）が書ける	-3
7	ひらがな、かたかなが書ける	-4

8	簡単な漢字が書ける	-5
---	-----------	----

表 2-1 の不一致度からわかるのは、ビジネス日本語の学習者は、コミュニケーションストラテジーやビジネスにおける文化・習慣を理解することを重要な項目であると考えてはいるが、実際のビジネスの現場においては、日本企業の求めるレベルに達していないということである。

清（1995）は、東京に勤務地がある外国人ビジネスパーソンに対する聞き取り調査で得られたビジネス日本語教育に対する考え方を分析することによって、待遇表現の使い分けを親疎・上下のように機械的に提示していたり、敬語を覚え込む事に力点が置かれている現行の上級者対象の会話教材は、学習者のニーズを満たしておらず、また、一般的に行われている新聞読解とディスカッション等の授業は、ビジネス現場で求められている日本語能力を養成するという目的からずれているという結論に達している。池田（2001：13）もまた、「ビジネス日本語会話の授業でありながら、テキスト中の文法事項や語彙の導入、文型定着のためのパターン練習などに時間をとられて、現実の場面に即した練習には十分な時間がとられていない」と述べている。このように、ビジネス日本語教育は、現実のビジネス場面に対応できるコミュニケーション能力の養成に十分対応していないという指摘があった。

その後、2007 年から 2010 年までに経済産業省と文部科学省がアジア人財を実施したことをきっかけに、ビジネス日本語教育を取り巻く環境は変化してきている。財団法人海外技術者研修協会 AOTS 日本語教育センター（2011：1）は、留学生に対するビジネス日本語教育は「就職活動から就職後までを視野に入れ、高度な日本語力の習得やビジネスの背景にある文化や考え方の理解とともに、社会人として生きていくための包括的な能力を育成するもの」と定義している。そして、大学教育や研究という分野で必要とされるアカデミック・ジャパニーズとの違いを踏まえたビジネス日本語教育の特徴として、以下の 5 点を挙げている。

- (1) ビジネスに関する語彙や表現を習得する必要がある。
- (2) 日本語そのものがうまくなることや、教養を深めることより、「日本での就職活動ができるようになる」「仕事ができるようになる」という実利的な目的がより注目される。
- (3) 日本のビジネスの背景にある文化や習慣、考え方、振る舞い方など、言語以外のこと

も学ぶ必要がある。

- (4) 社会人教育やキャリア教育という点を含めて考える必要がある。
- (5) 「大学と企業社会の違い」「留学生の母語と日本の違い」という2種類の異文化性を整理しつつまとめて学習する必要がある。

財団法人海外技術者研修協会 AOTS 日本語教育センター（以下、AOTS という）のビジネス日本語教育の定義とアカデミック・ジャパニーズとの違いから見出したビジネス日本語教育の特徴では、「包括的な人材育成」がその目標であり、その目標達成のためには社会人としての能力を身につけさせることが必要であると考えていることがわかる。この考えは、従来行われていたビジネス日本語教育で重視されてきたこと（上記(1)から(3)に当たる内容）の枠を超え、単に日本語を操れる外国人材ではなく、日本人であるか外国人であるかを問わず、一社会人として日本企業で働く人材の育成が必要との認識に立ったものであると思われる。これは、ビジネス日本語教育の対象が従来とは異なっているためであり、「従来のビジネス日本語教育が、ホワイトカラービジネスパーソンに対する『語学』としてのビジネス日本語教育だとすると、現在求められているものは、多様な学習者に対して、仕事をするための包括的能力育成を目指したビジネス日本語教育」であると説明している（表 2-2）。

表 2-2 従来のビジネス日本語教育と現在求められているビジネス日本語教育

(AOTS, 2011 : 4)

	従来	現在→これから
目的	すでに十分に仕事をする能力を持っている人が、日本語を使っても仕事ができるようにする。	社会人としての経験がないか、または経験が少ない人が、十全に仕事ができるようにする。
対象者	ホワイトカラーのビジネスパーソンが中心	就労者全般 就労希望者（定住外国人、学生等）
内容	ビジネス場面特有の日本語表現	仕事をするための知識や技能、態度、考え方等全般

アジア人財では、包括的能力の育成には、「キャリア教育」（社会に出て仕事をするため

の基礎的な態度や能力の育成),「ビジネス日本語教育」(ビジネス場面で求められる日本語能力の育成),「ビジネス教育」(言語使用の背景にある企業文化や商習慣,企業活動の理解等)の3つの視点から取り組むことが肝要であるとしている。

本論文で対象とするのは,国内外の大学等に在籍し,卒業後に日本語を使ってビジネス活動をすることを目指している日本語学習者である。彼らに対するビジネス日本語教育を考える際に,アジア人財のビジネス日本語教育が目標としている包括的能力の育成という視点は非常に重要であると評価できる。また,日本語を使って働く外国人材が多様化している社会的背景においてビジネス日本語教育を考える際に,いわゆる社会人基礎力の育成という目標を立てることも妥当であるし,そのための教育実践が必要であるということも論を俟たない。しかし本論文では,アジア人財のビジネス日本語教育が対象としている「キャリア教育」「ビジネス日本語教育」「ビジネス教育」の3つの視点のうち,狭義の「ビジネス日本語教育」に焦点を当てて論じることとする。

2.1.2.2 ビジネス日本語教育の内容

ビジネス日本語教育の内容について,神吉(2010)は,「就職活動のための日本語」と「就職後に業務を行うための日本語」の2つに分けられるとしている。「就職活動のための日本語」とは具体的に,自己分析,業界・企業研究,エントリーシートの書き方,面接練習などがあり,それには交渉,主張,説得,説明といったスキルの養成が挙げられる。

「就職後に業務を行うための日本語」の代表的なものとしては,電話対応,名刺交換,ビジネスマナー,ビジネススキルなどがあり,その背景には課題の発見,解決,交渉,主張,説得,説明,思考などがあるとしている。AOTS(2011)によると,アジア人財においても,就職活動前から就職活動へ向けての「前半」と,就職内定から就職までの「後半」に分けて考え,カリキュラムを作成している。

「前半」の「就職活動のための日本語」の授業実践報告としては野元(2005)がある。野元(2005)は,日本企業への就職を希望する大学の留学生に対し,日本での就職活動を支援すると同時に日本企業で日本人と仕事をする上で求められる基本的な事柄の理解を目標に設定し,日本企業で国際ビジネスの経験を持つ日本語教師とのチームティーチングで,実践を行っている。具体的な授業内容は,日本語担当とビジネス担当に分け,次のように説明している(pp.15-16)。

日本語担当

- 1) 就職活動の概要
- 2) 自己分析と自己アピール
- 3) ビジネス文書の基礎（特にメール文書）
- 4) 敬意表現とビジネスマナー
- 5) 模擬面接・ディスカッション

ビジネス担当

- 1) 日本企業の特徴
- 2) 人材開発と研修
- 3) 企業的意思決定過程
- 4) 社内における人間関係の構築
- 5) 社内プレゼンテーション

野元はこの実践の後の受講者アンケートの結果から、本実践について、「就職活動に関する情報を提供し、就職活動がうまくいくように日本語運用能力を高めるという目標が、今回のビジネス日本語教育の主眼であった。しかし、それだけで果たして十分だったであろうか。今回のアンケートでも明らかなように、学生が一番難しいと感じたのは、日本企業の中で何がおこなわれているのかに対する理解であった。そのためには、大学におけるビジネス日本語教育は、『日本企業からの内定獲得』という目標にとどまらず、『入社後の企業でうまくやる』という目標も設定されるべきではないだろうかと思われる」（p.22）と述べている。これは、大学におけるビジネス日本語教育の対象が「就職活動のための日本語」の教育だけでは十分とは言えず、「後半」の「就職後に業務を行うための日本語」の教育を含めるべきとの指摘であると考えられる。

本論文で扱う日本語教育におけるビジネス日本語は、ビジネス場面でのコミュニケーションを行うための日本語であるという認識の下で調査、考察を行うが、ここでのビジネス場面には就職活動における場面を想定しない。本論文ではアジア人財のビジネス日本語教育の「後半」に当たる「就職後に業務を行うための日本語」に焦点を当てて論じることとする。次節では、ビジネス場面でのコミュニケーション能力を測定対象とした大規模試験のひとつである **BJT** ビジネス日本語能力テストについて述べる。

2.1.3 BJT ビジネス日本語能力テスト

2.1.3.1 開発の背景と目的

BJT は、ジェトロビジネス日本語能力テストとして、日本貿易振興会（現・日本貿易振興機構：以下、JETRO とする）によって創設されたテストである。JETRO（2000）によると、JETRO は、1980 年代に貿易摩擦や市場開放問題が注目されるようになる中で、「関税以外に貿易の障壁となるもののひとつが『日本語』である」との指摘が外国政府からなされたことから、ビジネス日本語支援をひとつの事業として位置づけて取り組みを始めたという。

JETRO のビジネス日本語支援事業は、在日大使館員や海外の公的機関の職員を対象とした日本語スクールの運営、研修用教材の作成、インターンシップ支援など多岐にわたった。中でも、日本人とのビジネスコミュニケーションの手段として日本語を使用し学習する外国人の増加、また、日本企業で働く外国人や就職を希望する外国人の増加にしたがい、日本企業および外国人ビジネス関係者双方から外国人のビジネス日本語能力を測定するテストの実現が望まれるようになった。JETRO が 1993 年に行った日系海外進出企業に対するアンケート（計 260 社）の結果でも、60%がこの種のテストの実現を望んでいた。そしてそのニーズに応えるために開発されたのが、ジェトロビジネス日本語能力テストである。

ジェトロビジネス日本語能力テストは、日本企業の国際化・海外進出日系企業の活性化に資することを目的として、外国人を雇用する企業、日本企業への就職を希望する外国人に役立つように、ビジネス日本語のコミュニケーション能力を客観的に示すテストとして、企業の協力を得て開発された。1993 年に日本語教育の専門家、ビジネスマン、外国人から成る検討委員会が発足し、委員会での議論を経て、1996 年にその第 1 回試験が実施された。第 1 回のテストは国内 2 都市、海外 13 都市で実施され、受験者数は計 1,223 名であった。

聴解や読解の試験だけでなく、会話テストの部分も充実させていくべきという開発当初の要望に沿う形で、聴読解テスト（Japanese Listening and Reading Comprehension Test: JLRT）と合わせて口頭コミュニケーション能力を面接で測るテスト（Japanese Oral Communication Test : JOCT）³も実施された。

その後、テストは調査研究がすすめられ、2003 年には、レベル III からレベル I までの

³ JLRT の上位 2 段階以上のレベル（J1+および J1）の受験者が受験資格を得、その結果、6 段階に評価される。JOCT は 2008 年 3 月をもって実施を終了した。

段階別テストを統合し、また、素点による合否判定を項目反応理論による等化得点（2.2.6.2 節を参照）で示されるようになった。そして、2004 年から年に 2 度の実施へと変わり、2006 年にはその名称を「BJT ビジネス日本語能力テスト」と変更した。2006 年に政府の行政改革の一環として BJT の民営化が勧告されたことにより、2009 年に JETRO の BJT 事業は財団法人日本漢字能力検定協会へ売却され、以降、2011 年の上半期中断を挟んで、現在は公益財団法人日本漢字能力検定協会が実施している。

2.1.3.2 BJT の出題形式

2016 年現在、BJT は聴解、聴読解、読解の 3 部門で構成される（BJT ビジネス日本語能力テスト「BJT の特徴」）。さらに、それぞれの部門の下にセクションが設けられており、出題項目数は 100 項目、時間は約 120 分である（表 2-3）。複問形式（1 つの会話または文書に 2 つ以上の設問がある形式）はなく、出題はすべて四肢選択肢方式で、1 つの設問につき正答は 1 つである。写真、絵（イラスト）、文字情報等の視覚情報があり、聴解、聴読解部門には音声情報がある。聴読解部門は、視覚情報と音声情報を同時に処理する形式である。なお、解答はすべてマークシートに記入する。

表 2-3 BJT の構成

部門	セクション	項目数	時間
聴解	(1) 描写問題	10	約 50 分
	(2) 表現力問題	10	
	(3) 聴解問題	15	
聴読解	(1) 描写問題	15	約 30 分
	(2) 聴読解問題	15	
読解	(1) 語彙問題	10	約 40 分
	(2) 表現力問題	10	
	(3) 短文読解問題	15	
合計		100	約 120 分

2.1.3.3 等化得点とレベルガイドライン

BJT は、項目反応理論に基づく得点等化を行っているため、毎回のテスト得点は比較可能なものとなっている。得点は 0~800 点によって示される。また、得点によって J1+~J5 の 6 段階のレベルで評価される（表 2-4）。この認定基準は、BJT の得点結果と現実のビジネス場面における日本語運用能力との相関を示すおおまかな目安である。

表 2-4 BJT のレベルの認定基準

（BJT ビジネス日本語能力テスト「レベルガイド」を参考に作成）

得点	レベル	評価基準
800	J1+	どのようなビジネス場面でも日本語による十分なコミュニケーション能力がある。
600		日本語に関する正確な知識と運用能力がある。 どのようなビジネス会話でも正確に理解できる。 会議、商談、電話の応対などで相手の話すことが正確に理解できる。 対人関係に応じた言語表現の使い分けが適切にできる。 どのような社内文書やビジネス文書でも正確に理解できる。 日本のビジネス慣習を十分理解している。
530	J1	幅広いビジネス場面で日本語による適切なコミュニケーション能力がある。 日本語の知識・運用能力に問題が一部あるが、意思疎通に支障はない。 幅広いビジネス会話が正確に理解できる。 会議、商談、電話での応対などで相手の話すことがおおむね理解できる。 対人関係に応じた言語表現の使い分けがある程度できる。 日常的な社内文書やビジネス文書が正確に理解できる。 日本のビジネス慣習をおおむね理解している。
	J2	限られたビジネス場面で日本語による適切なコミュニケーション能力がある。

420		<p>日本語の知識・運用能力に問題が一部あり，意思疎通を妨げることがある。</p> <p>日常のビジネス会話がおおむね理解できる。</p> <p>会議，商談，電話での応対などで相手の話すことがある程度理解できる。</p> <p>対人関係に応じた言語表現の使い分けが少しできる。</p> <p>日常的な社内文書やビジネス文書がおおむね理解できる。</p> <p>日本のビジネス慣習に対する理解がある程度ある。</p>
320	J3	<p>限られたビジネス場面で日本語によるある程度のコミュニケーション能力がある。</p> <p>日本語の知識・運用能力に問題があり，意思疎通を妨げることが多い。</p> <p>日常のビジネス会話の簡単なものがおおむね理解できる。</p> <p>会議，商談，電話での応対などで相手の話すことが少し理解できる。</p> <p>対人関係に応じた言語表現の使い分けが断片的にできる。</p> <p>日常的な社内文書やビジネス文書の基本的なものがある程度理解できる。</p> <p>日本のビジネス慣習に対する理解が少しある。</p>
200	J4	<p>限られたビジネス場面で日本語による最低限のコミュニケーション能力がある。</p> <p>日本語の知識・運用能力に問題が多く，意思疎通できることが少ない。</p> <p>ゆっくり話された簡単なビジネス会話がおおむね理解できる。</p> <p>対人関係に応じた言語表現の使い分けはできない。</p> <p>日常的な社内文書やビジネス文書の基本的なものが断片的に理解できる。</p> <p>日本のビジネス慣習に対する理解が断片的にある。</p>
	J5	<p>日本語によるビジネスコミュニケーション能力はほとんどない。</p> <p>断片的な日本語の知識しかなく，日本語の運用能力はきわめて不十</p>

0	<p>分である。</p> <p>ゆっくり話された簡単な会話が部分的にしか理解できない。</p> <p>日常的な社内文書やビジネス文書は理解できない。</p> <p>日本のビジネス慣習に対する理解はほとんどない。</p>
---	---

新日本有限責任監査法人（2015）が日本企業に対して行った調査では、外国人留学生に求めている日本語能力を **BJT** のレベルで示しており、英語能力が無い場合は **J1+**程度が 25.9%，**J1** 程度が 40.6%，英語能力がある場合でも **J1+**程度が 20.5%，**J1** 程度が 38.6% であり、「英語能力に関わらず相応の日本語能力を求めている」（pp.47-48）と述べている。

2.1.3.4 運用実績と今後の展開

BJT の Web サイトによると、2016 年現在、**BJT** は 6 月と 11 月の年 2 回実施されており、試験実施都市は国内 7 都市、海外 21 都市である。2015 年度の受験者数の合計は 5,700 人強で、2003 年度からの累積受験者数は 8 万 3 千人を超えている。受験者は半数以上が学生であり、特に日本企業への就職を目指す大学生や大学院生が多く受験している。実際にビジネス現場で働く会社員等のビジネスパーソンの受験者の職種は、購買・生産管理・技術、営業・マーケティング等さまざまである。

試験結果の社会的活用状況については、日本漢字能力検定協会への聞き取り調査によると、2012 年度までに **BJT** を活用している企業数（累積）は国内において 575 社、中国において 717 社あるという。外国人の採用を積極的に行う日本企業においては、透明性の高い人事制度にするための指標として、また、外国人にビジネス日本語教育をしている大学においては就職支援の一環として、あるいは学習の到達目標として **BJT** が活用されている。

また、1.2 節で述べたように、アジア人財においては、受け入れ時の日本語能力の測定指標と、ビジネス日本語教育の到達度測定の目的で **BJT** を採用していた。さらに、2008 年より、「在留資格認定証明書交付申請」における審査の参考として採用され、**BJT** の 400 点以上が「留学」の在留資格に相応すると見なされる。「高度人材に対するポイント制による出入国管理上の優遇制度」においては、日本語能力の参考指標の一つとして活用されている。

このように、**BJT** はビジネス日本語能力を客観的に測定するツールとして社会的に広く活用されていることから、ビジネス日本語能力測定において妥当性を備えたシステムであ

ると認識されていると言えよう。

2.2 テストの信頼性・妥当性に関する先行研究

2.2.1 言語テスト研究について

言語テスト研究が対象とする領域は、人は言語をどのように運用しているか、言語をどのように学習するかということであるが、教育現場では教授または学習過程の評価という側面にまでその範囲が広がっている。Oller（1979：13）は、言語テストに関する研究が取り扱う問題として、次の4つを挙げている。

- (1) 言語運用能力のレベル、言語学習の段階、二言語使用の程度、あるいは言語能力をどのように定義すればよいか
- (2) 言語学習の諸段階をどのように区別するか
- (3) 教授(またはその他の環境的要因)が言語学習に及ぼす効果をどのように測定するか
- (4) 言語能力が教育の現場での知識の獲得にどのように関連しているか

そして、言語テストに関する研究は、言語テストそのものを研究対象とするもの、発話・反応を誘出する手段として言語テストを用い、学習者の特性を調査研究対象とするもの、言語を基盤とするいろいろな作業において言語運用の際に働く心理言語学的要素や社会言語学的要素についての仮説を調べるために、言語テストを研究の手段として使うことの3つの視点に分けられると述べている。本論文が対象としているのは、テストが測定しようとする能力はどのようなものか、その能力をテストは実際に測定しているのか、また、テストの測定結果はどの程度安定しているのか、すなわち、Oller（1979）の言う視点のうち、測定ツールとしての言語テストそのものに焦点を当てるものである。

2.2.2 測定と評価

中村（2002）は、言語テストで得られるテストデータについて、測定と評価それぞれに関する2つの側面があるとしている。測定の側面とは、測定ツールとしてのテストあるいはテスト項目の特性に関するもので、評価の側面とは、測定の結果に意味づけを行う際に必要となる、受験者の特性に関するものである。

一般に、教育評価（Educational Evaluation）とは、学習者に対してはどこまで理解できているか、どこが間違っているかといった情報を与えたり、教授者に対しては個々の学習者への教授法が適切であるかという情報を得たり、あるいは管理や研究を目的として行うもので、「測定や質的記述の結果に価値判断を加えること」（岸，2010）である（表 2-5）。これに対し、教育測定（Educational Measurement）は、教育評価のための客観的な情報を得ることで、定量的に表されるものである。つまり、「測定された結果に意味づけや価値判断をすることで重みづけがなされ、評価となる」（岸，2010）のである。中村（2002）は、「日本の教育文化の中では、測定と評価の分化が曖昧」であり、「テストデータの検討においても、評価、つまり受験者の特性に関するデータのみの検討に重きが偏りがちである」と述べている。

表 2-5 評価の目的と実施時期（バックマン & パーマー，2000）

	目的	時期
診断的評価 (diagnostic evaluation)	学習者の言語能力の長所や短所に関する情報を得る	学期やコースの前に行われる
形成的評価 (formative evaluation)	学習者の学習がどの程度進んでいるかの情報を得て、次に行うべき学習活動の指針を作る。また、教授者がより適した教授方法・教材を開発するための情報を得る	学期やコースの途中で行われる
総括的評価 (summative evaluation)	学習者の学習成果や進歩に関する情報を得る	学期やコースの終了後に行われる

言語の評価と言語教育との関わりについて、李（2015）は、アセスメント（assessment）としての評価（学習の成果を確認する目的で行われる評価）⁴は、教育に関わるあらゆる活動において日常的に行われていると述べ、大規模試験の場合は受験者が自身の言語能力の熟達度を確認できるため、学習の動機付けにもなるとしている。そのため、『教える』こ

⁴ 前述したエバリュエーション(evaluation)との違いについて、近藤ブラウン(2012)は、「エバリュエーション」は、「アセスメント」よりも広域の概念であり、教育の質や価値を維持、もしくは向上するのに必要な評価のすべてがその対象になりえると述べている。

と『評価する』ことは不可分の関係にあり、両者の効率的な循環は教育の成果に直結する」(p.9)と述べている。

以上のことから、言語教育においても、学習者の言語能力を正しく公平に「評価」し、学習者や教授者に価値ある情報をフィードバックするためには、「測定」によって客観性の高い情報を得ることが極めて重要であると言えよう。次節から、「測定」において客観的な情報を得るために必要である信頼性と妥当性について述べる。

2.2.3 テストの信頼性

信頼性 (reliability) は次節で述べる妥当性 (validity) とともに、測定の質を表す概念のひとつで、1.3 節でも述べたように、同じ受験者が同様の条件下で、難易度、質、量が同等のテストを行った場合に同じ結果が出るかどうかという、測定の安定性を表す。また、同じ受験者がある期間をおいて測定を繰り返したとき、1 回目と 2 回目で測定値が一貫しているか、同じテスト作成デザインに基づいて作成したテストの複数のフォームの間で得点が一貫しているとき、その測定値は信頼性が高いといわれる (南風原, 2002)。

BJT のような客観テストにおいて、その信頼性を脅かす原因となるのは、第一に体系的 (systematic) な理由によるもの、第二に非体系的 (unsystematic) な誤差によるものがある (Alderson et al., 1995)。体系的な理由とは、受験者が一定の間隔において同一のテストを受験した場合に、その間の学習効果や発達の変化により得点が向上するといった、いわゆる「真の変化」である (野口ほか, 2014)。このような「真の変化」が得点の変化に反映されることは望ましいことである。これに対して、受験者の体調がすぐれなかった、テストが長時間にわたったため受験者の集中力が欠けていた、外部から雑音があって音声情報がきちんと聞き取れなかったなどという場合には、受験者の本来の能力がテストに反映されなくなる。これを非体系的な誤差といい、測定誤差 (measurement error) として定義すると、(2.1) 式のように、測定値すなわちテスト得点 x は、個人の能力水準に起因し、誤差の影響がない受験者の本当の得点である「真の得点 (true score) t 」と、測定される能力以外の要因に起因する「測定誤差 e 」の和によって表わされる。これは、古典的テスト理論 (Classical Test Theory : CTT) の基本モデルであり、 e の平均は 0、 t と e は無関係であるという仮定を前提として成り立つ (Bachman, 1990)。

$$x = t + e \quad (2.1)$$

CTT は、テスト得点に真の得点が反映されている程度を評価することによって、テストの信頼性を評価しようとする理論であり、すなわち、信頼性の高いテストは、測定誤差 e を最低限に止めることを意味する。そのためには、Alderson et al. (1995) が述べているように、たとえばテストが常に同様の条件で実施されている、受験上の指示が明確であるという外的な要因に加え、採点に首尾一貫性がある、テスト項目にあいまいなところがない（設問が明瞭であり、受験者が何を求められているのかが明らかであるか、正答がひとつ、あるいは指示通りの数であるかなど）、意図する測定範囲を内容的にカバーするだけの十分な項目数がある等、テストの内的な要因についても検証が必要である。あいまいで欠陥のあるテスト項目は識別力（2.2.6.1 節、2.2.7.1 節で詳述）が低くなる傾向があり、識別力の低いテスト項目が多く含まれるテストは、信頼性も低くなる傾向があるという。これについては、第 5 章で扱う。

2.2.3.1 信頼性係数

テストの信頼性を検証するために、信頼性係数を定義する。CTT で信頼性係数を推定する方法は、単一の受験者にテストを 2 度繰り返して実施して推定する方法と、受験者集団に 1 度だけ実施して、項目一貫性を測定する方法がある（Alderson et al., 1995 ; 繁峙, 1998 ; 野口ほか, 2014）。

(1) 再テスト法 (test-retest reliability)

同じ受験者に期間をおいて再度同じテストを実施し、1 回目と 2 回目のテスト結果の相関係数を計算する方法。2 つのテスト結果がよく一致すれば、安定性が高いことを示す。再テスト法で計算される信頼性係数を ρ_1 、1 回目のテストと 2 回目のテストの結果の相関係数を r_1 とすると、(2.2) 式で表すことができる。

$$\rho_1 = r_1 \quad (2.2)$$

ただし、最初に受けたテストの結果を記憶していることが 2 回目のテストの結果に影響することがあるが、1 回目と 2 回目の間に長い時間を置けば、その期間に受験者の能力自体に変化が生じていることが考えられる。つまり、期間を空けずに実施すれ

ば受験者の能力の変化は小さいが記憶の影響を受けやすくなり、期間を空けて実施すれば、記憶の影響は小さくなるが、受験者の能力の変化が大きくなる可能性が高くなる。したがって、再テスト法で得られる 2 回の測定値が強平行測定による（真の得点と同じであり、誤差分散が等しい）という保証はない。

(2) 平行テスト法 (parallel-form reliability)

似通った 2 つのテスト（平行テスト）を作成して同じ受験者に行い、2 つのテスト結果の相関係数を計算する方法。平行テストは難易度、質、量が同等なものを用いるため、理論的に 2 つのテストの結果は一致することになる。(2.2) 式と同様に表される。

ただし、内容的にも量的にも同等のテストを 2 つ作成することは難しく、同等であることを客観的に示す方法もない。

(3) 折半法 (the split half reliability index)

テスト全体を半分に分け、それらを平行テストとみなし、両者間の相関係数を計算する方法。再テスト法や平行テスト法と違い、1 度の実施で信頼性係数を推定することができる。折半法を用いて計算する信頼性係数を ρ_2 とし、ひとつのテストを半分に分けた際の一方の結果ともう一方の結果の相関係数を r_2 とすると、(2.3) 式で表される。スピアマン・ブラウンの公式と呼ばれる。

$$\rho_2 = 2r_2 / (1 + r_2) \quad (2.3)$$

ただし、どのように半分に分けたかによって相関係数は異なることに注意が必要である。

(4) 内部一貫法

折半法は 1 つのテストを 2 つに分けてその両者間の相関係数を計算したが、この方法はすべての分け方で相関係数を計算し、その相関係数の平均を算出する方法。この方法は、テスト得点がどの程度一貫性があるかを評価することによって、信頼性係数を推定することができる。 n 項目から構成されるテストで、各項目の得点を X_j ($j=1, \dots, n$)、その合計点（テスト得点）を X としたときに (2.4) 式で表される。ここで $s^2(X_j)$ は項目 j の得点の分散、 $s^2(X)$ は合計点（テスト得点）の分散を表す。クロンバックの α 係数と呼ばれ、比較的によく用いられる。

$$\alpha = \frac{n}{n-1} \left\{ 1 - \frac{\sum_{j=1}^n s^2(X_j)}{s^2(X)} \right\} \quad (2.4)$$

n 個に分けられたそれぞれのテストが平行テストではなくても、各テストの真の得点間の差が一定であれば（たとえば一方のテスト得点に一律 10 点を加算するなど）、(2.4) 式は信頼性係数の推定値となる。真の得点間の差が一定でないとき、 α 係数は信頼性係数よりも小さい値を与える。すなわち、 α 係数は信頼性係数の下限であり、 α 係数が大きいことは信頼性係数も高いことを保証するものである（繁桝，1998）。 α 係数は 0 から +1 の間の数値となり、1 に近いほど信頼性が高いテストであると解釈できる。言語テストでの目安は 0.8 以上で、0.9 を超えればかなり信頼できるテストと見なされている（中村，2002；岸，2010 など）。ただし (2.4) 式を見てもわかるように、出題する項目数が多ければ α 係数も高い数値が出やすい傾向があり、 α 係数の数値には絶対的な大きさの基準はない。

2.2.4 テストの妥当性

テストの妥当性は、1.3 節でも述べたように、テストで測ろうとしているものと実際に測っているものがどれほど一致しているかといった概念である。米国における心理・教育測定に関するガイドラインとも言える AERA et al. (1999) の *Standards for educational and psychological testing* (『教育・心理検査法のスタンダード』，以下、スタンダードとする) における妥当性の定義では、「実証研究や理論によって、意図されたテストの使用目的に見合ったスコア解釈が裏付けられる度合い」とされている。

しかし、妥当性のとらえ方は変化してきた。表 2-6 は、APA (1954, 1966)、APA et al. (1974) のスタンダードが伝統的な観点での妥当性を分類したものである。この分類では、妥当性の 3 つの「タイプ」として、それぞれ個々の存在として三位一体的にとらえられており、それぞれの関係について詳細な考察は行われなかった (Hubley & Zumbo, 1996)。

表 2-6 伝統的観点からの妥当性の分類

妥当性の種類	定義	検証法
内容的妥当性 (content validity)	項目が、どの程度全体を偏りなく代表しているか。	その分野の専門家に判断してもらう。
基準関連妥当性 (criterion-related validity)		
1. 併存的妥当性 (concurrent validity)	外的な基準とどの程度関連があるか。	同時期に得た基準との相関を検証する。
2. 予測的妥当性 (predictive validity)	その尺度の実施以降の変化などをどの程度適切に予測しているか。	尺度の得点とその後の基準との相関を検証する。
構成概念妥当性 (construct validity)	測定しようとする構成概念をどの程度適切に反映しているか。 尺度の得点とモデルや理論との整合性がどの程度あるか。	信頼できる基準との相関や因子分析などによって検証する。

澤木（2011）によると、1920～1950年代は基準関連妥当性を中心とする妥当性の概念が発達しており、1950年代初め頃までには、内容的妥当性の観点もテストの妥当性検討に使用されるようになったという。構成概念妥当性が議論されるようになったのは1950年代に入ってからで、構成概念妥当性を中心とする妥当性理論が妥当性の一般的な定義として広く認められたのは1980年代前半になってからだという。1985年版のスタンダードでは、妥当性は単一の概念(unitary concept)であり、それまで基準関連妥当性、内容的妥当性、構成概念妥当性と呼ばれてきたものは、別個のものではなく、テスト妥当性の違った側面であると説明した。

Messick（1989）の妥当性理論では、テストの妥当性を単一の概念と定義し、構成概念妥当性をテスト妥当性の中心と位置付けている。そして、構成概念妥当性を、テスト得点の解釈の適切性であるとした。Messick（1996）は表 2-7 のような、構成概念妥当性の証拠集めのための6つの側面を挙げている。Messick はテストの妥当性の検証とは、テストを使用することで明らかとなっていく証拠を積み重ねていくプロセスであるとしている。

表 2-7 構成概念妥当性の 6 つの側面

妥当性の側面	妥当性の証拠の内容
内容的側面	項目の内容が測りたい特性に対応しているか。十分に代表しているか。
本質的側面	項目への反応が何らかの理論に基づいて説明できるか。
構造的側面	尺度内の項目間の関係が特性の下位概念や次元性などの理論的構造に一致しているか。
一般化可能性の側面	ある特定の条件のみでなく、他の集団や実施場面、実施時期に対しても不変であるか。
外的側面	当該尺度とほかの尺度との間に、理論上想定される相関が示されるか。
結果的側面	そのテストを使用した結果として生じた事態によって示される証拠。社会的影響。

Messick (1989) の妥当性理論は、テストの使用場面の適切さや社会的影響までを含め、妥当性の概念をそれ以前に比べて拡張しているのが特徴である (Kane, 2006)。

2.2.4.1 測定の妥当性の検証

テスト得点を使用したり、テスト得点から合否や人事評価をしたりする場合、その正当性を主張するためには、テストの測定の妥当性を検証することが必要である。Messick 以降、言語テストの分野においてもテストが果たす社会的な役割や意味をも考慮した上で、さまざまな検証の必要性やその方法が取り上げられてきている (清水, 2005)。表 2-6 で述べたように、テストにおける測定の妥当性を検証する方法には、専門家が主観的に判断する方法や、既存のテストなど外的基準との相関係数を求める方法、テストに含まれる項目群の因子分析を通して構成概念妥当性を検証する方法などがある。本節では、言語テストにおける妥当性検証について、論証モデルに基づく方法と社会的・認知的枠組みで行う方法について述べる。

2.2.4.1.1 Toulmin の論証モデルに基づいた言語テストの妥当性検証

Toulmin (2003) の論証モデルは、実社会の議論形態を分析して提唱したもので、「主

張(Claim)」を支える根拠を「データ(Data)」と「理由づけ(Warrant)」に分け、さらに「裏づけ(Backing)」,「反証(Rebuttal)」,「反証データ(Rebuttal data)」を加えた6要素を論証の基本要素として定式化している(図2-2)。

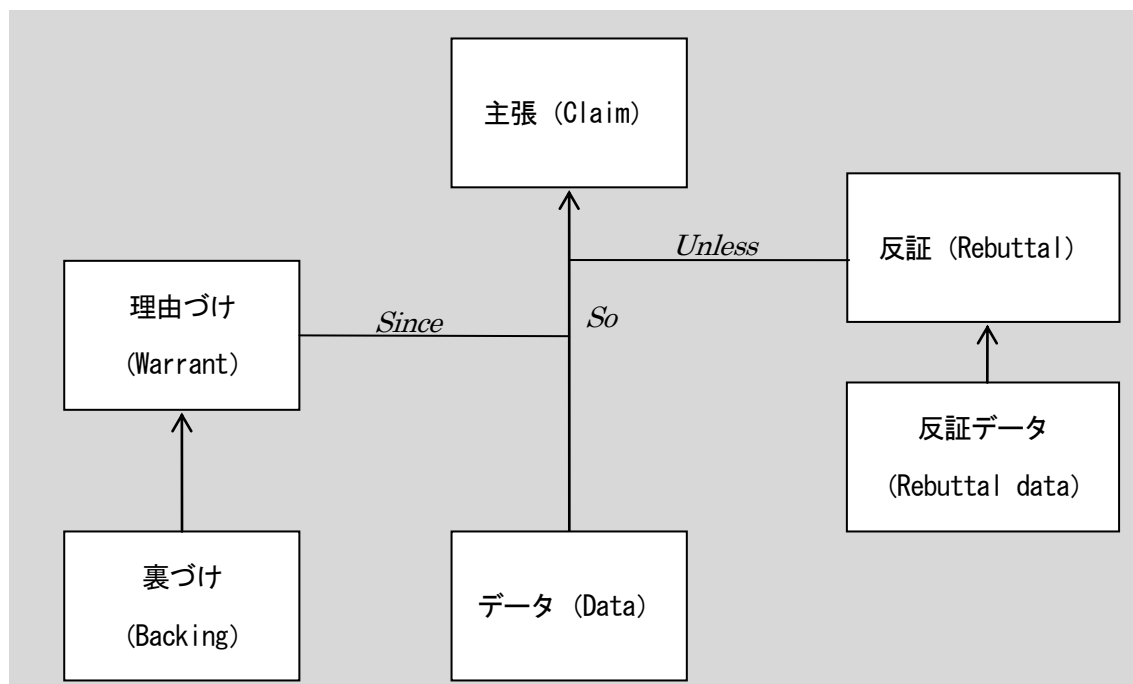


図 2-2 Toulmin (2003) モデルにおける論証の基本形

Chapelle (2008) によると, TOEFL iBT では, Toulmin モデルを使って妥当性を論証しているという。TOEFL の妥当性論証 (validity argument) の最終目的は, TOEFL の得点解釈が測定対象を示すのに妥当であるかどうかであり, 高等教育機関への入学・留学者選抜に適したものであるかどうかということを論証することである。そのため, 次の6つの推論を Toulmin モデルを用いて論証しているという(澤木訳, 2011)。

- (1) Domain definition (領域の定義) : TOEFL で観測される受験者のパフォーマンスは, 英語で授業を行う高等教育機関で学業に従事する際必要な知識やスキル, 能力の代表的なものを明確にするものである
- (2) Evaluation (評価) : TOEFL において観測される受験者のパフォーマンスを評価することによって, 目標言語能力を反映する得点が算出される
- (3) Generalization (一般化) : TOEFL で観測される得点は, 比較可能な課題や形式,

実施・採点条件下で測定を繰り返した際に期待される得点の推定である

(4) **Explanation** (説明) : 比較可能な課題や形式, 実施・採点条件下で測定を繰り返した際に期待される得点は, 構成概念(高等教育機関で勉強するのに必要な英語力)を反映するものである

(5) **Extrapolation** (推定) : TOEFL で測る学業に必要な英語力は, 受験者が高等教育機関において英語でどの程度学業をこなしていけるかを示すものである

(6) **Utilization** (利用) : TOEFL から得られる, 受験者が高等教育機関において英語でどの程度学業をこなしていけるかに関する推定値は, 留学・入学を希望する学生の選抜や教育指導の指針を立てるうえで役立つ

澤木 (2011 : 58) は, 「Toulmin の論証モデルを言語テストの妥当性検討に応用する場合は, 一つの推論に関して最初に予定した研究が一通り完了したとしても, さらに追加データ等を参照して裏付けを強化したり, 以前の研究結果を新しいデータで再検討したりするなど」して, 繰り返し再確認しながら進めていくことが必要になるとした上で, 「この TOEFL 妥当性研究の試みは, 大規模言語テストの妥当性研究をより効率的・組織的に計画・実施していく方向性を探るテストケースとして示唆に富むものである」と述べている。

澤木 (2015) は, TOEFL iBT の Toulmin モデルを用いた妥当性論証を日本語のプレースメント・テストに応用した例を紹介している。この例においては, 海外からの留学生を主な対象とした大学日本語教育プログラムが, 3 つのレベル (初級, 中級, 上級) で構成されており, そこで実施するプレースメント・テストの目的は, テスト成績を基に学生を能力別の 4 グループ (初級, 中級, 上級, 履修免除) に分けることであり, 具体的には以下の 2 点を設定している。そして, 6 つの推論を論証した (表 2-8)。

- ・日本語の授業を履修する必要があるかどうか。
- ・日本語の授業を履修するなら, 初級・中級・上級のどのレベルか。

表 2-8 日本語プレースメント・テストにおける妥当性論証（澤木，2015 より作成）

<p>(1) 推論 1 領域の定義</p> <p>日本語教育プログラムの授業で扱う指導内容が身に付いているかを評価するためのデータを、このテストから得ることができる。</p>	
理由づけ（Warrant）	：初級・中級・上級で扱う指導内容の主なものが中心に出題されている。
裏づけ（Backing）	：担当教員による①日本語教育カリキュラムの指導目標や指導内容，教科書の詳細な分析と②テスト内容の適切性の確認など。
<p>(2) 推論 2 評価</p> <p>算出されるテスト得点は，日本語教育プログラムでの教育内容が身につけているか評価する上で必要な日本語能力の様々な側面を，適切かつ正確に反映するものである。</p>	
理由づけ（Warrant）	：採点基準や採点結果は，日本語教育プログラムで身に付けるべき日本語能力を十分反映している。
裏づけ（Backing）	：テスト実施手順や受験上の注意の適切性の検証，指導目標に合った採点基準の設定など。
<p>(3) 推論 3 一般化</p> <p>このテストで得られる学習者の得点は，平行なタスクや問題冊子，評定者を使ってテストを繰り返し実施した場合に得られる期待される得点を反映するものである。</p>	
理由づけ（Warrant）	：実際のテストの得点は，平行なタスクや問題冊子，評定者を使って日本語能力を繰り返し測定した場合に観測される得点の正確な推定である。
裏づけ（Backing）	：信頼性・一般化可能性の分析，等化，テスト細目 ⁵ ・細目表の作成など。
<p>(4) 推論 4 説明</p> <p>このテストで測っている能力と，構成概念（日本語教育プログラムで育てることを目指す</p>	

⁵ Alderson et al. (1995) は，テスト細目 (test specification) とは，「テストで何をどのような方法で測定するかを明示した設計書 (p.20)」と説明している。

日本語能力) との関係が明らかである。

理由づけ (Warrant) : 期待される得点は、構成概念を反映するものである。

裏づけ (Backing) : 因子分析、バイアス分析、解答プロセスの分析など。

(5) 推論 5 推定

このテストの構成概念は、日本語教育プログラムで指導をする日本語能力を十分に反映するものである。

理由づけ (Warrant) : 初級・中級・上級それぞれで「B」の成績を修めるのに必要な日本語の到達度は、科目のレベルが上がると共に高くなり、到達度の高い学習者ほどこのテストでより高い点を取る傾向がある。

裏づけ (Backing) : 各レベルの授業における成績とプレースメント・テスト得点の関係など。

(6) 推論 6 利用

このテストは、日本語教育プログラムで指導の対象とする日本語能力を十分に反映するものであり、このプログラムにおける授業への学生のプレースメントや履修免除者の決定、また学生の日本語指導の上で役立つ。

理由づけ (Warrant) : 受験者は、授業における「B」の成績と同等またはそれ以上の到達度にあると考えられる最も高いレベルのすぐ上のレベルの科目にプレースメントされる。この決定は適切であり、またそのテスト結果は学生の日本語学習・指導に結びつく有益なものである。

裏づけ (Backing) : プレースメントされた授業の単位取得状況や成績の追跡調査、テストに基づくプレースメント結果と教員の評価の一致度、テスト結果に関する資料 (スコアレポート等) の学習・指導における有用性に関する学生・教員の満足度など。

Toulmin の論証モデルを日本語プレースメント・テストの妥当性検討に応用した澤木 (2015) の試みは、プレースメント・テストを作成する教授者とテストを受験する日本語学習者双方が、テストの目的や得点の解釈について理解しあう方策として有用なものであ

る。このような妥当性論証のプロセスが、結果的にプレースメント・テストの測定の質を高めることにつながると考えられる。

ところで、Chapelle (2008) と澤木 (2015) が、「説明」という推論の中で「期待される得点は構成概念を反映している」という理由づけに対する裏づけのひとつとしてバイアスの検討を挙げているが、Cole & Moss (1993) は、バイアスを「受験者の下位集団におけるテスト得点の解釈の差異妥当性 (differential validity)」と位置付けている。これは、バイアスがあるということは、テスト得点に与えられる意味が受験者の下位集団で異なるということを表すのであり、妥当性確認の理論がバイアス検討の概念的基礎を与えることになる」と述べている。Bachman & Palmer (1996 : 136) もまた、「得点が低い場合、それは受験者の言語能力が低いためなのか、あるいは彼らに話題の知識が欠けているためなのか、解釈があいまいになり構成概念の妥当性に支障をきたす可能性がある」と述べている。バイアスの分析・検討については、第 6 章で扱う。

2.2.4.1.2 Wier の社会的・認知的枠組みにおける言語テストの妥当性検証

Wier (2005) が提案している社会的・認知的枠組み (Socio-cognitive framework) は、Cambridge English Language Assessment を中心に、世界のテスト開発機関で妥当性検証の枠組みとして用いられているという (中津原, 2013)。Wier (2005) は、CEFR (Common European Framework of Reference for Language) が、学習者の認知プロセスが実際の学習者の能力レベルをどのように判別しているのかについて考慮していないという点、また、テストを CEFR と結び付けることと、テストの妥当性の検証をすることは別であり、テスト機関は、テストがレベルをどのように区別しているのかを、明確に提示するべきであると指摘した。そして、受験者の能力を認知的プロセスの表れであるとし、課題を遂行する際の言語使用を社会的現象としてとらえ、社会的・認知的枠組みを提案した (中津原, 2013)。この枠組みでは、妥当性を構成する要素とその要素同士が時間的、概念的にどのように関係しているかを、図 2-3 のように表す。図中の矢印は、要素同士の影響する方向を、時間軸は上から下へと流れる。

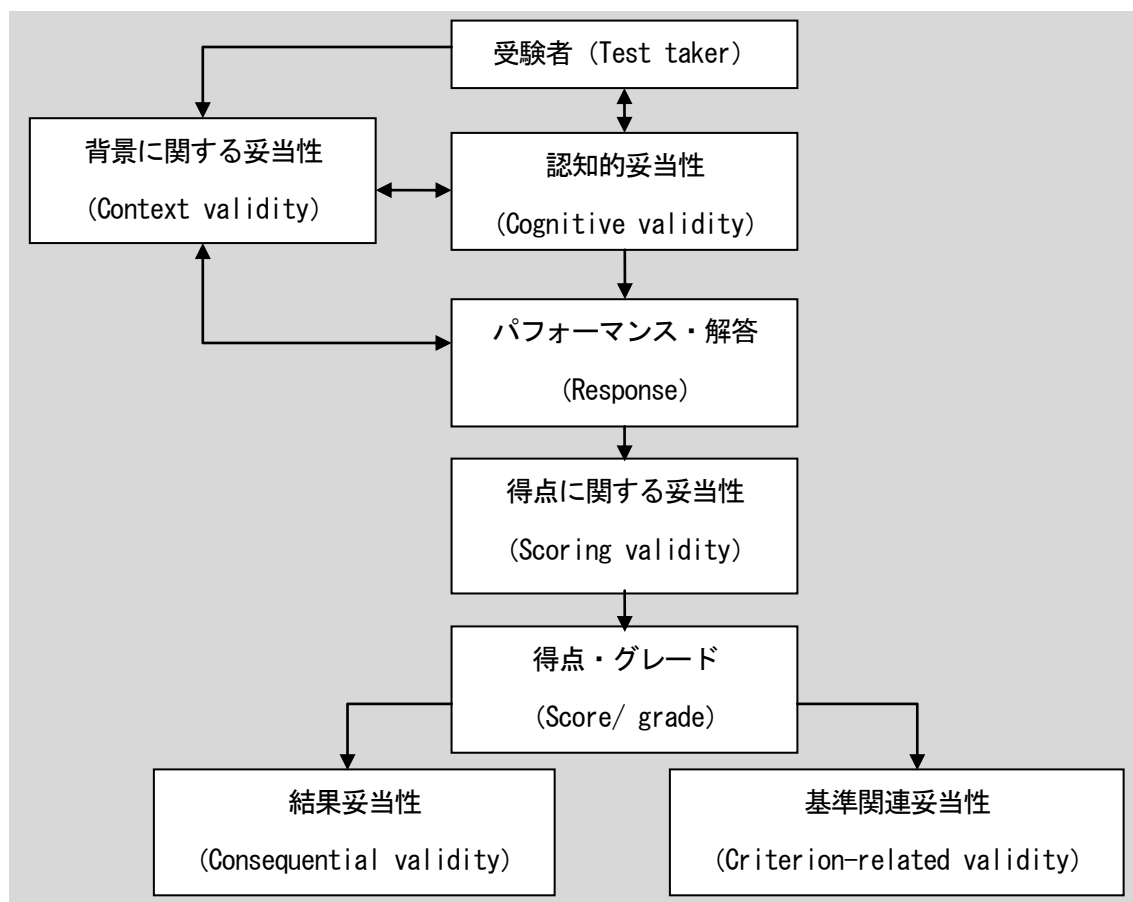


図 2-3 社会的・認知的枠組み (O'Sullivan & Weir, 2011 より作成)

「認知的妥当性」は、テストの課題によって要求される認知プロセスは、測定の目的と照らし合わせて適切であるかどうかを示す。「背景に関する妥当性」は、テストの課題が測定の目的と照らし合わせて適切であるかどうか、受験者にとって公平であるかどうか、また、テストの実施方法が適切であるかを示す。「得点に関する妥当性」は、テストから得られる得点が信頼できるものかどうかを示す。「結果妥当性」は、テストや得点が利害関係者 (stakeholders) にどのような影響を与えるかを示す。「基準関連妥当性」は、テストが外的基準と関係しているかどうかを示す。

中津原 (2013) は、社会的・認知的枠組みにおける言語テストの妥当性検証について、課題の困難度を理解し、学習者の言語能力レベルをより明確に定義するための研究を体系的に進行して総括するのに役立つと述べている。中津原 (2013) はまた、この枠組みで検証されるテストの妥当性のデータに基づく研究は、能力基準としての Cds (4.1.1.1.2 節で詳述) を開発・改善するための重要な資料となりうると述べている。本論文での Cds の開

発・改善については、第4章で扱う。

2.2.5 信頼性と妥当性

南風原（2002）は、テストで得られた測定値（ x ）を構成するのは「測りたい構成概念を反映する成分（ c ）＋系統誤差（ s ）＋偶然誤差⁶（ e ）」であり、妥当性を高めるには、「 c の分散を大きくし、 s と e の分散を小さくすれば良い」と述べている。そして妥当性と信頼性の概念的な定義を、

$$\text{妥当性} = \frac{\text{測りたい構成概念を反映する成分 (c) の分散}}{\text{測定値 (x) の分散}}$$

$$\text{信頼性} = \frac{\text{真値 (c+s) の分散}}{\text{測定値 (x) の分散}}$$

と示し、南風原（2011）では、「この定義から妥当性 ≤ 信頼性という関係が導けるので、妥当性と信頼性の間に、『信頼性が低ければ妥当性も低い』『信頼性が高くても妥当性が高いとは限らない』『妥当性が高ければ信頼性も高い』といった関係のある」こと、すなわち、信頼性は妥当性の必要条件（上限）であることがわかるとしている。繁樹（1998）もまた、測定の最終的な目標は、妥当性を持つことであり、その意味では、妥当性のほうが信頼性よりも重要であるが、安定性の欠けている測定値に妥当性を望むことはできず、信頼性は、妥当性の必要条件と位置付けることができると述べている。つまり、測定が妥当なものであれば、測定の繰り返しごとに結果が大きく変わることはなく、複数のテストフォームの間で結果が大きく異なることもないはずである。しかし、測定値が一貫していても、それが測りたい能力を測っていなければ、測定の妥当性は低くなる。すなわち、信頼性は妥当性の十分条件ではないということになる。

言語テストを作成する際に問題になるのは、高い信頼性を確保するために妥当性を犠牲にしなければならない、妥当性を高めようとすれば信頼性を犠牲にしなければならないということである。これは、2.2.3 節の 4) 内部一貫法で述べたように、 α 係数はテスト得点がある程度一貫しているかを表すが、それはつまり、内容ができるだけ似通った項目を集め

⁶ 系統誤差(systematic error)と偶然誤差(random error)はそれぞれ、2.2.3 節で Alderson et al. (1995) が述べた、体系的な理由と非体系的な誤差に当たる。

れば内部一貫性が高まり、信頼性が高まるが、測定したい特性（構成概念）の一部しか測定できない。一方で、測定したい特性には一定の概念幅（帯域幅）があり、妥当性を高めるには、その概念幅をできるだけカバーするような項目を作成するのがいいことになる。この場合は、構成概念の広範囲を測定できることになるが、内部一貫性（忠実度）は低くなる。これを「帯域幅と忠実度のジレンマ」と呼び、信頼性と妥当性の双方を同時に満たすことの難しさを表している。しかしながら、信頼性のない妥当性というのはありえない。Alderson et al. (1995) は、信頼性と妥当性について「どちらにより重きをおくかはテストの目的にかかっているものであり、不正確な結果を得るとどのような帰結をもたらすのかを考慮する必要がある」と述べている。測定においては、妥当性のある尺度を使って、信頼性の高いデータを得ることが重要である。

2.2.6 項目反応理論 (Item Response Theory : IRT)

IRT は、1940 年代に誕生し、Lord (1952) によって基礎が確立され、1960 年代から欧米において活発に研究が始まった理論である。IRT は、受験者の能力とテスト項目への正答確率との間に、ある種の関数関係を仮定して、それによって項目の特性と受験者の能力とを記述していこうとする統計モデルである (池田央, 2001)。例えば、ETS が実施する TOEFL など IRT で運用されている。日本のテストでは、医療系大学共用試験 (医療系大学間共用試験実施評価機構) や情報処理技術者試験 (通産省)、日本留学試験 (日本学生支援機構)、日本語のテストでは、日本語能力試験 (国際交流基金, 日本国際教育支援協会)、BJT 等で IRT が用いられている。

従来から用いられていたテスト理論である、古典的テスト理論 (CTT) では、素点 (正答数に基づく得点) を用いて受験者の測定結果を表示するため、同一の受験者であっても、テストに含まれる項目の難易度が異なれば、得点も変化することになる。また、同一のテスト項目であっても、受験する集団が異なれば、項目特性も集団ごとに異なる値を得ることになる。これに対し、IRT で算出され、等化 (2.2.6.2 節で詳述) された項目特性を表すパラメタは、テストを受験する誰にとっても同一の尺度として使用することが可能な、受験者に依存しない不変的なパラメタである。同様に、受験者の潜在特性を表すパラメタは、テストやテスト項目に依存しない不変的なパラメタである (中村, 2008)。つまり、テスト項目の難易度と受験者の能力を分離して検討することが可能となる。CTT と比較して IRT の特徴を示すと、表 2-9 のようになる (大友, 2009)。

表 2-9 CTT と IRT の比較（大友，2009：1010 より作成）

領域	古典的テスト理論（CTT）	項目応答理論（IRT）
テスト得点	順序尺度	間隔尺度
受験者能力	テストに依存	テストに依存しない
項目特性	受験者集団に依存	受験者集団に依存しない
測定の精度	受験者集団全体	受験者個人ごと

ただし、IRT の前提条件として、局所独立の仮定と一次元性の仮定を満たす必要がある。局所独立の仮定とは、あるテスト項目に正答できる確率が、ほかの項目に正答できる確率に影響を受けないこと、例えば、項目 j の結果を利用して項目 k を解くというように、「項目間に特別な関係がない」という条件で成り立つということである。局所独立であるということは、そのテストに含まれる項目群に共通の潜在特性がただひとつであることを意味しているため、テストが一次元を満たしているということに等しい（野口ほか，2014）。本論文では、塗師（1989）が述べている因子分析に基づく方法、すなわち、項目間の四分位相関係数行列の最大固有値として得られる第一因子の分散の大きさが、テスト得点の全分散に占める比率を調べる方法に基づいて、一次元性の仮定を満たしているかの確認を行っている。

2.2.6.1 項目特性曲線

IRT では、各テスト項目の統計的性質を、項目特性曲線（Item Characteristic Curve：ICC）で表す。ICC は「潜在特性尺度値 θ の値を連続的に変化させたときに対応する正答確率の変化の様子を視覚化したもの」（野口ほか，2014）で、横軸に受験者の潜在特性値 θ を、縦軸に対応する正答確率 $P(\theta)$ を図示した曲線である。ICC の確率分布モデルにはいくつかのモデルが提唱されているが、本論文で使用されているモデルは 2 パラメタ・ロジスティック・モデル（2PLM）および 3 パラメタ・ロジスティック・モデル（3PLM）である。

2PLM では、テスト項目 j の正答確率 $P^{[2]}_j(\theta)$ を次のように表す。

$$P^{[2]}_j(\theta) = \frac{1}{1 + \exp(-1.7a_j(\theta - b_j))} \quad (2.5)$$

受験者の潜在特性値 θ は、値が高いほど高い能力を持った受験者であることを示す。 $P(\theta)$ は潜在特性値 θ の受験者が項目 j に正答する確率である。 a_j は、項目 j の受験者の能力レベルを識別する力（識別力パラメタ）を表し、値が高いほど受験者の能力レベルを敏感に識別することを意味する。 b_j は、テスト項目 j の難しさ（困難度パラメタ）を表し、値が低いほど易しい項目、値が高いほど難しい項目であることを示す。 b_j と θ は同一の基準で比較することができ、例えば、 $\theta = t$ の潜在特性値を持つ受験者が項目 j に正答する確率が 50% であるとき、項目 j の困難度は $b_j = t$ となる。

2PLM の ICC の例を、図 2-4 に挙げる。縦軸は正答確率、横軸は潜在特性値 θ を表す。能力が高くなるにつれて、いずれの項目も正答確率が 1 に近づいていく。

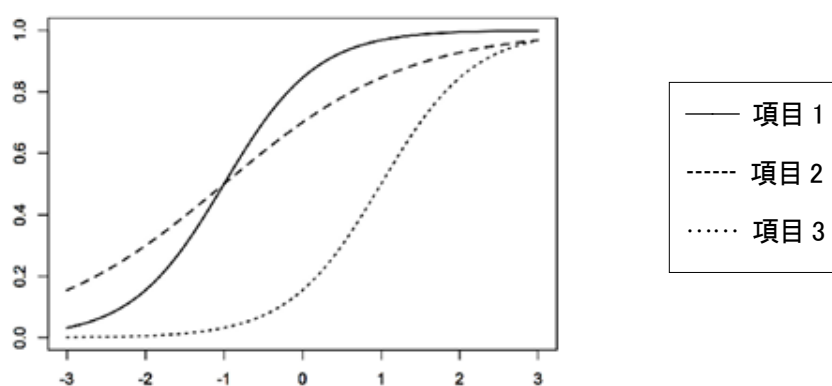


図 2-4 2PLM の ICC の例（奥村，2014）

実線で示した項目 1 と点線で示した項目 3 の識別力パラメタは 1.0、破線で示した項目 2 は 0.5 である。これを見ると、識別力パラメタが高いほうが曲線の傾きが大きくなっていることがわかる。次に、項目 1 と項目 2 の困難度パラメタは -1.0、項目 3 は 1.0 であり、パラメタの高い項目 3 は項目 1、2 と比べて右側に位置していることがわかる。つまり、難しい項目ほど右に位置し、易しい項目ほど左に位置する。

3PLM では、識別力パラメタ、困難度パラメタに、適当に選択肢を選択しても偶然に正答できる確率を表す当て推量パラメタ c_j （ c が大きいほど項目反応関数の下限値が大きくなるため、下方漸近母数とも呼ばれる）を加えたモデルで、テスト項目 j の正答確率 $P^3(\theta)$ を次のように表す。

$$P^{[3]}_j(\theta) = c_j + (1 - c_j) \times \frac{1}{1 + \exp(-1.7a_j(\theta - b_j))} \quad (2.6)$$

3PLM の ICC の例を、図 2-5 に挙げる。実線で示した項目 1 の識別力パラメタは 1.7, 破線で示した項目 2 は 1.45, 点線で示した項目 3 は 0.8 であり, 2PLM と同様に, 識別力パラメタが高いほど傾きが大きい。困難度パラメタは, 項目 1 が -0.8, 項目 2 が 0.5, 項目 3 が 0.1 であり, 最も困難度の高い項目 2 が最も右側に位置している。

当て推量パラメタは, 項目 1 が 0.15, 項目 2 が 0.26, 項目 3 が 0.08 である。最も当て推量パラメタの高い項目 2 の下方漸近線 (潜在特性値 θ の負の側の漸近線) が, θ が低くなっても 0 に近づかず, 潜在特性値の低い受験者でも, ある程度あてずっぽうで正答を得られることを意味している。

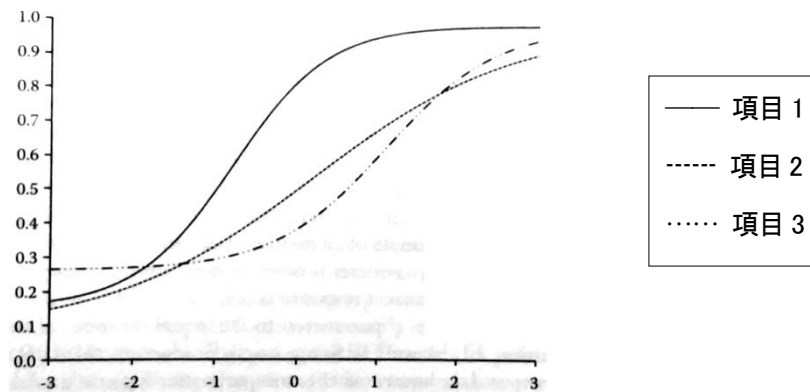


図 2-5 3PLM の ICC の例 (Nering & Ostini, 2010 : 25)

2PLM と 3PLM をひとつの座標軸に表すと, 図 2-6 のようになる (藤森, 1991)。なお, 図中の P_2 は, (2.5) 式を表す。

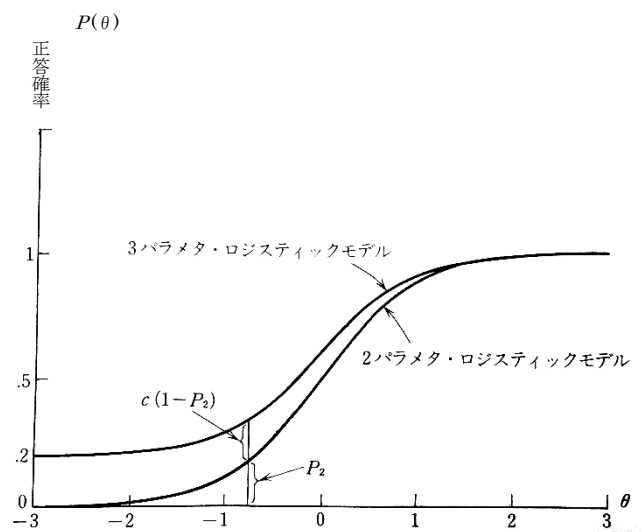


図 2-6 ICC の例 (藤森, 1991 : 36)

2.2.6.2 IRT に基づく等化

特に大規模試験では、問題漏えい対策や緊急時の対応などのセキュリティのために、異なるテスト項目で構成された複数のテストフォームを作成する必要がある。それら複数のテストフォームの得点を比較可能にするためには、等化を行う必要がある。

得点等化とは、異なる試験を受験した人の結果から、試験の性質（難易度等）による影響を排除して、受験者の能力を、共通の尺度上に表示して比較可能にすることである。IRT に基づく得点等化では、統計的性質があらかじめわかっているテスト項目（アンカー項目と呼ばれる）について、受験者の潜在特性値と各テスト項目に正答する確率が対応付けられており、これらのテスト項目に対して受験者が正答を得ているかどうかという情報に基づいて、それぞれの受験者の潜在特性値を推定する。IRT を用いて構築された項目バンク

（Item bank：テスト項目を保存しているデータベースで、新たにテストフォームを構成する際に適切に利用できるように、項目バンクに格納されているテスト項目は内容や難易度等によって整理・分別されている）に含まれる項目は、すべて共通の尺度上に位置づけられた項目パラメタを持つ。したがって、項目バンクに含まれるテスト項目を利用して複数のテストフォームを作成した場合、たとえそれぞれのテストフォームに含まれる項目が異なっても、推定される受験者の潜在特性値 θ は互いに比較可能である。ただし、新しいテスト項目を項目バンクに追加する際には、何らかの形で、新しいテスト項目の項目パラメタの値を項目バンクの尺度に合わせる必要が生じるが、この作業のことを項目パラ

メタの等化と呼ぶ。

IRT の尺度では、項目パラメタ値や潜在特性値を表現する尺度の目盛りの原点と単位とは線形変換の範囲内で自由に決定することができる（原点と単位の不定性）。このため、異なる原点と単位を持つ複数の IRT 尺度は適切な線形変換を行うことによって、すべて共通の原点と単位を持つ共通尺度に合わせられる。たとえば、新たにテストを実施するごとに、IRT 尺度は当該受験者集団で潜在特性値の平均が 0.0、標準偏差が 1.0 となるように原点と単位が定められ、その上で項目バンクの IRT 尺度に変換する（野口ほか，2014）。

等化には、共通受験者あるいは共通項目を用いる場合がある。共通受験者を用いる場合は、同じ受験者集団がほぼ同時期に複数の異なるテストフォームを受験することによって等化を可能にする。共通項目を用いる場合は、複数のテストフォームに共通のアンカー項目を挿入することによって等化を可能にする。これらはテストをどのようにデザインするかによって決定される。本論文で資料として扱う BJT については、項目パラメタならびに受験者の得点の等化は、共通項目を用いて行っている。

新しいテスト項目の項目パラメタを項目バンクの尺度に等化するには、等化係数を求める必要がある。等化係数は、新しいテスト X の項目の項目パラメタを項目バンクの項目パラメタに線形変換する際の定数（ k , l で示す）である。共通項目を用いた等化では、テストフォームごとに推定された共通項目の ICC がなるべく一致するような等化係数を求める。具体的には、新しいテスト X の IRT 尺度における潜在特性値 θ を項目バンクの尺度における θ_z に変換するとすると、識別力パラメタ a は、

$$a_z = \frac{1}{k} a \quad (2.7)$$

困難度パラメタ b は、

$$b_z = kb + l \quad (2.8)$$

当て推量パラメタ c は、

$$c_z = c \quad (2.9)$$

と変換される（ただし、 $k > 0$ ）。また、受験者の潜在特性値についても、

$$\theta_z = k\theta + l \quad (2.10)$$

と変換されるが、変換しても、項目の識別力、困難度、受験者の潜在特性値の度合いに変化はなく、それを表現する目盛りの原点と単位とが変わるだけである（野口ほか, 2014）。このことは、(2.6) 式を

$$P(\theta|a, b, c) = \frac{1-c}{1+\exp(-1.7a(\theta-b))} + c \quad (2.11)$$

としたときに、

$$P(\theta|a, b, c) = P(\theta_z|a_z, b_z, c_z) \quad (2.12)$$

であり、どちらの尺度においても確率は変わらないことを示している。

2.2.7 項目分析 (Item analysis)

テストの評価においては、受験者の解答データから得られる項目情報が貴重な手がかりとなる。したがってテストの実施後は、受験者から得られた各テスト項目の解答データを分析することによって、その項目に関する情報を得ることが重要である。特に項目バンクを持つような大規模試験において項目分析は必要不可欠なプロセスである。Livingston (2006) によると、大規模試験では、測定の過程の中で項目分析が行われるポイントが3つあり、どのポイントで行うかによって目的や結果の使われ方が異なるという。

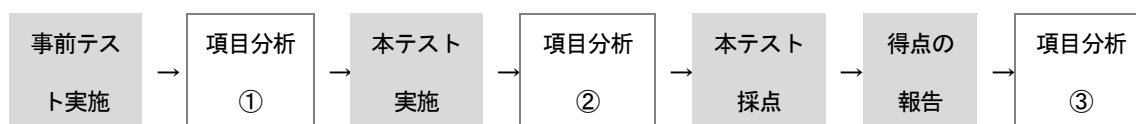


図 2-7 項目分析の実施ポイント

項目分析① 本テストを受験する人と似たテスト受験者グループに対して事前テストを実施することによって、本テストで実際に使うテストのフォームに含める前に

欠陥項目を特定したり直したりすることができる。

項目分析② 採点から除外する項目を特定する。事前テストを行わないテストについて特に重要であり、どんなテストにとっても有用な品質管理のステップである。

項目分析③ 将来のテストフォームで再利用するための項目を選ぶ助けとなる。

Livingston (2006) はまた、項目分析によって得られる項目情報として、CTT における困難度の情報である項目困難度 (proportion correct)、識別力の情報である双列相関係数 (biserial correlation) あるいは点双列相関係数 (point-biserial correlation)、また、IRT における困難度パラメタおよび識別力パラメタ、さらに DIF があるとしている。ETS では、実施するテストにおいてこれらすべての項目情報を得て、厳格な審査基準を満たしたテスト項目のみを出題しているという (ETS「ETS のテストに対するアプローチ」)。

2.2.7.1 BJT における項目分析

2.1.3.1 節でも述べたように、BJT では、IRT に基づく得点等化を行い、尺度得点を算出して受験者に提示している。得点等化を行うためには、過去に出題済みのテスト項目をアンカー項目として再使用することが必要となるが、アンカー項目として再使用が可能かどうかを検討する際には、CTT あるいは IRT による項目分析を行っている。BJT が項目分析を行う目的は、図 2-7 の項目分析③のような、アンカー項目としてテスト項目を再使用することが可能かどうかの検証ということもあるが、それだけでなく、テスト項目の統計的特性に異常値が発生した場合に、その原因を探ることによって今後のテスト項目作成の方針に具体的なフィードバックを与え、より精度の高いテスト項目の蓄積に役立てることでもある。

この分析には、日本語教育や教育測定の専門家委員で構成されるフィードバック委員会を設置し、「BJT ビジネス日本語能力テストフィードバックレポート (以下、フィードバックレポートとする)」を作成して行っている。フィードバックレポートは、統計的数値によって改善を要するかどうかの一定の基準を設け、それによって異常値であると判断されるテスト項目を抽出し、その解答の傾向の内実について検討した内容を記述したものである。フィードバックレポートは完成後、テスト項目作成関係者に開示され、テスト項目作成上の新たな問題点の発見や、項目のさらなる精緻化に役立てられた。

本節では、フィードバックレポートの対象項目の統計的な基準や設問回答率分析図を用

いた項目分析の方法，そしてフィードバックレポートの対象項目における質的な分析の方法について述べる。なお，本節で扱う BJT フィードバックレポートは，当時のフィードバック委員会によって作成されたものであり，現在の実施団体である漢字能力検定協会における BJT の公式見解を示すものではない。

2.2.7.1.1 フィードバックレポートの基準

(1) IRT 項目パラメタ

IRT にはいくつかのモデルがあるが，BJT では「識別力パラメタ」「困難度パラメタ」「当て推量パラメタ」の 3 つの項目パラメタを用いて推定をする 3PLM を採用しており，このパラメタが表 2-10 の基準に当てはまる項目について，フィードバックレポートの対象となっていた。

表 2-10 フィードバックレポートの基準(1)

IRT パラメタの基準（いずれかあるいは複数の条件に当てはまる）
・IRT 困難度パラメタ 3.0 以上または-3.0 以下
・IRT 当て推量パラメタ 0.4 以上

(2) 点双列相関係数による識別力

JETRO が設定した，フィードバックレポートの対象となるテスト項目のもうひとつの基準として，CTT において項目の識別力を示す点双列相関係数がある。点双列相関係数は，その項目の正誤とその項目が含まれているテストの得点との相関係数である。項目－総合点相関，あるいは IT (Item-Total)相関などと呼ばれることもある。この点双列相関係数で示される指標は，その項目がそのテストで測定している特性をどの程度とらえているか，つまり測定の内部一貫性を表すものである。ある項目の得点が，テスト全体で測ろうとしている能力の高さを反映しているならば，点双列相関係数はある程度高い正の値を示すはずである。そうでなければ，その項目が正常に機能していないことが疑われる。一般的に点双列相関係数の値は 0.2 以上が望ましいとされることが多いが，BJT においても同様に，表 2-11 の基準に当てはまる項目について，フィードバックレポートの対象となっていた。

表 2-11 フィードバックレポートの基準(2)

識別力の基準	
・点双列相関係数	0.2 以下

2.2.7.1.2 分析の方法と観点

(1) 設問回答率分析図

設問回答率分析図は、総得点や受験者の潜在特性値 (θ) にしたがって受験者を複数の群に分け、当該群ごとの正答率と各選択肢選択率をプロットしたものである。植野ほか(2007)は、設問回答率分析図について、低得点群ほど低い正答率で、高得点群であるほど高い正答率であれば、当該項目は測ろうとする学力を測定するための識別力があることの強い傍証となる (p.75) としている。同時に、誤答選択肢に関しては、少なくとも学力の最も高い群では、その選択率が正答選択肢の選択率を下回っていることがのぞまれる。

BJTにおいても、出題したすべてのテスト項目について設問回答率分析図を用いて分析を行い、受験者がどの選択肢をどの程度選択しているか、また、群ごとにどの選択肢を選びやすい傾向があるか、そして識別力の度合い (どの群を敏感に識別しているか) などを確認し、出題内容が妥当であったかどうかを検証していた。例として、四肢選択肢形式のある日本語テストにおける設問回答率分析図の典型的な例を図 2-8, 2-9, 2-10 に示す。図 2-8 は正答選択肢[1]の選択率 (正答率) が高得点群 (横軸の LL が最低得点群で HH が最高得点群) になるにつれて増加しており、増加の程度も直線の勾配であるため、受験者の能力の識別に有効なテスト項目であると言える。また、誤答選択肢の選択率については、おおよそ高得点群になるにつれて減少しており、誤答選択肢も有効に機能していることがわかる。

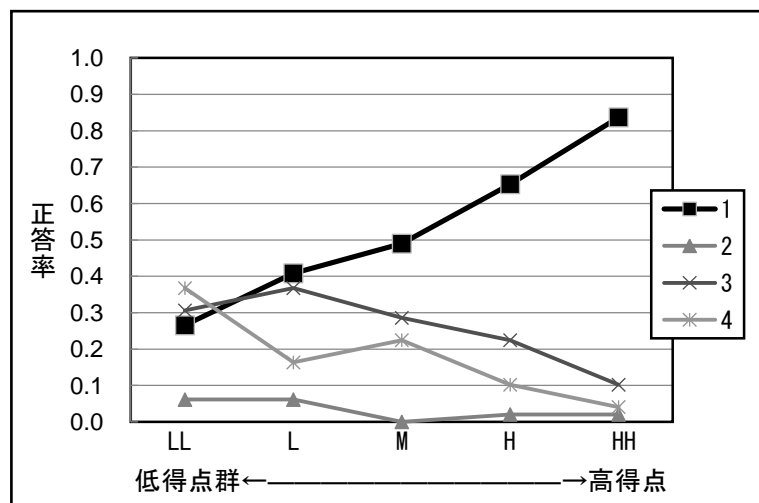


図 2-8 設問回答率分析図の例(1)読解問題[正答選択肢：1]

図 2-9 は正答選択肢[3]の選択率のプロットが折れ曲がり、最低得点群よりも中間得点群の正答率が低いテスト項目である。また、中間得点群の誤答選択肢[2]の選択率が前後の群と比べて高くなっている。これにより、中間得点群が誤答選択肢[2]を選択しやすい特有の原因があることが示唆され、設問や内容の検討を要する。

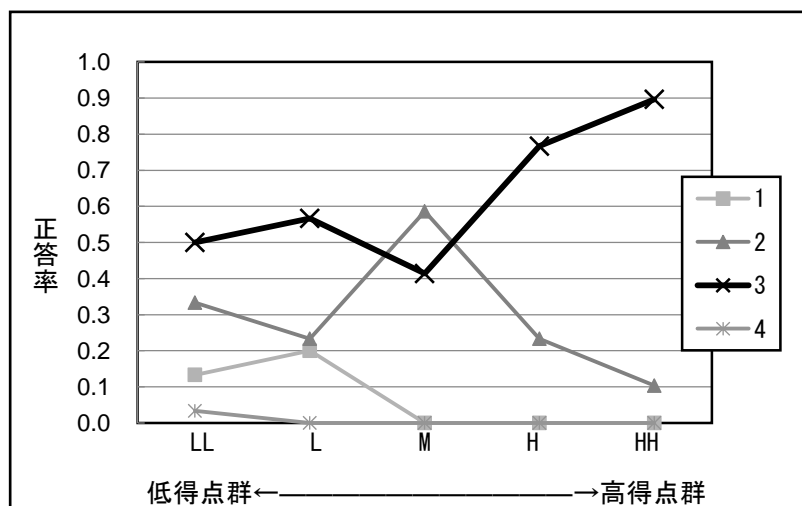


図 2-9 設問回答率分析図の例(2)読解問題[正答選択肢：3]

図 2-10 は正答選択肢[3]の選択率が最低得点群と最高得点群でほぼ同率で、それよりも間の得点群の選択率が低くなっており、受験者の能力を識別しにくいテスト項目である。難易度が高く、最高得点群でも正答率が 0.4 未満である。このような項目の場合、何らか

の要因で、潜在特性値の高い受験者が正答することを阻害する要因が働いており、設問や内容に受験者が読み違えたり誤解したりしやすい、あるいはケアレスミスを起こしやすい原因があったことが考えられ、検討を要する。

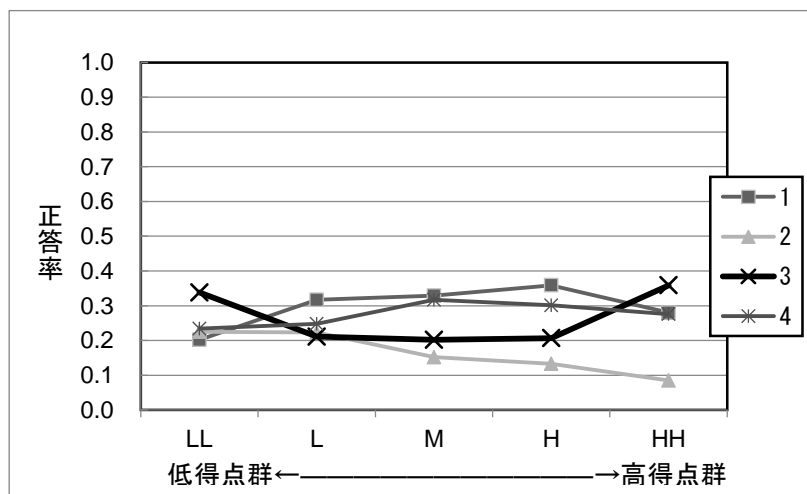


図 2-10 設問回答率分析図の例(3)聴解問題[正答選択肢：3]

(2) 質的分析

前述のBJTのフィードバック委員会では、IRTやCTTの統計情報や設問回答分析図を参照しながら、個々のテスト項目について検証を行っていた。その検証の観点としては、BJTの測定対象である構成概念と無関係の要因が含まれていないかということ、音声・文法・語彙など、言語要素に関する問題はもとより、設問の方法、選択肢の形式、録音のスピード、写真やイラストの情報量、「わたし・あなた」や「これ・それ・あれ」などのダイクティックな表現（意味が文脈や状況に依存して初めて決まる言語表現）の問題、漢字やカタカナ語などの表記の問題、パラフレーズなど正答に至る鍵となる情報の分布状態、そして受験者の視点や焦点など認知言語学的な問題も含んでいた。

2.2.7.1.3 BJT フィードバックレポートの分析内容

本節では、BJT フィードバックレポートの具体的な分析内容について紹介する。なお、JETROが作成したBJTフィードバックレポートには出題した全項目について統計量や設問回答率分析図が提示されているが、ここでは表 2-10、2-11 の基準で抽出されたものの中から再利用不可となったテスト項目をいくつか取り上げることとする。

(1) 聴解 セクション2 表現力問題

【設問】（音声情報）

プレゼンテーションをしています。参加者に質問があるか聞きます。参加者になんと言いますか。

【選択肢】（音声情報）

1. 何かご質問があるのでしょうか
2. 何かご質問がおありでしょうか
3. 何かご質問したいのでしょうか
4. 何かご質問なさりたいのでしょうか

（正答：2）

【視覚情報】



<統計情報>

IRT による：

識別力パラメタ 1.233

困難度パラメタ 2.769

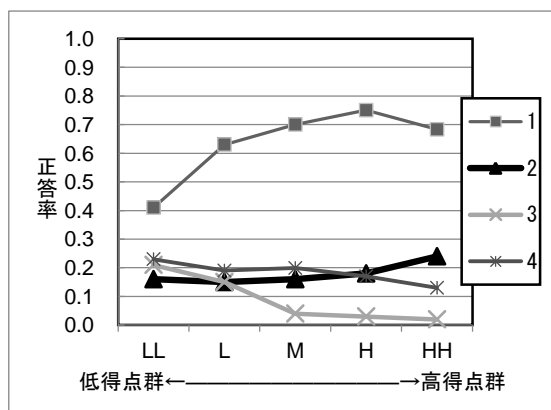
当て推量パラメタ 0.097

CTT による：

正答率 0.126

点双列相関係数 0.121*

<設問回答率分析図>



点双列相関係数が基準値より低い。正答率が0.126と低くIRT困難度パラメタも2.769で難しい項目である。最高得点群（HH）でさえ、誤答選択肢である[1]を選んだ者が63.8%と半数以上を占めた。その他の能力レベルの受験者もいずれも[1]を選んでいる者が最も多い。

選択肢については、正答選択肢[2]「何か質問がおありでしょうか」は、状況依存度が高い表現で最高得点群の中にも学習していなかった者がいたと思われる。また、最も多くの受験者が選んだ[1]の「何か質問があるのでしょうか」は、聴衆が質問をためらっているのではないかと懸念される場合に発するなど、状況によっては使用可能な表現であり完全に間違いとは言いきれないため、誤答選択肢として不適切である可能性がある。

また、視覚情報について、写真には男性一人しか写っておらず、その表情もリラックスしているように見えるため、フォーマルなプレゼンテーションの場であるという感じが伝わらない。背景に写っているホワイトボードも何があるのかよくわかりにくく、問題である。プレゼンテーションを聞いている人々を入れるなどすべきであった。

(2) 聴解 セクション3 聴解問題

【設問】(音声情報)

デパートの店員が、話しています。 部下は、このあと取引先に何を問い合わせますか。

【会話文】(音声情報)

女2: 課長、きのう納品されたカーテン、チェック終わりました。

男1: 新作のカーテン、30セットのことだね。

女2: はい、そうです。

男1: 今度のカーテンも人気が出そうだね。

女2: ええ、そうなんです。前回は約3週間で完売してますから、
今回の新作も同じくらいでの完売を見越しています。

男1: 新作はパステルカラーの3色模様だったよね。

女2: ええ、そうです。前は1色のグラデーション模様で、それはそれで好評でした。
今回の新作に関しては、すでにお客様からの問い合わせも来てます。

男1: うん、ゴージャスな感じだからな。じゃ、追加注文も考えたほうがいいかな。
ボーナスの時期だし、店頭に並べる数を増やすか。

女2: じゃ追加注文しておきますか。昨日在庫は確認済みですがまだ余裕ありますか。

男1: じゃ、とりあえず先方に、ファックスであとどのくらいこちらにまわせるか、
確認しておいてくれる？

女2: はい、わかりました。

【選択肢】(音声情報)

【視覚情報】

1. 商品の納品数
2. 商品の在庫
3. 商品の納期

4. 商品の色

(正答：1)



<統計情報>

IRT による：

識別力パラメタ 0.322

困難度パラメタ 2.887

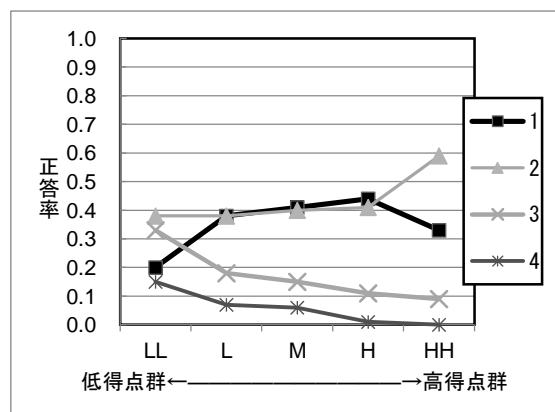
当て推量パラメタ 0.170

CTT による：

正答率 0.353

点双列相関係数 0.199*

<設問回答率分析図>



点双列相関係数が基準値より低く、また、IRT 困難度パラメタが高く、難しい項目である。最高得点群 (HH) が、誤答選択肢[2]を正答である[1]より多く選んでおり、それより能力が低い (H) から (L) は正答を多く選んでいることにより能力の識別が難しい問題となっており、再利用は不可能である。識別力が低い原因としては、1) 納品のプロセスを知らないと考えられる設定になっている、2) 「納品数」と「在庫」は密接に関係しており、切り離すことは難しいため正答が複数になっている、3) デパートの店員同士の会話であることがわかりにくく、受験者が状況をつかみにくい、ということが考えられる。

会話の人物の関係をよりわかりやすくし、選択肢[1]あるいは[2]を変えることによって、改善される可能性がある。

(3) 聴読解 セクション1 描写問題

【設問】（音声情報）

男の人と女の人が話しています。何について話していますか。

【会話文】（音声情報）

男2：これとこれ、どっちがいいと思う？

女2：これは聞いたときの感じがいいよね。

男2：いや、視覚的にはこっちのほうがインパクトあるよ。

女2：うーん、女性向けの商品だし、こっちじゃない、やっぱり。

【選択肢】（文字情報）

【視覚情報】

1. 新製品のスペック
2. 新製品のデザイン
3. 新製品のターゲット
4. 新製品のネーミング

（正答：4）



<統計情報>

IRTによる：

識別力パラメタ 1.504

困難度パラメタ 2.268

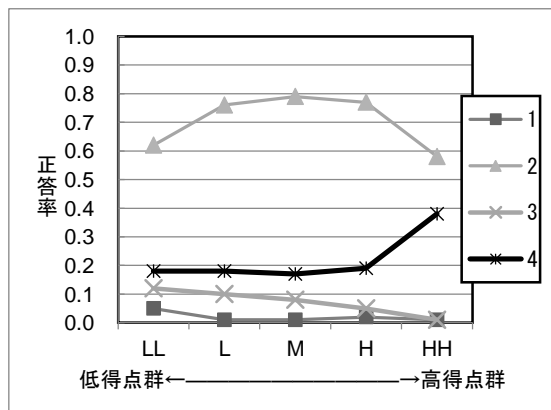
当て推量パラメタ 0.141

CTTによる：

正答率 0.190

点双列相関係数 0.142*

<設問回答率分析図>



点双列相関係数が基準値より低く、また、正答率およびIRT困難度パラメタから難しい項目であることがわかる。最高得点群を始め、すべてのレベルの受験者が正答[4]ではなく[2]を最も多く選んでいる。

会話文の女2の発話「聞いたときの感じ」と言う部分を聞ければ、正答である「ネーミング」につながるが、聞き逃した場合「視覚的には」の部分から「デザイン」につながり、

結果的に[2]が選ばれたものと思われる。正答選択肢につながるキュー（発話）が1か所しかないために、誤った解釈を起しやすいたことが考えられる。

(4) 聴読解 セクション2 聴読解問題

【設問】（音声情報）

上司と部下が、バスの広告のデザイン案について話しています。上司は、どの案がいいと言っていますか。

【会話文】（音声情報）

女2：課長、例の、広告のデザイン案が届きました。

男1：ああ、バスのボディ広告の件ね、どれどれ。

女2：ジュースの「ごくごく野菜」のです。案は、4つですが…。

男1：これは…街中を走るバスの車体の広告だから、目立つのがいいかな。

女2：はい。

男1：うーん。これは、商品名だけか…、面白みに欠けるな。

女2：はい。わたしは、個人的には、この、野菜の並んだのが気に入ってるんですが…。

男1：うん。確かに、視覚的に見せるデザインになってる感じだな。

 だけど、ジュースの商品名を、グッと押し出すには、インパクトに欠けるかな。

女2：そうですね。

男1：この、ジュースのキャラクター、元気そうで、いいね。

女2：ええ。最近、人気が出てきて、テレビのCMでもかわいいって評判が…。

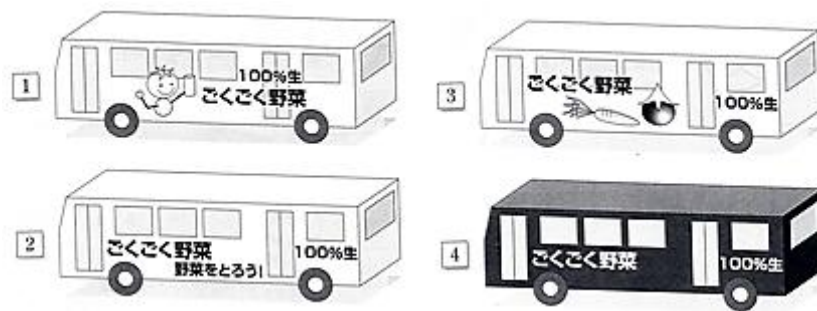
男1：こっちの商品名とキャッチコピーが並んでいるだけのもちよっといただけないな。

女2：味気ないですね。やっぱりジュースの人形ですかね。

男1：そうだな、やはり知名度優先だな。こっちので進めて。

女2：はい。わかりました。

【視覚情報・選択肢】



(正答：1)

<統計情報>

IRT による：

識別力パラメタ 0.402

困難度パラメタ -2.622

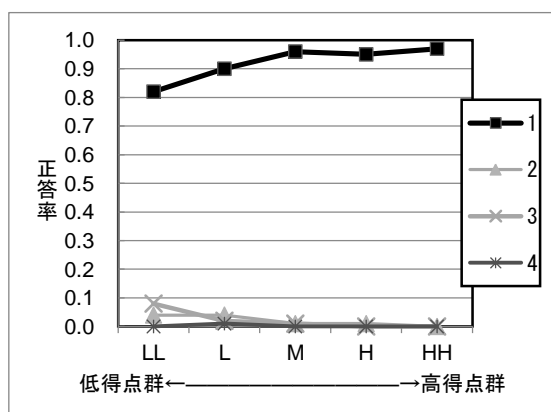
当て推量パラメタ 0.249

CTT による：

正答率 0.915

点双列相関係数 0.188*

<設問回答率分析図>



点双列相関係数が基準値よりも低く、識別力の低い項目である。最低得点群（LL）でさえ 80%以上が正答を選んでおり、IRT 困難度パラメタも-2.622 と易しい。易しくした原因としては、発話中に「かわいい」「元気そう」「キャラクター」「人形」など、正答につながるキューが数多く含まれていることが考えられる。

選択肢について、誤答選択肢がどの群にもあまり選ばれておらず、ほとんど機能していないことが考えられる。たとえばかわいい女の子がジュースを飲んでいる写真を誤答選択肢に入れて迷わしを増やす、または会話文中のキューを減らすなどすれば誤答選択肢の選択率が上がり、困難度や識別力が上がる可能性がある。

(5) 読解 セクション2 表現力問題

【設問】(文字情報)

最近の若手社員の言葉の乱れについては、() ものがある。

【選択肢】(文字情報)

1. 目にあまる
2. 目がくらむ
3. 目をみはる
4. 目にとまる

(正答：1)

<統計情報>

IRT による：

識別力パラメタ 0.876

困難度パラメタ 3.066*

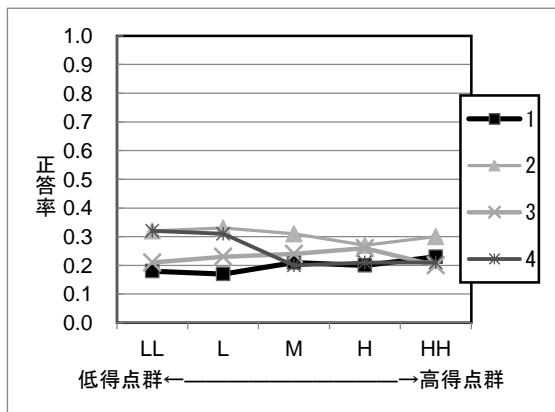
当て推量パラメタ 0.144

CTT による：

正答率 0.199

点双列相関係数 0.194*

<設問回答率分析図>



IRT 困難度が 3.066 と基準値より高く、また点双列相関係数も低い。最高得点群であっても正答率が 28.2%となっており、誤答選択肢である[2]を正答[1]より多く選んでいる。全体としてみても正答が最も選ばれておらず、能力を識別する力の弱い項目であるため、再利用は不可能である。

設問では「言葉の乱れ」に続くネガティブな表現を選択する項目であるが、誤答選択肢についてもどの選択肢が多く選ばれているといった傾向が見いだせないことから、「目にあまる」だけでなく「目がくらむ」「目をみはる」「目にとまる」といった表現を高得点群であっても学習していなかったことが考えられる。このようないわゆる慣用句的な表現を選ぶ項目は受験者の能力の差を反映しない傾向があるため、テスト項目作成には注意が必要である。

2.2.7.2 日本のテストにおける項目分析

ここまで、BJTにおける項目分析の、具体的な基準や分析内容について述べてきた。本節の初めに述べた、Livingston（2006）の項目分析実施のポイントには、事前テスト実施後、本テスト実施後（採点前）、得点の報告後の3つが挙げられているが、BJTのフィードバックレポートは先述したように3つ目の得点の報告後に行われる項目分析である。統計的な分析（IRT項目パラメタや設問回答率分析図を含むCTTによる分析）は本テスト実施後（採点前）にも行われる。しかしBJTをはじめ、我が国のテストではその多くがセキュリティの問題や情報公開の方針から事前テストを行わない。そのため、本テストで出題するテスト項目についての情報を事前に得ることができない。そのような状況下では、BJTのフィードバックレポートのように、テスト項目作成者、データ解析者、そしてテスト理論の専門家といった各方面の専門家によって得られる統計的な観点と質的な観点からの知見は、測定の質をより高めるためのテストやテスト項目作成の指針となりうるだろう。統計的、質的両側面からのテスト項目の分析については、第5章および第6章で具体的に示す。

第3章 構成概念としてのビジネス日本語能力

第2章で概観したビジネス日本語およびビジネス日本語教育に関する先行研究から、これまでのビジネス日本語のとらえ方や、外国人材に対するビジネス日本語教育の内容についての課題や知見を得られた。先述したように、ビジネス日本語やビジネス日本語教育を取り巻く環境は刻々と変化している。そこで本章では、外国人材を採用する企業の多くが、現在求めている日本語能力について、その具体的な内容について分析し、構成概念としてのビジネス日本語能力を明らかにする。ここで明らかにするビジネス日本語能力は、第4章から第6章の分析の基盤となるものである。

3.1 研究背景と研究目的

3.1.1 研究課題の設定

3.1.1.1 研究背景と目的

2.1 節で挙げた先行研究では、ビジネス日本語やビジネス日本語教育が指す範囲は広範にわたっており、ビジネス活動を行う上で必要であるとされているビジネス日本語能力についてもさまざまな指摘があることがわかった。本章では、企業が外国人材に求めるビジネス日本語能力について、各種調査結果から探る。第1章で概要を述べたが、日本企業が外国人材に求めている能力について、最も多いのが「コミュニケーション能力」であり、次に「日本語力」とあるという報告がある（株式会社 DISCO, 2015）。そのほかに、同様の調査を行った研究として、日本貿易振興機構人材開発支援課（2005）は、中国・大連の日系企業 58 社に対し、インタビュー調査を行っている。調査の中で、「日本語人材に望むこと」という質問項目において、主に次のような回答を得ている（自由回答形式、複数回答、有効回答数 33 社、（ ）内は業種）。中国の日系企業でも、日本国内の日本企業と同様にコミュニケーション能力を重視しており、その業種もさまざまである。

- ・ 会話能力と理解能力等のコミュニケーション能力
..... D 社（食品）ほか 6 社
- ・ 本社（本部）とのコミュニケーション（TV 会議、提出書類、電話、メール等）
..... H 社（観光）ほか 6 社

- ・メールと電話ができること K 社（工業製品）ほか 4 社
- ・顧客とのコミュニケーション能力がすべて
..... I 社（電子機器・コンサルティング）ほか 4 社
- ・接客対応の日本語。いろいろな問い合わせに対応できるより高い日本語能力
..... Z 社（航空）ほか 1 社
- ・ビジネスマナー N 社（医薬品）ほか 1 社
- ・機械設備のマニュアルを読むこと DI 社（化学）
- ・日本語での社内会議 DNT 社（加工）
- ・日本語で論文作成 DA 社（設備）
- ・駐在員とのやりとり DR 社（電気機器）
- ・経営の現地化を進めるにあたり、経営を任せられるには完全に近いほどかなりの日本語
が求められる O 社（電気機器）
- ・言語はツールのひとつでありそれほど重視しない
..... S 社（サービス）ほか 4 社
- ・現場（工場等）で敬語能力は重要ではない
..... M 社（産業機器）ほか 1 社

次に、海外技術者研修協会（2007）は、製造業を中心とする日本企業 288 社を対象とし、具体的なビジネス場面において外国人材に期待される日本語能力および現在の日本語能力について、5 段階評定の調査を行っている（「1.ほとんどできない」「2.あまりできない」「3.ある程度できる」「4.おおむね問題ない」「5.全く問題ない」）。その結果、技能別の平均値について表 3-1 のような結果が得られたとしている。要求度が一番高いのは、口頭コミュニケーション能力（「求められる能力」レベルの平均が 4.09）、次いで書記コミュニケーション能力（同 3.87）、複合的コミュニケーション能力（同 3.85）の順になっている。また、具体的なビジネス場面としては、社内会話、社内会議などの社内でのコミュニケーション以外にも、社外との直接的なやりとりや電話・メールによる対応についても求められていることがわかる。この調査結果から、社内・社外問わず、また、会話や読み書きの技能を問わず、企業は外国人材に、おおむね業務の遂行に問題のないレベルの日本語能力を求めていることがわかる。

表 3-1 企業が外国人材に求めるビジネス日本語能力（海外技術者研修協会 2007 より作成）

口頭コミュニケーション能力		
	現在の能力	求められる能力
敬語・丁寧語等	3.9	4.1
相手の依頼・苦情対応	3.8	4.1
依頼・主張	3.8	4.0
相手への確認	3.9	4.2
相談・報告	4.0	4.2
電話での用件連絡	3.6	4.0
電話の要点確認	3.8	4.0
書記コミュニケーション能力		
新聞読解	3.7	3.7
資料スキャニング	3.8	4.0
資料スキミング	3.7	3.8
報告文書作成	3.6	3.9
メール発信	3.7	3.9
受信メールへの返信	3.8	4.0
ビジネスレター作成	3.6	3.8
複合的コミュニケーション能力		
プレゼン理解と要点作成	3.7	3.9
資料作成とプレゼン実施	3.5	3.8
議論理解と意見発信	3.6	3.9
会議理解と報告作成	3.5	3.8

日本漢字能力検定協会（2013）は、東証一部上場企業等 1170 社の人事採用担当者に対し、Web 上で外国人留学生の採用と日本語能力に関するアンケート調査を実施し、139 社からの回答を得ている。調査結果によると、「採用選考時に、重視する日本語スキル」について、Speaking & Listening では、112 社が「相手の考えを理解する能力」、107 社が「自

身の考えを表現する能力」, 次いで 78 社が「TPO や人間関係に配慮して適切に表現する能力」を挙げている。同様に, Writing & Reading では, 108 社が「自身の考えを表現する能力」, 107 社が「相手の考えを理解する能力」, 78 社が「TPO や人間関係に配慮して適切に表現する能力」を挙げている。回答した企業の 9 割が四技能すべてにおいて「相手の考えを理解する能力」, 「自身の考えを表現する能力」を重視していることがわかる。それらを重視する理由としては, 「社内でのコミュニケーションで日本語が必要になるため」, 「日本人顧客の対応をとまなう業務があるため」であり, 当該調査の報告によると, それらの日本語スキルは, 「『できたほうがよい』ことではなく, 『業務上の必要性に迫られた必須の』スキルと言える」としている。また, 「TPO や人間関係に配慮して適切に表現する能力」については, 社外とのやり取りを伴わない業務を除き, そういった「高度なコミュニケーション能力が外国人社員にも求められている」としている。

これらの調査から, 企業は外国人材に対し, 必ずしも正しい日本語を書いたり話したりすることのみを求めているわけではなく, 日本語を道具として, 社内や社外と円滑にコミュニケーションを図れることを望んでいるということがわかる。つまり, ビジネスの現場においては, 言語的知識を用いてさまざまなやりとりの中で相手の考えを理解し, 社内外での人間関係を構築するためのコミュニケーション能力の重要性が指摘されている。

本章では, 外国語でのコミュニケーション能力に関する理論的枠組みを確認し, 企業の多くが外国人材に求めている「ビジネスコミュニケーション能力」について考察する。そして, 「ビジネスコミュニケーション能力」の構成要素⁷を同定し, 構成概念としてのビジネス日本語能力を明らかにすることを目的とする。具体的には, 以下の 2 つの課題を設定し, 研究を行った。

- (1) 外国人材にはどのような日本語能力が必要とされているか。
- (2) ビジネスコミュニケーション能力とは何か。

⁷ 構成概念を構成している要素を構成要素と呼ぶ。野口ほか(2014:43)によると, 熟達度テスト(proficiency test)の場合は, 「言語能力」を新たに定義する必要があり, たとえば 2010 年に改定された日本語能力試験の場合, 全体では「課題遂行のための言語コミュニケーション能力」を測るものとして設計され, その下位の構成要素では「言語知識と, それを利用して課題を遂行する能力」の 2 つからなると定義されているという。

3.2 先行研究

3.2.1 外国語でのコミュニケーション能力の理論的背景

言語や文化の異なる者同士が外国語を使ってコミュニケーションをとるためには、言語能力以外にも様々な知識や態度によって支えられているものと考えられる。ここでは主に Bachman (1990), Bachman & Palmer (1996) の意思伝達言語能力の理論的枠組みを基本として、外国語でのコミュニケーション能力を考察する。

3.2.1.1 コミュニケーション能力の構成要素

Bachman (1990), Bachman & Palmer (1996) は、個人の持つコミュニケーション能力は、まず、その個人の「知識構造（人間社会に対する知識）」「言語能力（言語に関する知識）」、そして「方略的能力」の影響を受けると述べている。さらに、「場面の状況」における個人のコミュニケーション言語能力は、性格や身体の状態などの「心理・生理的機能」の影響も受けると見なされる。外国語でのコミュニケーション能力は、これらによって成り立つと考えられる。

(1) 言語能力 (Language competence)

言語を使用する場面においては、文法力や語彙力といったものだけでなく、多くの能力が重要な役割を果たす。Bachman (1990), Bachman & Palmer (1996) は、言語能力の下位能力の主なものを、図 3-1 のように示した。

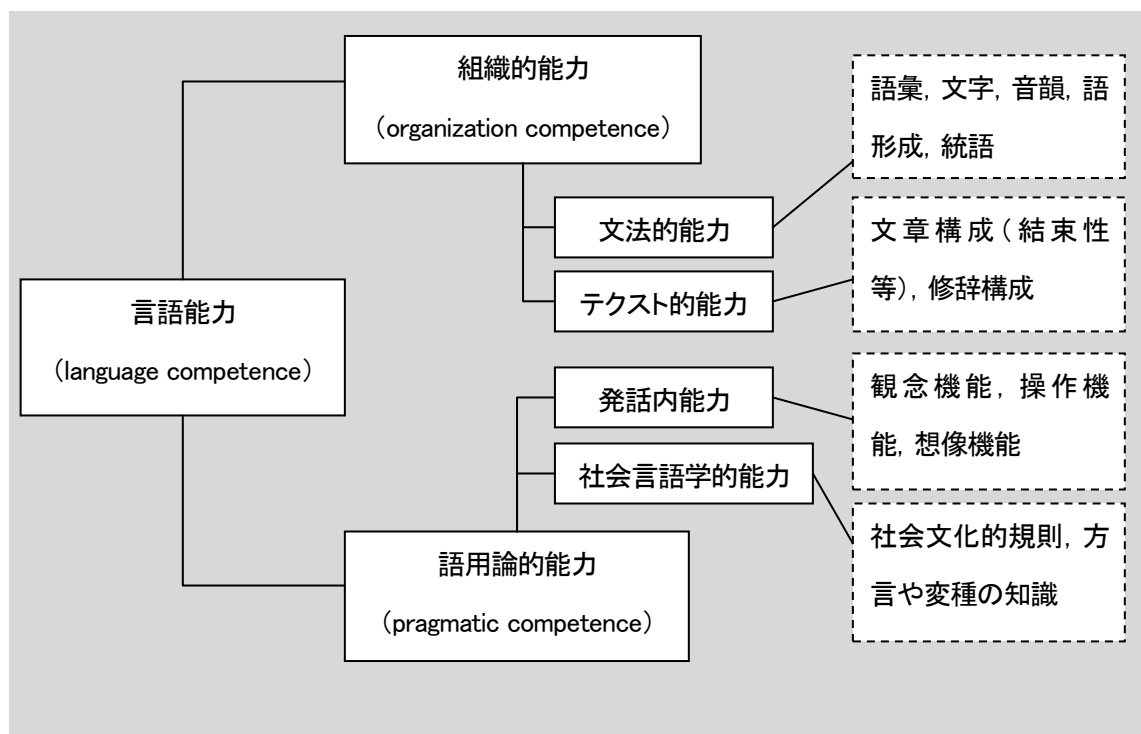


図 3-1 言語能力の構成要素 (Bachman, 1990 : 99)

図 3-1 で示されているように、言語能力は、言語のしくみに関する「組織的能力」と実際の場面で言語を適切に用いるための能力である「語用論的能力」に分けられる。「組織的能力」は、文レベルで言語を正確に理解し、表現することを可能にする文法的能力と、文を超えたレベルでの言語使用を可能にするテクスト的能力に分けられ、「語用論的能力」は、発話内能力と社会言語学的能力によって構成される。

1) 文法的能力

文法的能力とは、学習する言語の語彙、文字、音韻、語形成、統語などの言語体系を適切に使用する能力。また、文法規則などの言語知識を、現実の場面において使用できる能力を意味する。

2) テクスト的能力

ひとつひとつの文が正確だけでなく、まとまりのある文章を書いたり話したりできる能力。まとまりを生じさせる要素間の適切な結びつきは結束性 (cohesion) と呼ばれ、結束性に関する知識は、文章を理解する上で重要な働きをする。また、対話の開始、継続、終了、対話順番の交代 (turn-taking) などを示す表現を適切に使用する能力を意味する。

3) 発語内能力

言語のさまざまな機能を使用する能力。たとえば「情報を伝達する・求める」「意見・判断・考えなどを表現する」「説得する」などがある。

4) 社会言語学的能力

社会文化的な規則を理解して、話す相手や地域など状況によって、また、相手との人間関係や状況に配慮して、適切に言語を使い分ける能力。

(2) 方略的能力 (Strategic competence)

意思伝達方略などのコミュニケーション・ストラテジー(communication strategy)を使用する能力。技巧的で円滑にコミュニケーションを行える能力。

李 (2015 : 24) は、特に Bachman の研究において方略的能力の部分に重きを置いた点が重要であり、「方略的能力はあくまでコミュニケーション上の失敗やトラブルを補完するものとして考えられたが、Bachman (1990) における方略的能力は、言語使用者の既存の知識や言語使用が行われるコンテキストの特徴と関連付け、言語活動を実行することを主な機能とする要素として位置づけている」と述べている。Bachman のモデルは多くの言語テスト開発に利用されてきたが、ここにとどまらず、言語能力の構成要素は何か、あるいはコミュニケーション能力の構成要素は何かという検討を継続していくことがテスト開発・運営にとって重要であると考ええる。

3.2.2 ビジネスコミュニケーション能力

日本企業が外国人材に求める能力としてもっとも多いのがコミュニケーション能力であるが、ビジネスコミュニケーションの定義について、近藤 (2007 : 7) は、「ビジネスのための言語行動と非言語行動を含むコミュニケーション活動全般」としている。また、加藤 (2010 : 184) は、ビジネス活動におけるコミュニケーションにとって最も本質的なこととして、「相手の声の調子、表情やジェスチャー、目の前にある文書や図表・写真など、与えられた情報すべてを手がかりとして相手の意図をくみとり、自分の意思を伝える工夫と努力をし、円滑な人間関係を維持しながらビジネスの目的を達成すること」と述べている。

株式会社アスクが実施する「日本語ビジネスコミュニケーション能力測定テスト」では、

外国人社員の日本語でのビジネスコミュニケーション能力について、「日本語の知識・理解力がどのくらいあるか」「知っている言葉を適切な場面で正しく使う能力があるか」「日本の企業文化に関する理解があるか」「社内外において良好な人間関係を築きながら業務を行えるか」という評価基準により測定をしているという。

以上の知見を踏まえ、本論文では、ビジネスコミュニケーション能力を、「言語的・非言語的情報の助けを借りながら日本語を理解・運用し、ビジネス活動上の課題に対して適切に行動する能力」と定義する。

3.2.2.1 ビジネスコミュニケーション能力の構成要素

芳賀ほか（2015）は、日本国内の大手企業（日本に本社を持つ企業だけでなく、グローバルに拠点を持つ外資系企業も含む）計 32 社のトップ・エグゼクティブおよび人事担当責任者を対象に、半構造化面接法によって、実際のビジネスの場面でビジネスパーソンに求められるコミュニケーション能力についての調査を行っている。収集したデータをテーマ分析した結果、表 3-2 の 7 つのテーマが導きだされたという。

表 3-2 ビジネスの領域におけるコミュニケーション能力（芳賀ほか，2015：93-95 より引用）

テーマ名		定義	言説例
1. コミュニケーションの双方向性		受信力と発信力のバランスがとれている。	情報を獲得することが重要な場面とこちらの考えを伝えることが重要な場面がある。
			コミュニケーションは話すと聴くの片方だけということはない。バランスが大切である。
			コミュニケーション能力とは、相手のいう事をちゃんと聴き、訊かれた事に対してきちんと反応する能力。
2. 発信力	2-1 発信の仕方	話の核とタイミングを押さえ、論理立てて話	必要項目をきちんと考えて報告するという能力を身につけないと、何が必要

		すことで、相手に要点を明確に伝えることができる。	な情報で何が必要でないかもわからなくなる。	
			人に分かりやすく話せるかどうかという部分が非常に大切だと思う。	
		2-2 話の展開への積極的関与	本質を理解して発信できている。	物事の本質を追求する姿勢が重要。
				ぶれないで伝えきる力があること。抽象的な事を上手く伝えられる能力。
				1つ1つにきちんと対応し、的確にきちっと話すという事が大切。
			相手を話に巻き込むことができる。	コミュニケーションでは、人を巻き込む力がないとダメである。
周りの人を巻き込んで大きな成果を上げることができることがコミュニケーションの最終形だと思っている。				
3. アクティブ・リスニング		積極的に相手の話を理解しようという姿勢を持つ。不明なことは明確にする。	情報を収集し、ニーズを積極的にヒアリングしてというのが一番大事な事。	
			相手の聴きたいことを引き出せるか、聴く力がポイント。	
			相手の話が終わったら後に確認作業を行うことが大事。確認作業を怠ると、違う思い込みで双方が進むなど良くない結果になる。	
4. コミュニケーションの基盤	4-1 倫理観／礼節	組織の一員として倫理観かつ礼節を持って意思疎通が行える。	礼儀正しさはコミュニケーションの前提条件である。	
			組織人としての自覚ができていないコミュニケーションはダメだと思う。	
	4-2 開示力	ある程度の自己開示によって、相手との距離を近づけることがで	相手に話をさせるような熱意を持つことは基本の基本である。	
			信頼関係を築くか築けないかは、相手	

		きる。	を緊張させない力、相手に好印象を持つといったコミュニケーションの入り口の力にかかっている。
	4-3 能動性	対立を恐れず、伝えるべきことをしっかりと伝える。	悪い情報ほど早く的確に伝えること。 最近の若手は、コンフリクトを恐れてコミュニケーション自体を避けてしまっている。 わからないことはお互いにきちんと説明する。わかるまで訊くというコミュニケーションスタイルを身につけて欲しい。
	4-4 全体像を把握する能力	全体像をきちんと把握し、周囲の状況を客観的に捉えることができる。	自分のやっている事が外部とどのようにつながっているのかを考える。 何のためにここににいるのかを意識していない、全体像や自分の置かれている立場を理解しないまま説明するなど論外。
	4-5 文章力	自分の言いたいこと、言うべきことをしっかりと文章で表すことができる。	「拝啓」で始まり「敬具」で終わるような手紙を日常書いていない。 最近の若い人は、文章がとっても苦手な人が多いと思う。もっと文章とか文字を大切にしてほしい。
5. 率直性		遺恨を残さず率直に話すことができる。	上司に対しても、耳障りが良くない事をしっかり言える人が良い。 相手の事は配慮しながらも、尚且つきちつと言う事が重要。 言いにくい事をちゃんと言うタイプの人 がコミュニケーションができる人だと思う。
6. 曖昧さの回避		物事を曖昧にせず、	会議での理解の食い違いなどは、曖

	結論を明確にする。	味に伝えていることが原因の場合が多い。日本的な「あうん」や「曖昧」がまだあると感じる。
		曖昧にしないで、非常にクリアにすることが昔より求められている。
7. 報告の重要性	タイミング、順番、中身の信憑性などに基づき論理的に報告をする。	場面にあったコミュニケーションがきちんと図れるかが問題。
		理論立てて説明することの大切さを考えることが少なすぎると思う。
		ポイントを押さえて簡潔に報告ができる事はビジネスにおいて非常に重要。

芳賀ほか（2015）の調査結果は、企業で働く外国人材に求められるコミュニケーション能力を想定して行われたものではないが、調査によって引き出された7つのテーマについて、「組織で活躍でき、重要な一員として認知、評価を得るために必要なコミュニケーション能力」（p.95）であると述べている。そして、特に外国語教育、コミュニケーション能力の実践的指導では、「読む・書く・聞く・話す」の四技能を独立したものとして扱うのではなく、「相互に影響を与え、バランスが取れた状態となって初めて『コミュニケーション能力がある』と考えられていることがうかがえた」（p.96）と考察している。2.1.2.1 節で、AOTS（2011）のビジネス日本語教育の対象が、日本人であるか外国人であるかを問わず、一社会人として日本企業で働く人材の育成が必要との認識があると述べたが、そのような認識で行われている昨今のビジネス日本語教育においては、芳賀ほか（2015）の調査結果はビジネスコミュニケーション能力の養成を考える上で参考になるものである。

なお、田崎ほか（2015）は、芳賀ほか（2015）の調査で得られた、コミュニケーション能力の7つのテーマを基盤として、ビジネスコミュニケーション能力を測定するスキルテストの開発を行っている。「日本企業のトップ・エグゼクティブを通じて浮かび上がった理想的なコミュニケーション像を基に、ビジネスの現場で実際に起こっている状況を加味した実践型のテスト」（p.137）の開発を目指して作成した90項目のスキルテストは、IRTによる項目分析の結果、十分な妥当性と信頼性を有していることが示されたという。田崎ほか（2015）の報告は、7つのテーマで示されたビジネス場面でのコミュニケーション能

力の構成要素をもとに、ビジネスコミュニケーション能力という構成概念を測定するテストを作成した実例であり、今後、コミュニケーション能力を測定するテストの開発に有益な示唆を与えている。

池田（2001）は、ビジネス日本語教育において実用的なコミュニケーション能力を養成するために、コミュニケーション・ストラテジーの能力が重要であるとして、初級段階からの指導の方法について提案し、その習得における有効性の検証を試みている。池田（2001）の指摘は、3.2.1.1 節で述べた Bachman（1990）、Bachman & Palmer（1996）が外国語でのコミュニケーション能力の構成要素として挙げた「方略的能力」に焦点を当てたものであると考えられる。

3.3 考察 構成概念としてのビジネス日本語能力

Bachman & Palmer（1996：131）は、「（言語能力の）構成概念の定義には、言語能力のうちの特定の構成要素を含めることになる」と述べている。これを踏まえて、構成概念としてのビジネス日本語能力について以下に考察する。

2.1.1 節で述べたように、本論文においてビジネス日本語能力が指す内容は、「ビジネス場面でコミュニケーションを行うための日本語の能力」であり、すなわち「ビジネス日本語能力」は「ビジネスコミュニケーション能力」と同義である。そして、3.2.2 節では、ビジネスコミュニケーション能力を「言語的・非言語的情報の助けを借りながら日本語を理解・運用し、ビジネス活動上の課題に対して適切に行動する能力」と定義した。したがって、ここで検討する構成概念は、単に言語能力としてのビジネス日本語能力ではなく、ビジネス場面での日本語によるコミュニケーションの要素で構成された概念ということになる。

3.2.2 節では、ビジネスコミュニケーション能力とその構成要素について、先行研究の定義や日本語学習者、企業関係者に対する調査から探った。これらを、3.2.1 節で示した、Bachman（1990）、Bachman & Palmer（1996）の外国語でのコミュニケーション能力の枠組みで考えると、ビジネスコミュニケーション能力の構成要素は次のように同定されることが考えられる。

(1) 言語能力

- ・ビジネス場面で使用される語彙や表現，文法に関する知識
- ・文書や対話から，情報を読み取ったり聞き取ったりする能力
- ・自分の考えを筋立てて伝達し，それに対する相手の要求や意図を読み取る双方向の能力
- ・目的，状況，相手との関係，話題などを考慮しながら，適切に日本語を使い分ける能力
- ・ビジネス課題に直面したときにその状況や相手に配慮しながら，場面に適した行動，自分の取るべき行動を選択できる能力

(2) 非言語能力

- ・相手の表情やジェスチャーを読み取る能力
- ・日本の企業における文化的背景を理解し，調整する能力
- ・能動的かつ率直にビジネス活動に取り組む姿勢

(3) 方略的能力

- ・コミュニケーション・ストラテジーを用いてコミュニケーション上の問題点を解決する能力

日本語教育および企業関係者に対する調査の結果から示される，「外国人材に求められているビジネス日本語能力」は，以上に挙げたビジネスコミュニケーション能力の構成要素から成る構成概念であると考え，これを本論文における構成概念としてのビジネス日本語能力とする。しかし，2.1.2.1節で述べたように，昨今のビジネス日本語教育の目標が，「日本語や日本文化に精通している人材の育成」から，一社会人としての「包括的能力の育成」と変化している背景においては，ここに挙げたビジネスコミュニケーション能力の構成要素をすべて身に付ければ十分であるというものではないだろう。グローバル人材一般に求められる，異文化理解能力（Cross-cultural communication skills），グローバルな視点（Global perspective），リーダーシップ，主体性など（Adler & Bartholomew, 1992；グローバル人材育成推進会議，2011；鈴木，2012）も，ビジネスコミュニケーション能力の構成要素として含めるべきかもしれない。ただし，ビジネスコミュニケーション能力は，現実的には，職種や業務形態，企業環境などによって，また，個人や集団，組織によって必要となるものは異なると言えるし，日本や世界の経済状況によって変化もするだろう。

つまり、ここで挙げるビジネスコミュニケーション能力は、田崎ほか（2015）が述べているように、個別のビジネスシーンの間で共有される通状況的な傾向を示しているに過ぎない。さらに、近藤ブラウン（2012：33）が述べているように、コミュニケーション能力とは「いくつかの構成要素（一般教養と言語に関する諸知識）からできているだけでなく、与えられた状況場面や、話者の心理身体的な状態によっても変化する、非常に複雑な構成概念」と言えるだろう。

3.4 第3章のまとめと今後の課題

本章では、まず、外国人材に求められている日本語能力について、その具体的な内容を先行研究の調査により明らかにした。その結果、企業は外国人材に対し、必ずしも正しい日本語を書いたり話したりすることのみを求めているわけではなく、日本語を道具として、社内や社外と円滑にコミュニケーションを図れることを望んでいるということがわかった。このことから、ビジネスの現場においては、言語的知識を用いてさまざまなやりとりの中で相手の考えを理解し、社内外での人間関係を構築するためのコミュニケーション能力が重視されていることが示唆された。

外国語でのコミュニケーション能力とは何であるか。3.2.1.1 節では、個人の持つコミュニケーション能力に関する Bachman（1990）、Bachman & Palmer（1996）の理論的枠組みを整理した。次に、ビジネスコミュニケーション能力とその構成要素について、先行研究の定義や日本語学習者、企業関係者に対する調査からその内容を探った。そしてその結果を Bachman（1990）、Bachman & Palmer（1996）の理論的枠組みを用いて整理し、本論文で扱うビジネスコミュニケーション能力の構成要素として（1）言語能力、（2）非言語能力、（3）方略的能力の3つを同定した。

ただし、コミュニケーション能力は複雑な構成概念であるとの指摘があるように、本章で行った分析によって構成概念としてのビジネス日本語能力がすべて解明されたわけではない。しかし、外国人材を採用する日本企業の多くが外国人材のビジネスコミュニケーション能力を重視する背景においては、ビジネスコミュニケーション能力の構成要素を検討することは、外国人材のどのような能力を評価・測定することが必要であるかといった、評価・測定方法の検討にもつながるものであると考える。さらには、ビジネス日本語教育の現場においては、どのような能力を養成することが必要であるかといった、カリキュラ

ム・シラバスの決定、教授法、教材の決定の際の参考資料となりうるとも考える。今後は、実際にビジネスの現場で働く外国人ビジネスパーソンに対するコミュニケーション能力に焦点を当て、ビジネス場面でのパフォーマンスとそれに対する上司・同僚の評価はどのようなものか、外国人材を採用した企業の生の声を精緻に記述しながら、構成概念としてのビジネス日本語能力をより詳細に把握することを目指したい。

一方で、Bachman (1990 : 47) が、「言語能力を測定するための最初の段階は、測定したい構成概念を明確に、詳細に、そして多義的にならないように定義づけ、これを他の類似する構成概念から区別することである」と述べているように、大規模試験の開発・運用を考える際は、テストが測定すべき構成概念は明確に位置付ける必要がある。1.4 節の研究の目的でも、テストを開発する際のはじめのステップとして、テストで測定すべき構成概念を決定し、詳述する作業が必要であることを述べた。しかし、テストが測定すべき構成概念は、テスト開発過程で一度決定したら変更を加えないというものではない。本論文で扱うビジネス日本語能力は、その時代や経済情勢によって変化するものであり、ビジネス日本語テストで測定すべき構成概念も、それに合わせて適宜修正していく必要があるだろう。

本章で述べた構成概念としてのビジネス日本語能力について、次章では、BJT が測定する構成概念に触れ、さらに Cds による能力記述でその具体的な内容を示すことを試みる。

第4章 ビジネス日本語能力の評価・測定

第3章では、外国人材に求められる能力である、構成概念としてのビジネス日本語能力の一端を明らかにした。本章では、BJTをより精緻に分析し、BJTが測定するビジネス日本語能力の内容を、構成要素だけでなく、テストのスケールや要求される言語技能、部門別のタスクなども明らかにする。そして、BJTで得られた得点を解釈するための資料として、また外国人材のビジネス日本語能力を評価するための指標として、BJTが測定するビジネス日本語能力を参考にビジネス日本語 Cds を開発し、その対応関係についての検証の結果についても報告し、ビジネス日本語能力を評価するシステムとしての有効性を検証する。

4.1 研究目的と研究方法

4.1.1 研究課題の設定

4.1.1.1 先行研究と本章の意義

Cds とは、ある現実の行動が「できる」か「できない」かを、自己評価あるいは他者評価によって回答させるものである。アルク教育総合研究所（2015：134）によると、たとえば「飲食店などで、注文の間違いなどの問題が生じたとき、ある程度詳しく状況を説明し、苦情を言うことができる」「子どもが通っている学校からの行事の報告や学習内容などのお知らせを読んで、あまり苦勞せずに内容を理解することができる」などの Cds によって、「動詞の活用や漢字をどれだけ知っているか、ではなく、『何ができるか』が目標」になり、「学習者は次に何が必要かがわかり、目標設定がしやすく」となるという。

TOEIC や日本語能力試験など大規模試験の受験者に実施される Cds は、ある得点を得た受験者が、現実場面でどれだけの運用能力があると自己評価しているかを示すことができる。ETS（2008）は、1990 年代後半に、TOEIC の受験者の得点とその受験者を対象とした Cds 調査を行い、その結果をもとに TOEIC の得点と受験者が遂行できる言語行動を対応させたことを報告している。三枝（2004）は、自己評価としての日本語 Cds を開発し、その性能と有効性について、外在基準として日本語能力試験および教育機関ごとに行われているプレースメント・テストの結果との関係を明らかにしている。また、島田ほか（2006）は、日本語能力試験に関連して、Cds 調査を通して 1 級および 2 級合格者の遂行

可能な言語行動を明らかにした。

Cds は、また、教育や研修の事前と事後に学生や研修生が自己評価をし、指導担当者や上司が他者評価することによって教育や研修の目標設定を確認したり、教育や研修の効果を検証したりするためにも用いることが可能である。たとえば、国際交流基金は 2010 年 3 月に『JF 日本語教育スタンダード 2010』を発表し、これについて、日本語で何がどれだけできるかという熟達度がわかるだけでなく、コースデザイン、教材開発、試験作成などにも活用できるとしている。

ビジネス日本語 Cds の開発については、Cds を用いてビジネス日本語の評価基準を作成しようとした試みとして、AJALT ビジネス日本語評価基準作成プロジェクト（2008）がある。国際日本語普及協会（以下、AJALT という）は 80 年代末以降、ビジネス関係者に対する日本語研修において、その成果を客観的に評価して報告することが企業から求められるようになり、企業関係者にわかりやすい表現で書かれた評価基準の作成を開始したという。その目的は、「企業関係者にわかりやすい表現で職務遂行能力と直結した評価基準を作成することで、企業関係者と日本語教育の専門家が効率よく意思疎通できるようになること」、そして日本語学習者が「学習目標や学習成果をより現実的に把握することができ、学習目標を高めること」（p.37）であると述べている。AJALT はビジネス日本語の評価基準を作成するにあたって、まず「ビジネス日本語 Cds（ある職務についてその遂行に必要な行動群）」を作成している。その中から職務遂行のために日本語力と関連が深い部分に着目し、それに関する日本語サンプルを収集して分析し、そのサンプルと同レベルの日本語を運用できる日本語力を、CEFR を参照して記述している。この記述から、職務遂行に必要な日本語能力の判定として、7 種類のスキルを段階的に設定した「ビジネス日本語基本 Cds」を作成している（p.37）。

- (1) 口頭表現力：一人で一方的に話すための日本語力
- (2) 情報収集力：文字情報、音声情報を理解するための日本語力
- (3) 情報交換力：文字・音声で、事実確認中心の情報交換をするための日本語力
- (4) 合意形成力：文字・音声で、合意形成中心のやりとりをするための日本語力
- (5) 文書作成力：文書を作成するための日本語力
- (6) 関係構築力：対人関係を維持・構築するための日本語力
- (7) 通訳・翻訳力：通訳または翻訳をするための日本語力

この「ビジネス日本語基本 Cds」は、日本語教育の専門家が「ビジネス日本語 Cds」に記述された行動を遂行するための日本語力を分析するために使用するものだという。また、この「ビジネス日本語基本 Cds」とは別に、ビジネス日本語教材を資料として、日本語の語彙力・構文力の目安を作成している。職務遂行に要する語彙、表現、文法事項の目安として、「ビジネス共通語彙」「専門語彙」「一般語彙」「産出力 (Output)」「理解力 (Input)」に分けて記述している (表 4-1)。

表 4-1 職務遂行に要する語彙、表現、文法事項の目安 (AJALT, 2008 : 38)

網掛けはすべての職務遂行が可能なレベル

	ビジネス 共通語彙	専門語彙	一般語彙	Output	Input
A1		現場別 超高頻度語彙	基本語彙 約 500	単語 短文	サバイバル 情報
A2		現場別 高頻度語彙	日本語能力試 験 3 級語彙	高頻度基本 (初級) 文型	教材用 加工資料
B1	ビジネス 基本語彙	業界別業種別 高頻度語彙	(同) 2 級 機能語彙	基本文型	生資料 (事実)
B2	ビジネス 一般語彙		(同) 2 級 内容語彙	高頻度 表現文型	生資料 (解説)
C		低頻度 専門語彙			

AJALT (2008) の作成したビジネス日本語評価基準は、企業関係者との意思疎通ということを目的としており、わかりやすい表現での記述が試みられているところが興味深い。また、日本語レベルによって遂行可能な職務の目安のほかに、語彙や文法項目の基準についてもビジネス日本語教育のシラバスや学習項目の決定の際に有用であると考えられる。

しかしながら、AJALT (2008) のビジネス日本語評価基準作成においては、次の点で未だ課題があると見られる。それは、表 4-1 で示した「職務遂行に要する語彙、表現、文法事項の目安」の設定が、AJALT のこれまでの指導実績に基づく経験知から作成された

ことである。基礎資料となったものは、AJALT が作成した、『外国人のための専門別漢字導入と練習』および『Japanese for Professionals』（講談社インターナショナル）の 2 種類の教材であり、AJALT（2008：39）は、これについて、「今回、評価基準作成プロジェクトの基礎資料として使用することが妥当だと判断したのは、これらの教材によって多くの学習者が実際に日本語で実務担当が可能なレベルへ到達しているから」と説明しているが、「これらの教材の語彙群による語彙カバー率の確認を進め、必要な調整作業を進めていく」としており、今後のさらなる検証の結果が待たれる。また、表 4-1 の「一般語彙」の目安として日本語能力試験の級別の語彙を挙げていることから、現実のビジネス場面で必要となる語彙力・構文力を反映しているかどうか疑問が残る。

ほかにビジネス日本語 Cds の開発に関する研究として、堀井ほか（2013）は、大学院ビジネス日本語コースにおける目標達成度の測定を目的としたビジネス日本語 Cds の開発を試みており、ある特定のコース評価に関しては有効性が示唆されたとしているが、ビジネス日本語教育での汎用性を得るためには更なる検討が必要であると述べている。また、BJT では、第 8～15 回の受験者、合計 14,634 名を対象に、話し言葉、書き言葉およびその複合で構成される 40 項目のアンケート調査を実施し、その結果を「Can do レポート」として公表している。しかし、その調査項目は、「簡単な自己紹介ができる」「物の値段や時間など数字が聴き取れる」など一般的な生活日本語のレベルにとどまっており、ビジネス日本語に特化した Cds とはなっていない。このように、ビジネス日本語 Cds についての検証は未だ十分とは言えない状況である。

1.2 節で述べたように、日本企業における外国人材の活用が進まない要因のひとつとして採用時の外国人材の能力の判定が難しいことが指摘されている。そこで本章では、ビジネスコミュニケーション能力を評価するシステムを構築することを目指し、ビジネス日本語 Cds を開発する。本章で開発するビジネス日本語 Cds は、BJT が測定するビジネス日本語能力を参考にして項目を作成する。このビジネス日本語 Cds によって、BJT が具体的にどのような能力を測定しているか、あるいは、BJT である得点を得た受験者が、具体的なビジネスコミュニケーション活動において日本語を使用して具体的にどのようなことができるのか、どのような点が不足しているのかを知ることができる。

ビジネス日本語 Cds はまた、BJT で得られた得点を解釈するための資料となるだけでなく、外国人材のビジネス日本語能力を評価するための指標として、具体的な内容を示すこともできる。そのため、第 3 章で論じたビジネスコミュニケーション能力全般を広く評価

できるよう、BJT が直接測定していない、「話す」能力や「書く」能力に関する項目も作成し、日本語を使ってビジネス活動を行う上での運用能力も評価できるような指標の作成を目指した。

島田 (2010:269-270) は、「Cds による自己評価の結果が大規模試験やプレースメント・テストなどの客観テストの結果を反映しているかどうかということを検討する研究は、Cds の妥当性を検討するものと言える」としているが、本章では BJT が測定するビジネス日本語能力の内容を参考にビジネス日本語 Cds の項目を構築し、各項目が実際に BJT の結果を反映しているかどうかを検証し、BJT が測定するビジネス日本語能力を反映する Cds としての有効性を検討する。繰り返しになるが、BJT に基づくビジネス日本語 Cds が果たす役割は、BJT の得点からのみでは推し量れないことを明らかにすることである。つまり、BJT 受験者がビジネス日本語コミュニケーション活動において、日本語を使用して具体的にどのようなことができるのか、あるいはどのような点が不足しているのかを知ることができる点であると考えられる。さらには、BJT が測定する能力が、具体的にどのようなビジネス場面で、どのような日本語を使用する際に必要とされるものなのかということも明らかにすることもできる。

4.1.1.2 研究目的

本章において開発するビジネス日本語 Cds は、外国人材のビジネス日本語能力を評価するシステムになりうると同時に、ビジネス日本語教育の現場において、教員や学習者にとって明確な目標レベルが提示できるなどの利点があると考えられる。本章において開発するビジネス日本語 Cds は以下のような点に資することを目標としている。

- 1) ビジネス日本語教育に携わる教員がシラバスを設定する際、教育内容や到達目標を決定する指針となる。
- 2) ビジネス日本語教育を受ける学習者に学習項目や到達目標を意識化させる。
- 3) ビジネス日本語教育の教育過程において教員は教育効果が確認でき、さらに指導すべき学習項目の設定に役立つ。
- 4) ビジネス日本語教育の学習過程において学習者は学習効果が確認でき、さらに習得すべき学習項目が把握できる。

本章は、以上のような観点から開発したビジネス日本語 Cds について、予備調査によってその有効性の検証を行い、修正を加えた後、再検証を行った結果について報告することを目的とする。具体的には、以下の 2 つの課題を設定し、研究を行った。

- (1) BJT が測定するビジネス日本語能力とは何か。
- (2) ビジネス日本語テストが測定する能力を反映する Cds をどのように開発するか。

4.1.2 研究方法

4.1.2.1 BJT が測定する能力の分析

BJT が測定するビジネス日本語能力を明らかにするために、まず、JETRO の資料から、BJT が測定する（ことを目指している）ビジネス日本語能力の領域・技能、BJT のスケール、BJT の設問で要求される課題と情報処理の内容について整理する。そして、ビジネス日本語 Cds の項目の構築のため、BJT のテスト項目を用い、各出題形式においてどのようなビジネス日本語能力が測定されているかを、部門やセクションごとに分析し、BJT が測定するビジネス日本語能力の構成要素を記述する。

4.1.2.2 ビジネス日本語 Cds の予備調査の概要

BJT のそれぞれの部門・セクションにおいて求められる課題（タスク）を分析した結果から、ビジネス日本語 Cds 調査票の質問項目として再構築した。「聴く」技能 10 項目、「読む」技能 10 項目の計 20 項目を作成し、予備調査を行った。回答は各項目について、1（ほとんどできない 0%）から 5（問題なくできる 100%）までの 5 段階評定尺度の該当する項目を丸で囲むようになっている。項目はすべて日本語で書かれており、漢字にはルビが振られているものを使用した（巻末の資料 1 を参照）。

本章は、就労前の外国人留学生のビジネス日本語能力を評価する基準としての Cds 構築を目指していることから、日本国内の大学または日本語学校でビジネス日本語を学習している学習者計 69 名に対し、BJT 模擬テスト⁸を実施し、次いでビジネス日本語 Cds 調査を行った。

⁸ 加藤(2010)をBJT模擬テストとして使用した。これは、正規のBJTと同じ項目数、出題形式、難易度で構成されており、その得点は簡易な換算表を用い、正規のBJTの等化得点に限りなく近い値を80%の確率で推計するものである。

4.2 結果

4.2.1 BJT が測定するビジネス日本語能力の領域・技能

JETRO (2000) によると、JETRO は、ビジネス日本語テストの開発に際し、ビジネスマンや外国人で構成される「ジェトロテスト検討委員会（以下、検討委員会とする）」において、現実の日本人ビジネスパーソンの言語行動にできるだけ近い内容のテストであるべきとの方針を立てたという。また、海外の日系企業へのヒアリング調査の結果から、企業は外国人ビジネスパーソンに対し、必ずしも正しく高度な日本語を書いたり話したりすることのみを求めているわけではなく、日本語を道具として使用し、社内や社外と円滑なコミュニケーションを図れることを望んでいるということが明らかになった。

検討委員会では、ビジネス場面において、円滑なコミュニケーションを行うためには、ことば、地域・国籍、人種、文化、社会規範、商習慣などの異なる相手を対等の立場で相互に尊重しあうことが重要な鍵であり、お互いを理解し、ビジネス上の課題を相互に調整し、当初の目的を達成することが必要であり、また、日本語でビジネス活動を行う外国人ビジネスパーソンにとっては、日本語の知識の量や正確さよりも、さまざまな局面で日本語を使って意思疎通ができ、そして、結果として当初の目的のビジネス活動が達成できるということが重要だとの指摘がなされた。

そこで、BJT では、ビジネス場面におけるコミュニケーション能力をその主要な測定対象として位置づけ、ビジネスに関する知識の領域ではなく、その知識を使う能力、つまり技能の領域に比重を置いて開発された。図 4-1 は、加藤 (2010) がビジネスコミュニケーション能力の領域のうち BJT が測定している能力領域（白抜き台形部分）を概念的に表したものである。四象限マトリクスによって示されているが、加藤 (2010) の説明によると、図中の対称に位置するものが二項対立関係であることを意味しているというのではなく、この図が説明しているのは、あくまで、楕円の右半分に示された「ビジネス知識」「専門的知識」が前提知識として位置づけられ、左半分に示された「コミュニケーション・スキル」「ビジネス・スキル」がその主要な測定対象として位置づけられており、BJT が測ろうとしているものは、「知識」よりも「技能」、「専門的」な事項より「実用的」な事項であることだという。また、白抜き部分によって、BJT が測れるものは、「たかだか、日本語によるビジネス活動全体の中のほんの一部」(p.174) であることを示しているに過ぎない。

加藤(2010)が、BJT で測定される技能として示しているのが図 4-2 である。加藤(2010)

の説明によると、BJTは「聞く・読む」を中心とした言語技能、さらにその下位技能である情報の抽出、総合、推測などの技能があるかどうかを測るとしている。図4-1と同様に、図4-2においても、日本語に関する知識、ビジネスや社会に関する知識は前提知識として示されており、開発時においてBJTは、日本語の運用能力を測定することを目指していたことがわかる。

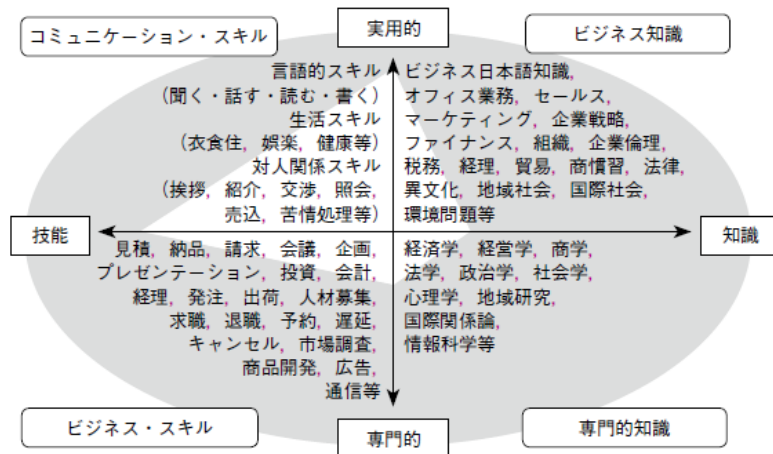


図4-1 BJTが測定している能力の領域（加藤，2010：174をもとに作成）

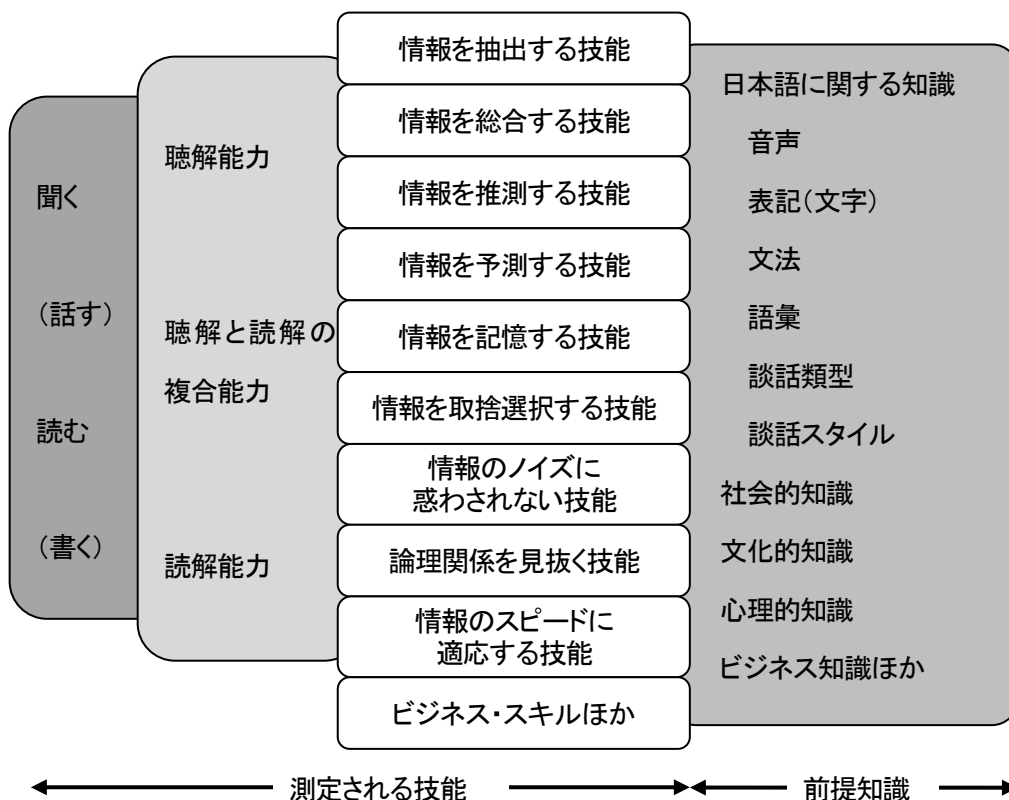


図4-2 BJTで測定される技能（加藤，2010：179をもとに作成）

4.2.1.1 BJT のスケール

加藤（2010）によると、人間の記憶はスキーマ（schema）と呼ばれる高次の構造からなっており、それぞれのスキーマは特定の事物や事象に関する知識を1つのカードのようなものに収めており、スキーマを発動させながら、人は情報を理解し、処理していると考えられるとし、スキーマはある種の階層構造をもって発動されることが多いと考えられ、この階層の深さが深ければ深いほど情報の処理は難しいということが言えるとしている。そして、BJTでは、ある情報を処理するためにスキーマというカードをたくさん使えば使うほど難易度の高いテスト項目になるとしている。

たとえば、打ち合わせを何時から始めるかという連絡の社内メールから、打ち合わせの開始時間の情報を読みとるという単純な情報処理よりも、複数の指示の書かれた上司からのメールを読んで、最も優先して処理しなければならないことは何かを判断するというほうが、高度な情報処理の遂行を要求される。こうした考えから、BJTのテスト項目の難しさを決定するものは、言語知識（漢字、語彙等）やビジネスの知識の難しさではなく、そうした知識を前提に、わからないことがあっても推測したり取捨選択する処理を行ったりしながら、日常的なビジネス場面における課題を達成することの難しさであると考え、それがテストの能力レベルを判定するスケールに反映されているととらえている。

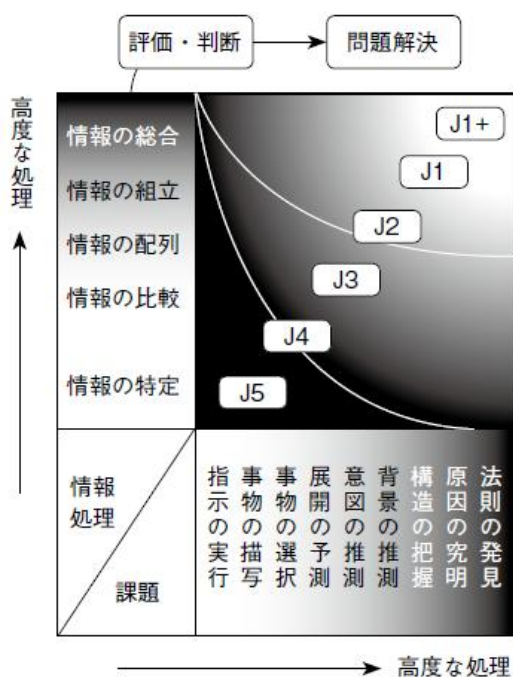


図 4-3 BJT の設問で要求される課題と情報処理（加藤，2010：192 を元に作成）

図 4-3 は、BJT の設問で要求される課題と情報処理の内容を概念図で示したものである（図中の J1+から J5 については、2.1.3.3 節を参照）。この概念図においては、横軸が課題を表し、右に行くほど高度で難しい課題であるとしている。また、縦軸は情報の処理を表し、上にいくほど高度で難しい処理であるとしている。ビジネスパーソンは、ビジネスコミュニケーションの過程において、横軸の課題に対して縦軸の情報処理を行いながら、評価や判断を加え、問題解決行動を行っていると考えられる。

4.2.1.2 BJT の設問で要求される課題と情報処理の内容

BJT の設問で要求される課題と情報処理の内容について、加藤（2010）によると、図 4-3 に示されている「情報処理」と「課題」の具体的な設定としては次のような例が挙げられ、BJT ではこのような能力を測定するテスト項目を出題しているという。ただし、以下に示すものは、単独で必要となる場合も複合的に必要となる場合もある。また、ある能力領域と他の領域とが明確に区別されているわけではなく、多くの能力が他の能力領域と一部重なり合っている。

（１）情報の特定

- ・単一の情報を特定し、それに反応する。

例：「お声が遠いようですが」という発話を聞いて、声を大きくして話す。

- ・単一の情報（指示）を理解し、実行する。

例：書類の訂正を指示され、訂正箇所を特定し、訂正する。

- ・単一の情報を特定し、その是非を判断する。

例：「検討します」という発話を聞いて、それが否定的見解だと理解する。

- ・複数の情報を特定し、その是非を判断する。

例：企業買収をもちかけられ、相手の信用、財務状況と照らし合わせ、その是非を判断する。

- ・複数の情報（指示）を理解し、実行する。

例：ある物（形状、色、大きさなど）を探すように言われて、それを探す。

（２）情報の比較

- ・複数の情報を比較し、単純な判断をする。

例：複数の商品の説明を聞いて、どれを購入するか判断する。

- ・複数の情報を比較し、複雑な判断をする。

例：複数の商品の説明を聞いて、予算ないし財政状況を考慮し、どれを購入するかを決める。

- ・複数の情報を比較し、事実の照合を行う。

例：複数のメーカーに電話して、ある機械の部品と別の部品とが代替可能かどうか調べる。

- ・複数の情報を比較し、背景をさぐり、問題を究明する。

例：製品の売り込みがうまく行かない原因を種々のソースから突き止める。

(3) 情報の配列

- ・複数の情報を把握、時系列的な配列を施し、指示を実行する。

例：上司のスケジュールを聞いて、部内の週間予定表を作成する。

- ・複数の情報を把握、空間的な配列を施し、指示を実行する。

例：訪問先への道順、オフィスの模様替え配置図などを作成する。

- ・物事の手順・手続きに関する情報を把握・配列し、指示を実行する。

例：切符の買い方、機械の使い方、仕事の手順、通関手続きの仕方、見積もり・納品・請求・領収書の書き方を調べて、指示を実行する。

- ・複数の情報や基準に従い、情報を配列し、展開を予測する。

例：住所録（氏名・住所・所属など）、株式相場、外国為替レート表、製品の販売実績などを作成し、そこにある規則性を発見し、そのあとに起こる事態の展開を予測する。

(4) 情報の組立

- ・複数の情報を組み立てて、指示を実行する。

例：社内会議の議事録の内容を項目別にまとめる。

- ・複数の情報を組み立てて、事物を選択する。

例：人事採用、プロジェクトの人材抜擢など、人物・事物の特徴、性質、背景に関する情報を組み立てて、特定の条件に合う人物・事物を探し出す。

- ・複数の情報を組み立てて、原因を究明する。

例：製品の不具合に対するクレームから、その原因を複数のソースから突き止める。

(5) 情報の総合

- ・複数の情報を総合して，指示を実行する。

例：プロジェクトの完成記念式典をいつどこでやるか，アンケートをとって企画立案を行う。

- ・複数の情報を総合して，展開を予測する。

例：得意先に納入した商品に関するクレームについて，問題点を洗い出し，その後の処置をどう行うべきか，対策を立てる。

- ・複数の情報を総合して，法則性を発見する。

例：ある企業の株式相場のこれまでの推移を図示し，世界の政治動向や外国為替相場，原油価格の推移と照合しながら，その企業の株式相場の変動にどのような規則性があるかをさぐる。

- ・複数の情報を総合して，原因を究明する。

例：ある販売システムに問題があり，いろいろなソースからその原因をさぐり，ある評価・判断を下して，どのような解決策が良いか提案する。

4.2.1.3 BJT の部門別のタスク

BJT の出題形式については，2.1.3.2 節で述べたとおり，3 部門 8 セクションで構成されている（表 2-3 再掲）。

表 2-3 BJT の構成（再掲）

部門	セクション	項目数	時間
聴解	(1) 描写問題	10	約 50 分
	(2) 表現力問題	10	
	(3) 聴解問題	15	
聴読解	(1) 描写問題	15	約 30 分
	(2) 聴読解問題	15	
読解	(1) 語彙問題	10	約 40 分
	(2) 表現力問題	10	

	(3) 短文読解問題	15	
合計		100	約 120 分

ビジネスの分野は多岐にわたっており、現実のビジネスでは、法律、財政、税務、企業、会計、貿易、保険、金融、財務、投資、融資、外国為替、信託、不動産、証券など、それぞれ専門的なビジネス知識が要求されるが、BJT では、そうしたビジネス知識について、解説をともなって出題される場合はあるものの、知識の有無を直接問うようなテスト項目は出題されていない。また、BJT では、テストの場面や設定が、現実のビジネス場面を反映しているかどうかという真正性（authenticity）を重視しており、受験者が課題達成のための視点をフォーカスしにくい設定の出題はされていない。

黄（2016）の分析によると、BJT の聴解および聴読解部門のテスト項目で設定されている場面は、上司指示、社内打診、発表提示、社外打診、表敬挨拶、会社訪問、部下報告、面談面接、社内雑談、電話対応、対外連絡、法則発見、構造把握、物事選択の 14 種類に分類されるという。特に、聴解部門では上司指示と社内打診が、聴読解部門では社内打診、構造把握が高い頻度で出題されているとしている。

ここから、各出題形式においてどのようなビジネス日本語能力が測定されているか、また、受験者に求められるタスクを分析した結果をそれぞれの部門・セクションごとに記述する。

4.2.1.3.1 第 1 部 聴解問題

（1）セクション 1 描写問題

音声の指示にしたがい、問題冊子にある写真を見て、それがどのような場面であるかを判断し、音声で提示される選択肢から正答を選ぶ問題群である。このセクションでは、ビジネス場面で行われるさまざまな活動について、それを描写する音声表現を理解できるかどうかを測る。

（2）セクション 2 表現力問題

問題冊子にある写真を見ながら、その場面ではどのように言うことがふさわしいかという音声の設問に対し、音声で提示される選択肢の中から正答を選ぶ問題群である。このセクションでは、ビジネス上のある状況において、適切な言語表現を用いて言いたいことを

伝えられるかどうか問われる。伝える相手は上司や顧客などさまざまである。

(3) セクション3 聴解問題

問題冊子にある絵（イラスト）を見ながら、音声の設問を聞いたあと、短い会話を聞いて、音声で出された設問に対し、音声で提示される選択肢の中から正答を選ぶ問題群である。このセクションでは、セクション1，2と異なり、ダイアログやモノログの会話を聞き、その会話がどのような状況で行われているかを把握することがまず問われる。状況を把握した上で、さらに設問で要求されている課題を処理・遂行することが求められる。このとき、受験者は、会話をだれの立場で聞かかということを判断しなければならない。たとえば、モノログの場合は、セミナー参加者、商品説明会の出席者などであり、ダイアログの場合は、会話に参加している人物のうちどちらか一方、あるいは会議での議論の場面では会議参加者という立場で会話を聞くことになる。特にダイアログの場合は、会話をしている人物同士と、その会話に参加して、あるいは第三者の立場で会話を聞いている自身との対人関係（上下・公私・親疎・内外）についても把握する必要がある。

4.2.1.3.2 第2部 聴読解問題

(1) セクション1 描写問題

音声の指示に従い、問題冊子にある写真や図表を見てそれがどのような場面であるか、あるいはどのような文書情報であるかを判断し、文字で提示された選択肢の中から正答を選ぶ問題群である。現実のビジネス場面では、ほかの人の会話に途中で参加したときに瞬時に話題を把握したり、また、メモや図表を一読して何について書かれているかを瞬時に把握したりすることが求められることがしばしばある。このセクションでは、ごく短い時間で音声情報と読解情報の両方を処理する能力が問われる。

(2) セクション2 聴読解問題

問題冊子にある絵（イラスト）や図表を見ながら音声の設問を聞き、そのあと、短い会話や説明文を聞いて、設問で出された課題に対して、文書情報または図中に提示される選択肢の中から正答を選ぶ問題群である。日常のビジネス活動の中には、会議資料やスライド画面などの視覚情報を見ながら会話を聞く場面が多い。このセクションでは、ダイアログやモノログを聞きながら、文書、図表、パンフレット等の視覚情報を読み、設問で

要求される課題を処理・遂行する能力が問われる。聴解セクション3と同様に、受験者はこのとき、会話をだれの立場で聞くかということを判断しなければならない。

4.2.1.3.3 第3部 読解問題

(1) セクション1 文法・語彙問題

文書やメール等のビジネス場面での使用が想定される短い文章を読んで、空欄にあてはまるもっとも適切な語彙あるいは文法項目を、文字で提示された選択肢の中から選ぶ問題群である。このセクションでは、文章全体の意味やニュアンスを把握しつつ、適切な文法項目・語彙を選択できるかどうかが問われ、日本語の文法や語彙の知識を運用して文章を読む能力を測定する。

(2) セクション2 表現力問題

文書やメール等のビジネス場面での使用が想定される短い文章を読んで、空欄にあてはまる最も適切な表現を、文字で提示された選択肢の中から選ぶ問題群である。このセクションでは、文章全体の意味やニュアンスを把握しつつ、適切な表現を選択できるかどうかを問われ、日本語の言語表現の知識を運用して日本語の文章を読む能力を測定する。

(3) セクション3 読解問題

ビジネス場面でよく使われる掲示、図表、文書、メール等を読み、設問に対し、文字で提示された選択肢の中から正答を選ぶ問題群である。現実のビジネス場面では、社内で閲覧されるお知らせや掲示物、出張やイベント等の報告書、また、社外から送られてくる依頼文書やメール、さらにはプレゼンテーション資料、商品カタログ、企業のホームページ等々、実に多種多様なものを読んでいる。その際は当然、ただ文字を目で追うだけでなく、内容を把握し、情報を整理し、さらに、自身は何をするべきか、あるいは相手の希望や要求は何であるかを判断しなければならない。このセクションでは、さまざまなビジネス関連文書を読んで、設問で要求される課題を処理・遂行する能力があるかどうかを測定する。

4.2.1.4 BJT が測定する構成概念

ここまで、BJT の出題項目を分析し、どのようなビジネス日本語能力が測定されているか、また、受験者にどのようなタスクの処理が求められているかを明らかにした。本節の

はじめに、BJT が測定するビジネス日本語の領域・技能を、BJT の開発資料から図 4-1 および 4-2 をもって示したが、BJT の実施が開始され約 20 年を経たため、ここでは、近年実際に出題されたテスト項目の分析結果から、BJT の測定する構成概念を、第 3 章で示したビジネスコミュニケーション能力の構成要素の枠組みの中で、改めて考えてみることにしたい。

第 3 章で、本論文で扱うビジネスコミュニケーション能力の構成要素として (1) 言語能力、(2) 非言語能力、(3) 方略的能力の 3 つを同定したが、BJT が測定する構成概念は、(1) 言語能力を主として、以下の点で重なっていると言える (3.3 節を参照)。

(1) 言語能力

- ・ビジネス場面で使用される語彙や表現・文法に関する知識
- ・文書や対話から、情報を読み取ったり聞き取ったりする能力
- ・目的、状況、相手との関係、話題などを考慮しながら、適切に日本語を使い分ける能力
- ・ビジネス課題に直面したときにその状況や相手に配慮しながら、場面に適した行動、自分の取るべき行動を選択できる能力

(2) 非言語能力

- ・日本の企業における文化的背景を理解し、調整する能力

つまり、(2) 非言語能力の一部（「相手の表情やジェスチャーを読み取る能力」、「能動的かつ率直にビジネス活動に取り組む姿勢」）あるいは(3) 方略的能力（「コミュニケーション・ストラテジーを用いてコミュニケーション上の問題点を解決する能力」）を BJT で直接測定することはできない。すなわち、BJT は、第 3 章で先行研究の定義や日本語学習者、企業関係者に対する調査から探った、社会的要請としてのビジネス日本語能力の構成要素すべてを測ることのできるテストではない。また当然ながら、「話す」「書く」といった産出技能や、3.3 節で示した、異文化理解能力やリーダーシップ、主体性などといったグローバル人材一般に必要な能力を測ることもできない。この点は加藤 (2010) が「日本語によるビジネス活動全体の中のほんの一部」と注釈した通りである。ただし、野口ほか (2014 : 43) は、「言語テストの構成概念は、一般に、目標言語使用領域に基づいて、言語の形式に関わるさまざまな言語能力や知識を測定されるように定義されるが、言語形式

の持つ語用論的、社会言語学的な知識は、当該言語が使用されている社会文化に関する知識と切り離せないものである。また、言語使用課題によっては、言語が用いられる背景となる百科事典的知識によって、言語の理解が補われることもある。そのため、構成概念では定義されていなくても、言語そのものの知識以外の知識が、テスト結果に影響を及ぼす可能性は否定できない」と述べており、実際にはより広い範囲を測定している可能性も考えられるのである。

しかしながら、紙と鉛筆を用いた（2016年現在）言語テストであるBJTでは直接測定することのできないビジネス日本語能力の構成要素を評価・測定するためには、その構成要素に焦点を当てたCdsやパフォーマンステストなど、別のツールを用いる必要があるのではないだろうか。そこでその一端として、次節では、日本語を使ってビジネス活動を行う上での運用能力も評価できるようなシステム、指標としてのビジネス日本語Cdsの開発について論じる。

4.2.2 ビジネス日本語Cdsの開発と検証

4.2.2.1 予備調査の結果

ビジネス日本語Cdsの予備調査は、2009年9月～2010年2月に実施した。分析の結果、当該予備調査で実施したビジネス日本語Cdsについて、以下の2点が明らかになった。

- 1) 信頼性係数（ α 係数）は0.864と高く、内部一貫性が認められた。

しかし、

- 2) ビジネス日本語Cdsの合計（全20項目×各5段階評価＝最大値100）とBJT模擬テストの尺度得点（0～800点）との相関係数は0.151と低く、BJT模擬テストの結果を反映する尺度とは言えないということが示された。

そのため、「BJTに基づくビジネス日本語Cds」を構築するためには、4.2.1.4節で示したBJTの構成概念が適切に反映されているかどうかについて、各項目の内容を精査し、その結果に応じて項目に修正を加え、より多くの被験者に更なる調査を重ねる必要性が指摘された。

4.2.2.2 ビジネス日本語Cds項目の修正・削除

予備調査の結果を基に、修正・削除を必要とする項目を特定した。修正対象項目の特定および修正は以下のようなプロセスを経て行われた。

- 1) 当該ビジネス日本語 Cds の全項目について、内容や表現に関し think aloud 方式でヒアリングを行い、疑問点、修正についての提案や意見を収集し、修正点として反映させた。ヒアリングの対象者は日本語非母語話者 10 名で、国籍は、中国、台湾、韓国、ドイツ、フランス、モンゴル、ブラジル、アルゼンチン、タイである。ヒアリングは、調査者のいる部屋に母語別に対象者が入り、修正前のビジネス日本語 Cds の項目を一つずつ確認し、気づいたこと、疑問点、思ったことをその場で声に出して説明するという方法で行われた。その際の音声はすべて録音し、後に文字化して資料とした。
- 2) 1) の資料に基づいて項目を修正したビジネス日本語 Cds を作成した。また、4.1.1.1 節で述べたように、ビジネスコミュニケーション能力全般を広く評価できるように、BJT の聴解と聴読解部門で出題されているビジネス場面の分析から「話す」10 項目を作成し、聴読解と読解で出題されているビジネス文書の分析から「書く」10 項目を新たに作成し、ビジネス日本語 Cds 調査票に追加した。
- 3) 2) を 11 ヶ国語に翻訳した。翻訳した言語は、中国語簡体字、中国語繁体字、韓国語、ドイツ語、フランス語、スペイン語、モンゴル語、ポルトガル語、ロシア語、タイ語、ベトナム語である。翻訳にあたっては各言語につき 2 人の母語話者が担当し、1 人が日本語を母語に訳し、次いでもう一人が訳されたものを再度日本語に翻訳し直し、最終的には面談して内容を確認して修正を加えるという手順を踏んだ。

三枝（2003）は、開発した日本語 Cds について、Cds が日本語能力の評定尺度として有効に機能するかを検討するため、日本語能力試験受験者の試験結果および A 大学のプレースメント・テスト受験者のテスト結果を Cds 評価の結果と比較している。そして、Cds の四技能を比較すると、「読む」と「聴く」の受容技能の方が、「書く」と「話す」の産出技能よりも客観的に評価しやすいためか、能力差を反映する結果となっていると分析している。一方で、Cds 調査において能力差を反映しない要素として以下の 4 点を挙げている（p.105）。

- 1) 日本語以外の能力、知識に関する項目
- 2) 具体的にどのような場面でどのような言語行動を指しているかが、回答者によって判断が異なる可能性のある項目
- 3) 授業（特に初級レベル）で学習経験のあるタスクに関する項目

4) 異文化間コミュニケーションにおける情意面に関係し、能力が高くなるほど難しさを
実感する項目

本章で構築するビジネス日本語 Cds において、三枝（2003）の 1) に該当すると思われる項目については削除し、BJT で測定しているスキルにより近い項目に置き換えることとした。また、2) に該当すると思われる項目については、ヒアリング中、回答者によって判断が異なる可能性があると思われた項目の記述をより明確になるように修正した。

以下に、ビジネス日本語 Cds の修正や削除を必要とした項目について、具体的にどのような変更を加えたか、技能別に記述する。

（1）「聴く」の項目の修正・削除

「聴く」の項目について、修正を加えたのは表 4-2 の 5 項目である。「聴く」の修正項目を見ると、項目 1, 2, 4, 6 については、回答者によってタスクのとらえ方が異なっていたため、より具体的でわかりやすい記述に変更した。項目 7 については、三枝（2003）が挙げている「1)日本語以外の能力、知識に関する項目」に該当すると考えられたため削除し、BJT で測定しているスキルにより近い項目を新たに作成して置き換えた。

変更を加えた部分は網掛けで示した。「修正前」は予備調査で実施した際の Cds 項目の記述文であり、「修正後」は再検証で実施した Cds 項目の記述文である。「ヒアリング結果」は think a loud 調査の際の発言の抜粋である。

表 4-2 「聴く」の項目の修正

項目		修正の内容（網掛け部分は変更箇所）
1	修正前	上司からだされた業務に関する指示が理解できる。
	ヒアリング結果	「業務」の定義は何か。自分自身の担当であるということがわかりにくい。
	修正後	上司からだされた自分の担当業務に関する指示が理解できる。
2	修正前	社内で同僚同士が話している会話を理解し、トピックが特定できる。
	ヒアリング結果	会話に加わらず、傍で聴いているということがわかりにくい。
	修正後	社内で同僚同士が話している会話を近くで聴いて、どんな話題について話しているかわかる。

4	修正前	自分の担当分野に関する上司や先輩からのアドバイスやコメントの内容が理解できる。
	ヒアリング結果	アドバイスとコメントの違いは何か。「理解できる」とはどのレベルを言うのか。
	修正後	自分の担当分野に関する上司や先輩からのアドバイス（助言）やコメント（感想）の内容が理解できて、仕事に生かせる。
6	修正前	会議において、議論の内容を聴き、主張の異なる点が理解できる。
	ヒアリング結果	主張が異なる点、という部分がはっきりしない。
	修正後	会議のときに、議論の内容を聴きながら、賛成や反対など主張の違いがわかる。
7	修正前	商品や製品の使い方や仕様など、技術的な説明が理解できる。
	ヒアリング結果	仕様ということばになじみがない。やったことがない。
	修正後	留守番電話のメッセージを聴いて、相手の要求や説明が理解できる。

（２）「読む」の項目の修正・削除

「読む」の項目について、修正を加えたのは表 4-3 の 5 項目である。主な修正点は、「業務メモ」「特定する」「洗い出す」「推し量る」など、ヒアリング対象者が「わかりにくい」とした用語をよりわかりやすくしたことと、「短時間」という回答者によって判断が分かれた部分をより具体的にしたことなどである。「短時間」を「2, 3 分」としたのは、BJT の聴読解問題セクションにおいてビジネス文書を 2 分強で特定する設問があることからである。また項目 6 については、「聴く」の 7 と同様の理由で項目を新たに作成して置き換えた。

表 4-3 「読む」の項目の修正

項目		修正の内容（網掛け部分は変更箇所）
4	修正前	社内に出されたお知らせや回覧文書、業務メモなどが短時間で理解できる。
	ヒアリング結果	短時間とはどのくらいかわからない。回覧文書になじみがない。業務メモがどういうものかイメージできない。

	修正後	社内に出されたお知らせや、 回覧されている文書 、 仕事の指示を書いたメモの内容が2, 3分読んで理解できる。
5	修正前	社外から来たビジネス文書の種類が短時間で特定できる。
	ヒアリング結果	短時間とはどのくらいか明記する必要あり。「特定できる」がわかりにくい。
	修正後	社外から来たビジネス文書を 2, 3分で読んで、内容によって分類できる。
6	修正前	社内に導入された新しい機械の取扱い説明書が理解できる。
	ヒアリング結果	日本語の知識以外の部分を測っている。日本語能力と関係がない。
	修正後	社内 で提案された新しい企画についての稟議書を読んで、企画の概要や提案理由がわかる。
7	修正前	一連のビジネス文書を比較して読み、異なる点を洗い出せる。
	ヒアリング結果	「一連」「洗い出す」がわかりにくい。タスクがイメージできない。
	修正後	社外の人と数回やり取りした文書やメールを読んで、 新しく決まったこと、変更があったことは何かがわかる。
9	修正前	婉曲的な表現で書かれた社外からのビジネス文書の意図を推し量ることができる。
	ヒアリング結果	婉曲的、推し量るがわかりにくい。(意味を確認し翻訳に反映)
	修正後	丁寧な印象を与えるために婉曲的な表現で書かれた社外から来たビジネス文書の意図がわかる。

(3)「話す」の項目の修正

「話す」の項目について、修正を加えたのは表 4-4 の 5 項目である。そのうち項目 10 以外の 4 項目は、三枝 (2003) が挙げている「(4) 異文化間コミュニケーションにおける情意面に関係し、能力が高くなるほど難しさを実感する項目」に該当すると考えられたため、より詳細な状況を設定するなどの修正を加えた。

表 4-4 「話す」の項目の修正

項目		修正の内容（網掛け部分は変更箇所）
1	修正前	社内と社外で話している相手との関係を考慮し、適切な待遇表現を使うことができる。
	ヒアリング結果	社内、社外をはっきり説明。待遇表現という言葉がわかりにくい。
	修正後	社内と社外で話している相手との関係を考慮して場合に 応じた適切な話し方を選ぶことができる。
4	修正前	上司に対して疑問点や問題点を婉曲的かつ正確に説明することができる。
	ヒアリング結果	「かつ」という表現は不適。婉曲的がわかるかどうか。もう少し具体的に記述。
	修正後	上司に対して疑問点や問題点を感じたときに、婉曲的な言い方だが、 正確に自分の思っていることを説明できる。
7	修正前	取引先の人のお話がわからないときに、気分を害さずに対応することができる。
	ヒアリング結果	気分を害さず、がわかりにくい。状況説明を詳しくする。
	修正後	取引先の人と話しているときに、よく理解できないことがあれば、 相手の気持ちを考えながら、丁寧に聴き返すことができる。
8	修正前	社外からの依頼や苦情を理解して、相手の気分を害さずに対応することができる。
	ヒアリング結果	気分を害さず、がわかりにくい。状況説明を詳しくする。
	修正後	社外からの依頼や苦情の内容を理解して、相手の気持ちを考えながら、 きちんと対応できる。
10	修正前	自分の担当している商品や企画について、簡潔かつ効果的にプレゼンテーションできる。
	ヒアリング結果	「簡潔かつ効果的に」の「簡潔に」という表現がわかりにくい。
	修正後	自分の担当している商品や企画について、長くなりすぎないように 簡潔に、効果的にプレゼンテーションできる。

(4)「書く」の項目の修正

「書く」の項目について、修正を加えたのは表 4-5 の 6 項目である。項目 1 は「読む」の項目 4、5 と同様に「手早く」という用語が具体的にどのようなことを表しているかわかりにくいという意見が多かったため、修正を加えた。項目 5 に関しては、「議事録」という用語は各国語で翻訳する際に支障があるという意見があったため「議事録」を意味する言葉に置き換えた。

表 4-5 「書く」の項目の修正

項目		修正の内容（網掛け部分は変更箇所）
1	修正前	業務依頼や業務の内容について手早くメールやメモを書くことができる。
	ヒアリング結果	手早くというのは具体的にどのくらいの時間か。
	修正後	同僚や上司に業務を依頼したり、業務の内容を伝えたいときに、あまり時間をかけないで、さっとメールやメモを書くことができる。
3	修正前	苦情や問い合わせのメールや文書に相手の気持ちを考慮して返事を書くことができる。
	ヒアリング結果	「相手の気持ちを考慮して」という表現がわかりにくい。
	修正後	苦情や問い合わせのメールや文書が来たときに、相手の気持ちを考えながら、丁寧な返事を書くことができる。
5	修正前	会議に出席して正式な議事録を作成することができる。
	ヒアリング結果	議事録ということばになじみがない。
	修正後	会議に出席して、会議の記録をとり、あとで報告書を作ることができる。
7	修正前	競合他社に関する情報を収集し、自社と比較した報告書を書くことができる。
	ヒアリング結果	競合他社ということばについて説明が必要。
	修正後	自社の競争相手の競合他社に関する情報を集め、自社と比較した報告書を書くことができる。

8	修正前	社外からの依頼文書に相手の気持ちを考慮しつつ断りの手紙やメールを書くことができる。
	ヒアリング結果	どのような「断り」なのかをもっとはっきりさせた方がいい。
	修正後	社外から依頼の文書が来たときに、相手の気持ちを考えながら、丁寧な断りの手紙やメールを書くことができる。
9	修正前	自分の担当している商品や企画を売り込む文書を書くことができる。
	ヒアリング結果	「売り込む」をもっとはっきりさせる。宣伝という言葉を使うなどする。
	修正後	自分が担当している商品や企画を他社や顧客に宣伝し、売り込む文書を書くことができる。

（５）項目の技能分類の変更

以上のように各項目の記述を修正した後、「聴く」「読む」「話す」「書く」の４分類（各 10 項目の計 40 項目）について、課題（タスク）の技能による分類をより明確にするため精査した。そして、「聴く」「聴きながら読む/読みながら聴く」「読む」「話す」「書く」「聴いて話す」「聴いて書く」「読んで書く」の 8 分類に変更し、それぞれを 7 項目、3 項目、10 項目、4 項目、4 項目、6 項目、2 項目、4 項目の計 40 項目のビジネス日本語 Cds として再構築した。

4.2.2.3 本調査の実施と分析結果

修正したビジネス日本語 Cds の再検証のための本調査は、2010 年 8 月～2011 年 11 月にかけて日本国内の大学および日本語学校でビジネス日本語を学習している学習者 446 名を対象として実施した。被験者は BJT 模擬テストを受験した後、修正を加えたビジネス日本語 Cds のそれぞれの母語に翻訳された調査票に回答した。

4.2.2.3.1 ビジネス日本語 Cds の基本統計量

表 4-6 の通り、被験者を BJT 模擬テストの得点を用いて 3 つの水準に分け、それぞれの得点水準でビジネス日本語 Cds の合計（全 40 項目×各 5 段階評価＝最大値 200）の基本統計量を算出した。また図 4-3 に、表 4-6 で示した得点水準別に、ビジネス日本語 Cds の

合計の平均値と BJT 模擬テストの平均点を示した。これによると、BJT の得点水準 (Level) が上がるにつれ Cds の平均値が高くなっていることがわかる。なお、被験者の BJT 模擬テストの得点と Cds の合計値の間には、正の相関が見られた ($r=0.399$)。

表 4-6 BJT 模擬テストの結果による得点水準とビジネス日本語 Cds の基本統計量

	有効回答	Level 1 (下位)	Level 2 (中位)	Level 3 (上位)
人数 (人)	446	148	147	151
BJT 模擬テスト得点範囲		149～399	400～440	441～672
BJT 模擬テスト得点標準偏差	64.5	35.0	11.6	49.6
Cds 平均	130.10	116.38	135.12	138.66
Cds 標準偏差	25.83	25.64	23.53	22.47
Cds 最小値	48	48	69	76
Cds 最大値	191	185	191	191

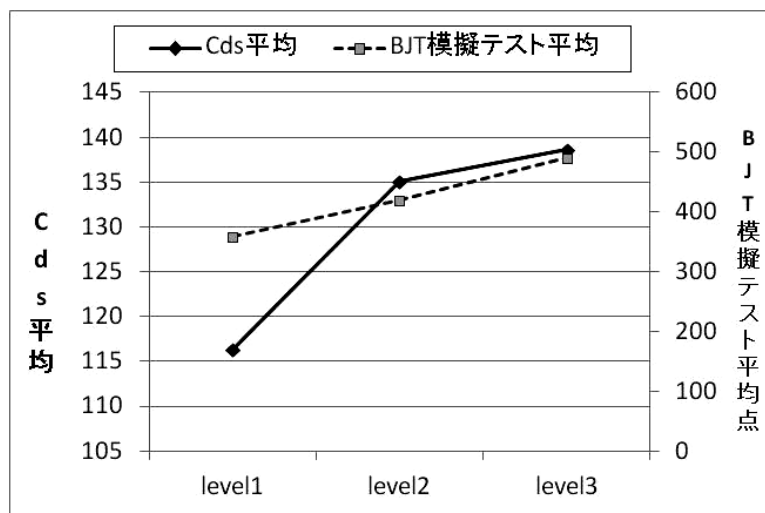


図 4-3 ビジネス日本語 Cds の平均値と BJT 模擬テストの平均点 (等人数による三水準)

4.2.2.3.2 ビジネス日本語 Cds の各項目と BJT の得点水準との関連

次に、ビジネス日本語 Cds の各項目について、BJT の得点水準との間に対応関係が見られるかどうかを検証した。まず、表 4-6 に示すように、BJT 模擬テストの得点を用いて等

人数になるように3つの水準に分け⁹、各水準のビジネス日本語 Cds の平均値を算出した。表 4-7 は、「聴く」「聴きながら読む」「読む」の各項目について、また、表 4-8 は、「話す」「書く」「聴いて話す」「聴いて書く」の各項目について、BJT 模擬テストの得点水準ごとのビジネス日本語 Cds の平均値を示している。この三群で分散分析を行った結果、ビジネス日本語 Cds の平均値の差は、すべての項目において 5%水準で有意であった。ただし、多重比較を行ったところ、Level 1 と Level 2 の平均値の差はすべての項目において有意差が確認されたものの、Level 2 と Level 3 の平均値の差が有意であった項目は6項目のみ(表内に†で表示)であった。

表 4-7 ビジネス日本語 Cds の課題と各水準の平均値 「聴く」「聴きながら読む」「読む」

Cds 項目 番号		課題(タスク)	Level 1	Level 2	Level3
聴く	1	上司からだされた自分の担当業務に関する指示が理解できる。	3.41	3.84	3.89
	2	社内で同僚同士が話している会話を近くで聴いて、どんな話題について話しているかわかる。	3.36	3.63	3.87 [†]
	3	会議や打合せで報告を聴き、何が重要なポイントなのかが理解できる。	3.13	3.65	3.74
	4	自分の担当分野に関する上司や先輩からのアドバイスやコメントの内容が理解できて、仕事に生かせる。	3.06	3.61	3.80 [†]
	5	会議のときに、議論の内容を聴きながら、賛成や反対など主張の違いがわかる。	2.93	3.50	3.65
	6	留守番電話のメッセージを聴いて、相手の要求や説明が理解できる。	3.16	3.51	3.70 [†]
	7	社外の人からの苦情や要求の内容が理解できる。	3.04	3.63	3.70

⁹ 後で述べる通り、本来は BJT のレベルガイドラインに沿って 3 水準に分けるべきだが、表 4-6 を見るとわかるように、本調査の受験者集団では BJT 模擬テストの得点の標準偏差が比較的小さかったため、BJT のレベルで群を分けるとサンプルサイズが偏ると考え、初めに等人数で分けて分析することにした。

聴きながら読む	8	自分の担当分野についてのプレゼンテーションの内容が理解できる。	3.09	3.61	3.83 [†]
	9	テレビの経済やビジネスに関するニュースの内容を理解することができる。	3.01	3.32	3.52 [†]
	10	ビジネスに関する研修及び講演の内容が理解できる。	2.75	3.20	3.34
読む	11	広告、パンフレット・カタログなどから必要な情報を見つけることができる。	3.64	4.02	4.07
	12	自分の仕事に必要な情報をインターネットで検索することができる。	3.57	4.05	4.11
	13	自分の担当分野に関係する記事やニュースが理解できる。	3.32	3.79	3.88
	14	社内で出されたお知らせや、回覧されている文書、仕事の指示を書いたメモの内容が2, 3分読んで理解できる。	3.01	3.59	3.70
	15	社外から来たビジネス文書を2, 3分で読んで、内容によって分類できる。	2.88	3.38	3.46
	16	社内で提案された新しい企画についての稟議書を読んで、企画の概要や提案理由がわかる。	2.94	3.47	3.56
	17	社外の人と数回やり取りした文書やメールを全部読んで、新しく決まったこと、変更があったことは何かがわかる。	3.05	3.65	3.75
	18	自分の担当業務に関する様々なビジネス文書を比較し、情報の優先順位をつけることができる。	2.98	3.41	3.62 [†]
	19	丁寧な印象を与えるために婉曲的な表現で書かれた社外から来たビジネス文書の意図がわかる。	2.84	3.38	3.52
	20	専門用語を使用した契約文書の内容が細部まで間違いなく理解できる。	2.47	2.92	2.94

表 4-8 ビジネス日本語 Cds の課題と各水準の平均値

「話す」「書く」「聴いて話す」「聴いて書く」「読んで書く」

(網掛けは Level2 と Level3 の平均値が逆転していることを示す)

Cds 項目 番号		課題(タスク)	Level1	Level2	Level3
話 す	21	社内と社外で話している相手との関係を考慮して、 場合に応じた適切な話し方を選ぶことができる。	2.85	3.26	3.43
	22	敬語に注意しながら上司や先輩に依頼ができる。	2.91	3.31	3.45
	23	専門的で複雑な内容でも、電話で用件をわかりやすく まとめて伝えることができる。	2.53	3.01	3.00
	24	自分の担当している商品や企画について、長くなり すぎないように簡潔に、効果的にプレゼンテーショ ンできる。	2.68	3.16	3.10
書 く	25	同僚や上司に業務を依頼したり業務の内容を伝えたり したいときにあまり時間をかけないで、さっとメ ールやメモを書くことができる。	2.78	3.16	3.28
	26	上司や先輩に敬語を使用して業務の報告書を書くこ とができる。	2.66	3.05	3.19
	27	自分が担当している商品や企画を他社や顧客に宣伝 し、売り込む文書を書くことができる。	2.53	2.97	2.92
	28	プレゼンテーションのための資料をわかりやすく作 ることができる。	2.80	3.28	3.18
聴 い て 話 す	29	営業などの業務で社外の会社を訪問したときや社外 の会議で集めた情報をまとめて、社内の人に報告す ることができる。	2.70	3.17	3.26
	30	会議や打合せで議論の内容を理解して、自分の意見 や疑問点を言うことができる。	2.80	3.26	3.28
	31	上司に対して疑問点や問題点を感じたときに、婉曲 的な言い方だが、正確に自分の思っていることを説	2.69	3.25	3.17

		明できる。			
	32	電話で話した内容の要点をまとめ、相手に確認することができる。	2.83	3.39	3.45
	33	取引先の人と話しているときに、よく理解できないことがあれば、相手の気持ちを考えながら、丁寧に聴き返すことができる。	2.81	3.44	3.40
	34	社外からの依頼や苦情の内容を理解して、相手の気持ちを考えながら、きちんと対応できる。	2.66	3.16	3.23
聴いて書く	35	会議やプレゼンテーションを聴きながら要点をメモすることができる。	2.80	3.31	3.38
	36	会議に出席して、会議の記録をとり、あとで正式な報告書を書くことができる。	2.57	2.91	3.05
読んで書く	37	苦情や問い合わせのメールや文書が来たときに、相手の気持ちを考えながら、丁寧な返事を書くことができる。	2.78	3.18	3.29
	38	新聞、インターネットなどから集めた情報をまとめて、報告書を書くことができる。	2.93	3.38	3.50
	39	自社の競争相手の競合他社に関する情報を集め、自社と比較した報告書を書くことができる。	2.74	3.18	3.25
	40	社外から依頼の文書が来たときに、相手の気持ちを考えながら、丁寧な断りの手紙やメールを書くことができる。	2.71	3.12	3.21

表 4-7 および表 4-8 は、多重比較の結果も示すように、多くの項目で、第一水準 (level1) と第二水準 (level 2) の間にはビジネス日本語 Cds の平均値の差が見られるが、上位の、第二水準 (level 2) と第三水準 (level 3) の間では差がないという傾向を示している。そこで、BJT のレベルガイドラインに沿って、J3 以下 (下位)、J2 (中位)、J1 以上 (上位) の 3 つの水準に分けたところ¹⁰、図 4-4 に示すように、上位の 2 つの水準間の差は少し拡

¹⁰ 各水準の人数は、J3 以下 (下位) が 225、J2 (中位) が 195、J1 以上 (上位) が 26 であった。

大するものの、やはり下位の2つの水準間の差を下回るものとなった。この三群で同様に分散分析を行ったところ、すべての項目において有意差が確認された。多重比較を行ったところ、Level 1 と Level 2 の平均値の差はすべての項目において有意差が確認されたものの、Level 2 と Level 3 の平均値の差が有意であった項目は2項目のみであった。ただし、これは各水準の人数の差が影響しているものと考えられ、実際の平均値の差は等人数による各水準間より大きくなっている。

なお、表 4-8 に網掛けで示した「話す 23」「話す 24」「書く 27」「書く 28」「聴いて話す 31」「聴いて話す 33」については、第二水準 (level 2) の平均値が第三水準 (level 3) をやや上回っている。これらの項目の課題 (タスク) を分析してみると、「話す 23」や「書く 27」は、「専門的で複雑な内容」「自分が担当している商品や企画」という記述からわかるように、専門性や個別性が高く BJT が評価基準として設定している能力の範疇に当てはまらないことが考えられる。また、「書く 28」については、「プレゼンテーションのための資料」を作ることが、ビジネス日本語能力以外の能力 (たとえばプレゼンテーションスキルや効果的な資料作りのノウハウなど) を測っている可能性があることや、「わかりやすく作る」の程度が回答者によって判断が分かれた可能性が考えられる。

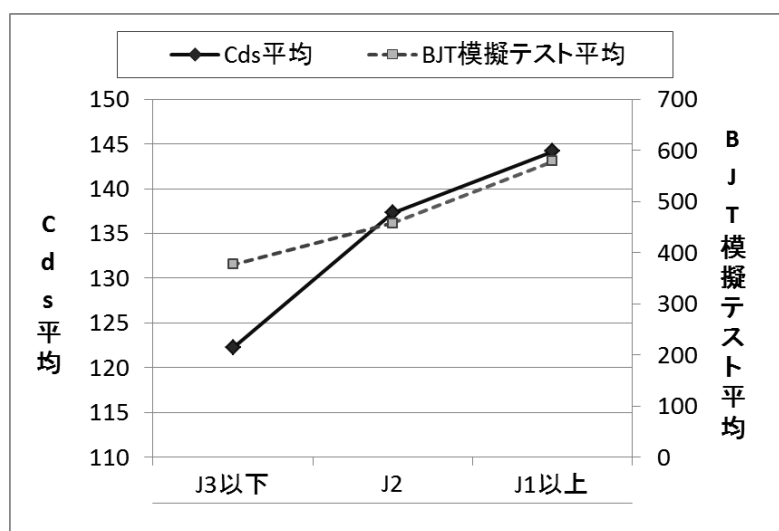


図 4-4 ビジネス日本語 Cds の平均値と BJT 模擬テストの平均点
(BJT レベルガイドラインによる三水準)

4.3 第4章のまとめと今後の課題

これまで繰り返し述べてきたように、日本企業による外国人材の活用が進まない要因のひとつとして、採用時の外国人材の能力の判定が難しいことが指摘されている。企業が外国人材に求めるビジネス日本語能力を客観的に評価・測定することができれば、採用や社内での活用をする際の判断基準となり、さらには大学や各自治体による効果的な就職支援にも役立つであろう。

そこで本章は、まず、ビジネス日本語能力を客観的に測るテストである **BJT** について、**BJT** が測定するビジネス日本語能力の内容を、開発当時の資料を参照しながら整理した。また、**BJT** のテスト項目を分析し、出題形式ごとにどのような能力を測定しているのかを明らかにし、**BJT** が測定するビジネス日本語能力の構成要素を記述した。その上で、ビジネス日本語能力を評価するシステムの構築をめざし、**BJT** が測定するビジネス日本語能力を参考に **Cds** を開発し、予備調査の結果に基づいて修正したビジネス日本語 **Cds** について再検証を行った。その結果、ビジネス日本語 **Cds** について、

- 1) 内部一貫性を示す信頼性係数（ α 係数）は0.978であり、修正前より高くなった。
- 2) 被験者の **BJT** 模擬テストの得点と **Cds** の合計値の間には、正の相関が見られ($r=0.399$)、修正前より高くなった。
- 3) 受容機能である「聴く」「聴きながら読む／読みながら聴く」「読む」の各項目について、**BJT** 模擬テストの結果による得点水準ごとに平均値を比較したところ、得点水準が上がるにつれて平均値も高くなっていた。このことから、**BJT** で測定されているビジネス日本語能力を反映した尺度として改良された。

以上のことから、本章で示したビジネス日本語 **Cds** の修正の手続きが、ビジネス日本語能力を評価する尺度の改善のために有用であることが確認された。しかし一方で、

- 4) 産出機能である「話す」「書く」「聴いて話す」の項目においては、得点水準が中位の群と上位の群の間に関連性が見られない傾向があった。

本章のビジネス日本語 **Cds** 調査では、**BJT** の得点による3つの水準のうち下位の2つ

の水準については被験者のビジネス日本語能力の識別に有効に機能しており、BJT を反映した尺度としての有効性を確認できた。上位の 2 つの水準についてより敏感に識別するためには、高度なビジネス日本語能力を習得している人材に対するさらなる調査が必要であろう。また、BJT が直接測定していない産出能力に関する項目では、回答者の能力をどの程度反映しているかについては、実際のビジネス場面でのパフォーマンスと自己評価の結果を比較する必要があるだろう。

一方で、本章で挙げた調査のほかに、2010 年 8 月から 2011 年 2 月にかけてアジア人財の修了生 10 名とその上司 8 名に対して行った当該ビジネス日本語 Cds 調査およびインタビュー調査によると、ビジネス日本語 Cds は「一般的なビジネスタスクで構築されているが業種によって求められていることが異なる」「業務上未体験のタスクについてどのように評価すべきかはっきりしない」などといった指摘がなされた。本章は、BJT が測定する能力を参考にビジネス日本語 Cds の項目を構築したが、今後は、業務経験が自己評価にどのように影響を与えるのかといった分析や、BJT で直接測定していない産出機能について、インタビュー調査やパフォーマンステストなど質的調査を実施するなどの検証を続けることによって、より汎用性の高いビジネス日本語 Cds に改善できるものと考えられる。そしてこのような手続きを経て最終的に完成したビジネス日本語 Cds は、日本企業への就職を希望する日本語学習者に対するビジネス日本語教育の実践や、さらには日本企業で働くビジネスパーソンの人事評価に資するものとなることが期待される。

本章で Cds によって示されたビジネス日本語能力の内容について、次章では、そのビジネス日本語能力を測るテストの、測定の質を高めるための具体的な方策を明らかにする。

第5章 ビジネス日本語テストの測定の質と改善

第4章では、BJTの出題項目を分析し、さらにBJTの測定する能力を参考にビジネス日本語Cdsを開発し、BJTの結果を反映する尺度としての有効性を検証した。そして、テストで測るビジネス日本語能力の具体的な内容を明らかにした。本章では、そのビジネス日本語能力を測るテストそのものに焦点を当て、測定の精度と質を改善する方策について明らかにする。

5.1 研究目的と研究方法

5.1.1 研究課題の設定

5.1.1.1 研究の背景と目的

BJTのほかに日本の公的な機関が実施する大規模な日本語試験としては、日本語能力試験、日本留学試験の日本語科目などがあるが、これらの試験はIRTを利用して作成されており、年に複数回、異なるテストフォームを用いて実施されているにも関わらず、その結果は互いに比較可能なものとなっている。そして、そのキーとなっているのが、項目バンクの存在である(2.2.6.2節を参照)。しかし、試験運用の観点から、項目バンクの構築とその維持には膨大な労力が投入されている。IRTによるテスト開発においては、新作項目の統計的性質を精査し、そのチェックを通った項目のみが項目バンクに追加され、再利用されるのであるが、新作項目の歩留まりは、決して良くないと推測される。すなわち、IRTを用いてテストの開発・運用を行う現場では項目バンクに質の高いテスト項目を十分に確保することは容易ではない。

これまで、公的な大規模日本語試験においては、その信頼性や実用性についての検証は多数行われてきているが(日本語能力試験実施委員会ほか、2011など)、項目バンクの構築の実際に関する研究は少ない。そこで、本章では、BJTのテスト項目を用いて項目修正の調査・分析を行う。具体的には、テスト項目の項目統計量が許容範囲を超えるために、項目バンクへの追加が適わなかったテスト項目(以下、特異項目とする)を抽出し、テスト項目の形式・内容を検証し、それらのテスト項目のどこに瑕疵があり、どのように修正することで項目統計量が改善するのかという点を、実証的に明らかにする。具体的には、以下の2つの課題を設定し、研究を行った。

- (1) どのような要因がテストの質に影響するか。
- (2) テスト項目の修正によって、測定の質は改善されるか。

テスト項目の項目統計量を基にしてテスト項目の選択を行うということは、項目バンクに含まれる項目の質を保証する作業であり、従って、そこから作成されるテストフォームの質の保証を意味する。しかし、テスト項目の質は一意に定義されるものではなく、少なくとも、テストの妥当性ならびに信頼性の二つの側面から検証されるべきものである。本章では、後述するように、主にテストの信頼性に関連した、テスト項目の統計的性質を基にして特異項目を選出し、改善を目指している。

5.1.1.2 先行研究と本章の意義

テスト項目の編集や再検討について、Baranowski (2006) は、テスト項目を適切に編集することによって、テスト全体の妥当性を向上させることができると述べた上で、その具体的な事例として、設問文で否定的な用語を避ける、選択肢は構造や文の長さをだいたい同じにする、設問文や選択肢にあいまいな用語の使用をさける、などを挙げているが、修正の効果を統計的に示している訳ではない。また、Abedi (2006) は、特に非母語話者の受験者の測定においては、不必要な言語的複雑さ (linguistic complexity) を避けるべきであり、英語のテストにおいて影響を与えるその言語的特徴を挙げ、語の使用頻度や親密度の高いものを使用する、動詞句は受動態より能動態を使用するなど、修正の例示をしながら説明している。しかし、これらの指摘は、一般的にテスト項目の作成時に意識すべき点として共通の認識とすべきものであるが、日本語能力を測定する試験に特化したものではない。

テスト項目の内容や表現がテストの測定の質に与える影響についての研究には、安永・石井 (2011, 2012) などがある。安永・石井 (2011) は、日本と韓国の中学生を対象とし、群馬県児童生徒学力診断テストの国語読解テストのうち、自由記述式問題を用いて、設問文中の単語の難しさがテスト項目の得点率 (難易度) や識別力に与える影響について調査を行っている。設問文中の記述について具体例を挙げて説明するタイプの 2 項目を比較したとき、一方の項目は安定した難易度を示したのに対し、もう一方の項目は日韓ともに低い得点率となっているのは、「設問の細かい文言の違いが項目特性値の違いの原因である可能性が推察される」(p.108) とし、得点率の低い項目の設問文にある「人間の外的な

身体能力の拡張」の「拡張」を「拡がり」と変えた設問文を作成して日本の中学3年生703名に実施した。その結果、設問文中の単語の難しさは得点率及び識別力に影響を及ぼさないことが示唆されたと報告している。

安永・石井（2012）は、小中一貫教育校の中学3年生169名を対象とし、群馬県児童生徒学力診断テストの国語読解テストの複数の項目を操作したテストを実施し、それらの操作が受験者の解答に及ぼす影響を検討している。項目に加えた操作とは、1) 相対する単語が入る空所の位置（本文と同じ順序・本文と反対）と、ヒントの位置（空所の横・空所の下）の変更、2) 本文の文頭に記号を振り当てて記号を選択する選択肢問題だったものを、実際に本文から一文抜き出す記述式問題の設定に変更、3) 空所の数と図の提示を変更、4) 空所前の単語における説明の有無（単語に関する説明の繰り返し）である。調査の結果、1) に関する設定の違いが受験者の解答傾向に影響を与え、得点率と識別力に変化が見られたことが示されたものの、読解能力の高い受験者を対象としたことにより、項目操作の影響を検出できなかった可能性も考えられる、と述べている。これらの研究は、テスト項目の識別力や難易度がどのような要因に影響されるのかを実証的に示したもので、興味深い。

外国語としての日本語を学ぶ日本語学習者を対象とし、大規模な項目バンクを持つようなテストにおいて、このような項目操作を行った実証研究の例は管見の限り見られない。また、安永・石井（2011, 2012）においてテスト項目に加えた操作は、設問文内の一部の語句や空所の位置などの表面的な部分である。調査の対象となる受験者数は200～700名程度であり、解答データの項目分析は古典的テスト理論の枠組みで行っているが、本来の受験者と調査対象とした受験者の能力分布の違いが結果に影響している可能性にも言及されている。

本章は、日本語学習者を対象とし、不適切なテスト項目を修正することによって適切な項目に改善することを目的とし、その具体的な方策に関する知見を得るための研究である。テスト項目に加える操作の方針は、日本語教育の観点から不適切なテスト項目の特徴を分析し、音声や語彙、文法といった言語的な要素による問題点、あるいは受験者の視点や焦点といった認知言語学的な要素による問題点にも着目して決定し、その方針に沿って修正を行う。また、比較的規模の大きいデータを収集し、IRTに基づく等化を行い、項目統計量を共通の尺度に位置付けた上で比較検証する。このようにテスト項目に関する質的・量的両側面からの検証結果は、「不適切なテスト項目をいかにして適切なテスト項目に修正す

るか」ということに対する具体的な方策を示すことになり、質の高いテストあるいはテスト項目の作成・改善に寄与するものである。

5.1.2 研究方法

本章では、以下の手順で検証・分析を行う。

- (1) 過去の BJT の出題項目から、項目統計量が特異であったテスト項目（特異項目）を抽出し、その解答の傾向の内実を検討し、項目統計量が特異であった要因についての検証を行う。
- (2) 上記(1)の検証結果に基づいて特異項目に修正を加える（以下、修正項目）。
- (3) 過去の BJT の出題項目で、項目バンクに保存されており、項目統計量に異常がない項目（以下、共通項目）と、修正項目を合わせた実験テストフォームを作成する。
- (4) 国内外の日本語学習者に実験テストフォームを実施する。
- (5) 上記(4)で得られた項目反応データを、共通項目を用いて BJT の規準集団に等化する。
- (6) 特異項目と修正項目の項目パラメタを比較することによって、特異項目に施した修正が項目統計量の改善に寄与したか、定量的・定性的に分析する。

なお、(3)の共通項目については、トピックや形式、難易度のバランスを考慮して項目バンクから選んだ。また、本章で行った IRT による分析について、BJT では 3PLM (2.2.6.1 節を参照) を採用しているため、本章においても、同モデルを用いて分析を行う。

5.1.2.1 特異項目の抽出と修正内容の検討

最初に、過去にBJTで出題されたテスト項目の中から、項目統計量が特異なもの、すなわち特異項目を表5・1に示す基準によって特定し抽出した。項目統計量とは、テスト項目の古典的識別力を表す項目－総合点相関、また、IRTによる困難度パラメタ、当て推量パラメタを指す。この特異項目の基準は、JETROがBJT実施後に、出題したテスト項目の検証を行う際、再利用不可項目の基準として実際に採用していたものであるが(2.2.7.1節を参照)、基本的には、項目特性曲線を描いた時に、潜在特性値 θ の有効範囲 (BJT の場合は およそ $-3 < \theta < 3$) で正答確率の値の変化が少ない項目を抽出するための基準であ

る¹¹。BJT は 3PLM を採用しているため、IRT識別力パラメタを用いるよりも項目―総合点相関を手がかりとした方がこの目的に沿った特異項目を抽出できることが経験的に判明している。

なお、特異項目への解答データのサイズは項目によってばらつきがあるが、おおむね1500～8000で、延べ数は47855である。

表 5-1 抽出した特異項目の基準

特異項目の基準(いずれかあるいは複数の条件に当てはまる)	該当項目数
項目―総合点相関 0.2 以下	24
IRT 困難度パラメタ 3.0 以上または-3.0 以下	2
IRT 当て推量パラメタ 0.4 以上	6
特異項目 合計	28

次に、抽出した特異項目それぞれについて、項目統計量が特異であった要因についての検証を行った。ここでの検証は、5.1.1.2節で挙げた先行研究の指摘を踏まえ、日本語能力一般ならびにビジネス日本語能力を測定するテストに特化して、項目統計量と項目の形式・内容の詳細を対照しながら行った。音声・文法・語彙など、言語要素に関する問題はもとより、設問の方法、選択肢の形式、漢字やカタカナ語などの表記の問題、パラフレーズなど正答に至る手がかりとなる情報の分布状態、受験者の視点や焦点など認知言語学的な問題などの観点からも分析を加えた。

5.1.2.2 特異項目の修正方針

前節で述べた特異項目の検証結果をもとに、その項目統計量が特異であった要因を克服する形で形式・内容を修正した修正項目を含む実験テストフォームを作成した。特異項目の検証の過程で見出された修正方針は、以下のような点である。なお、以下に挙げた修正方針について、1点だけでなく複数の点を反映させたテスト項目もある。

¹¹ 受験者の能力特性値は正規分布していると考えられるため、有効範囲外の受験者数は少ない。したがって、有効範囲で正答確率の値の変化が少ないテスト項目は、能力特性値を推定する上で機能を果たさないものと言えるため。

- 1) ビジネス場面で使われる抽象的な表現が、困難度を上げ、識別力を下げる可能性が考えられるため、テスト項目のテキスト部分に複数の情報を配置し、その複数の表現から手掛かりを得られるようパラフレーズすること。
- 2) 聴読解テスト項目では、視覚情報（文字情報）と聴解情報との整合性を保つようにすること。
- 3) 読解テスト項目の設問文では、いつ、だれが、どこで、だれに対して、どのような状況で当該文書を書き、それを読む者は、いつ、どこで、どのような状況で読むのかなど、状況や文脈を適正に設定すること。
- 4) 読解テスト項目においては、受験者の母語の干渉が識別力を低くする要因となりうるため、選択肢の作成では、たとえば表記や発音などの類似・非類似性といった対照言語的観点から、受験者の諸母語（主に中国語、韓国語、英語を想定）を意識すること。
- 5) キーワードのマッチングで正答を選べるテスト項目は、日本語能力の低い受験者も正答を得やすく、識別力が低くなるため、できるだけ避けること。
- 6) そのほか、1)から 5)以外の微調整

実験テストフォームの構成は表5-2の通りである（テストの内容は巻末の資料2を参照）。

表 5-2 実験テストフォームの構成と項目数

部門	§	共通項目 (BJT 項目バンク と共通の項目)	修正項目 (BJT 項目バンク の特異項目を修 正した項目)	計
聴解	1	3	3	6
	2	3	3	6
	3	5	4	9
	計	11	10	21
聴読解	1	5	4	9
	2	5	4	9
	計	10	8	18
読解	1	3	3	6
	2	3	3	6

	3	5	4	9
	計	11	10	21
合計		32	28	60

5.1.2.3 調査の実施

調査は、2009 年から 2010 年にかけて、国内および国外で日本語を学習する日本語学習者計 1128 名に対して実施した。国内外の受験者の内訳は表 5-3 の通りである。なお、受験者の日本語能力レベルは、中級（日本語能力試験 N2 レベル）以上で、ほとんどがビジネス経験のない学生であった。BJT の実際の受験者の約 7 割が大学生・大学院生、専門学校生・日本語学校生であることを考慮すると、このサンプルは実受験者の集団をほぼ反映していると考えられる。

表 5-3 受験者数の内訳

調査実施地			受験者数計
国 外	タイ, 韓国, 中国, 香港の大学・日本語学校(計 18 機関)		688 名
	国籍別	中国(431 名) 韓国(176 名) タイ(81 名)	
国 内	東京近郊, 大阪, 石川, 沖縄県の大学・日本語学校(計 12 機関)		440 名
	国籍別	中国(台湾含む)(334 名) 韓国(38 名) タイ(3 名) その他(65 名)	
合計			1128 名

受験者は、実験テストフォームの問題冊子と音声（聴解および聴読解）によって解答した。解答は、BJTと同様、すべてマークシートにより行った。

実験テストフォームの解答データの分析にあたっては、特異項目の項目統計量と修正項目の項目統計量を比較可能なものとするために、共通項目のIRT項目パラメタを用いて、共通項目法による尺度等化を行った（前川，1991 ならびに Arai & Mayekawa, 2011 に示す方法を用いた）。この方法は、複数のテストフォームを異なる集団に実施し、そのテストフォームごとに推定された共通項目の項目特性曲線がなるべく一致するような等化係数を求めるものである。本章の場合には、実験テストフォームの受験者の解答データから、

その受験者の潜在特性値 θ の平均が 0, 分散が 1 という仮定の下で項目パラメタを推定し, そこで得られた共通項目の項目パラメタの推定値を, 項目バンクの値と比較することにより, 等化係数を求めた¹²。

図5-1 から図5-3 に, 共通項目の, 項目バンクにおける項目パラメタと, 実験テストフォームで得られた項目パラメタの散布図を示す。つまり, 共通項目について, 項目バンク上で登録されていた項目パラメタと, 今回の実験テストフォーム受験者のみで算出された項目パラメタを比較したものである。図5-1はIRT識別力パラメタ, 図5-2はIRT困難度パラメタ, そして図5-3はIRT当て推量パラメタを, それぞれプロットした。これら3つのパラメタはどれも, モデル上では直線上に乗るべき値である。したがって, 直線からのズレは, データとモデルとの乖離を示す誤差であると考えられるが, そこにシステムティックなズレは見られず, 共通項目を用いた項目パラメタの等化が行えると判断した。ただし, 図 5-3 に示す実験テストフォームで IRT当て推量パラメタの値が 0.5 以上として推定された4 項目は, 等化係数の算出からは除外した (図中に△で表示)。なお, 得られた等化係数 (2.2.6.2節を参照) は, 実験テストフォームの尺度 θ と項目バンクの尺度 θ_Z の間に $\theta_Z = 0.224 + 0.740 \theta$ という関係を表すものであった。

¹² 等化係数の具体的な求め方については,
http://www.ms.hum.titech.ac.jp/mayekawadocs/Rpackages/HTML_help/lazy.irt/calr.html
に公開されているパッケージの中関数を使用した。

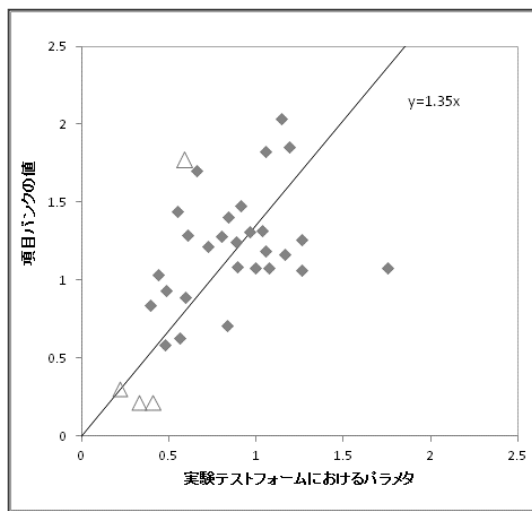


図 5-1 共通項目の IRT 識別力パラメタ

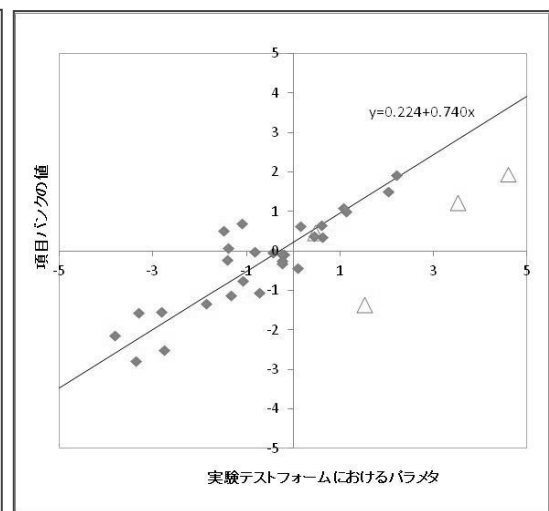


図 5-2 共通項目の IRT 困難度パラメタ

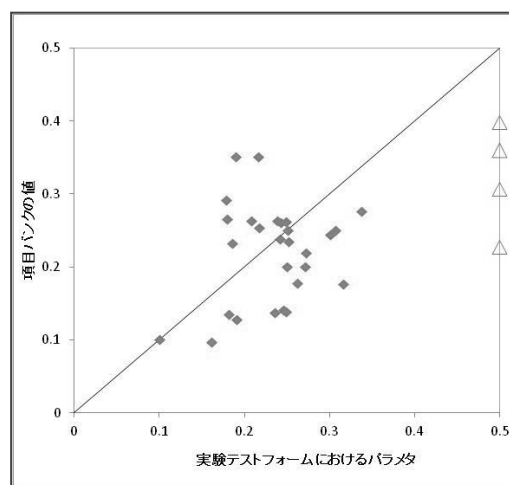


図 5-3 共通項目の IRT 当て推量パラメタ

5.2 結果

5.2.1 基本統計量

受験者の得点分布および基本統計量を、表 5-4 および図 5-4 に示す。図 5-4 に示す実験テストの得点分布は、BJT の規準集団に等化して得られた受験者の潜在特性値を、BJT の得点変換表を用いて 800 点満点に換算して得られたものである。BJT の規準集団での平均

が 400 点であることから、本調査における実験テストの受験者の能力は、BJT の規準集団における受験者と比較し、 θ で 0.224、得点にして約 14 点高い集団である。

表 5-4 実験テストの基本統計量

平均値	414.3
標準偏差	76.9
最大値	719
最小値	153
α 係数	0.85

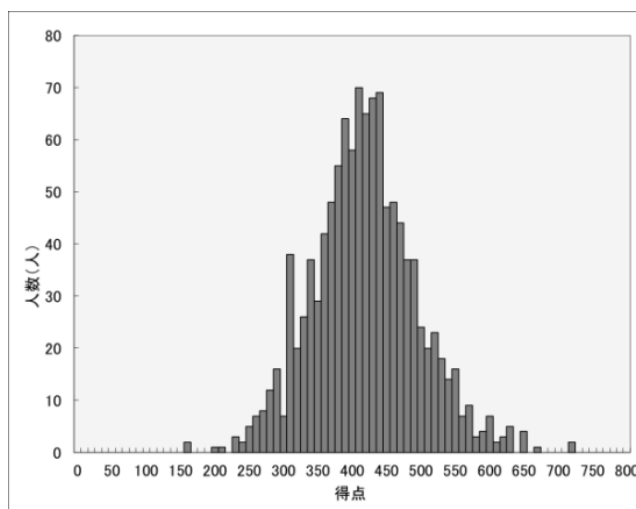


図 5-4 実験テストの得点分布

5.2.2 項目統計量の比較

本節では、修正前（特異項目の値）と修正後（修正項目の値）の項目統計量を比較し、修正が項目統計量の改善に寄与したか否かをチェックした。識別力の指標としては、古典的な項目－総合点相関と IRT 識別力パラメタを用いた。また、それに加えて、IRT 困難度パラメタ、IRT 当て推量パラメタの項目統計量に関しても、修正前後を比較した。ただし、項目－総合点相関に関しては、総合点の計算が、各項目が含まれるテストフォームに依存するため、ここでは特異項目・修正項目共に項目の正誤と IRT により推定された潜在特性値 θ との間の相関係数（項目－特性値相関とする）を用いた。なお、特異項目に関して、その項目－総合点相関と項目－特性値相関はきわめて近い値であり、両者の相関係数は 0.905，差の絶対値の平均は 0.025 であった。

表 5-1 に示した特異項目の基準を用いて修正項目のチェックを行った結果、28 個の特異項目のうち、修正後も未だこの再利用不可の基準内に入るものは僅か 5 項目であった。

図 5-5 から図 5-8 に修正前後の項目統計量の変化を示した。横軸が特異項目の値すなわち修正前のパラメタ、縦軸が修正項目の値すなわち修正後のパラメタを表す。

図 5-5 は、項目－特性値相関の変化を表しており、項目－特性値相関の値は、修正後基準である 0.2 をおおむね超えている。図中に丸で囲われた範囲に入る 5 項目は、修正後

も 0.2 を超えていない。聴解 2-5, 読解 1-1 については改善されていないが, 聴解 2-1 を含む残り 3 項目については若干の改善が見られる。

図 5-7 は IRT 困難度パラメタの変化を表す。これを見ると, 修正後は全項目が -3.0 から 3.0 の適正值の間に収まっている。図 5-8 は IRT 当て推量パラメタの変化を見ると, 修正後は 2 項目を残して 適正值である 0.4 以下に収まっているが, 図中に示したように, 聴読解 2-1 および聴解 2-5 については改善されていなかった。

図 5-6 を見ると, IRT 識別力パラメタについては, 特異項目の基準には含まれない範囲でパラメタ値がやや低くなった項目もあるが, 修正前に 1.0 を下回っていたものについてはおおむね改善されている。ただし, これらの項目は主に, IRT 困難度パラメタが基準を超えていたために特異項目に分類されていた (IRT 識別力パラメタには異常はなかった) が, IRT 困難度パラメタの絶対値が極端に大きい場合は, 同項目の IRT 識別力パラメタの推定が不安定になることがわかっている。そのため, このような項目における修正前 (IRT 困難度パラメタが改善される前) と修正後の識別力パラメタの比較は意味を持たない。また IRT 識別力パラメタ値は低くなっているが, 困難度パラメタや当て推量パラメタは適正の範囲である。

以上のように, 項目の修正によって項目統計量は全体として改善していることがわかる。

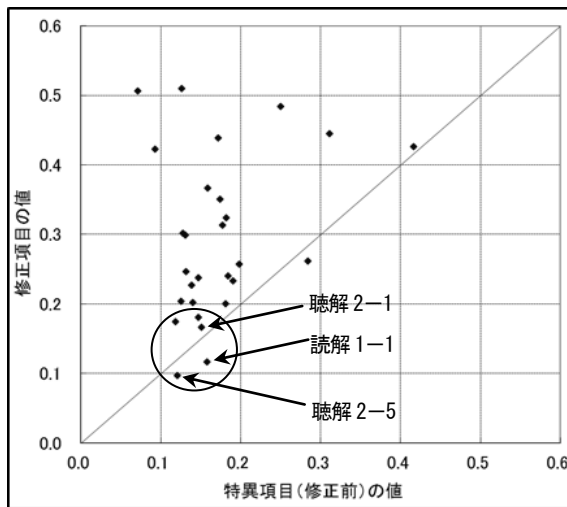


図 5-5 特異項目と修正項目の
項目-特性値相関

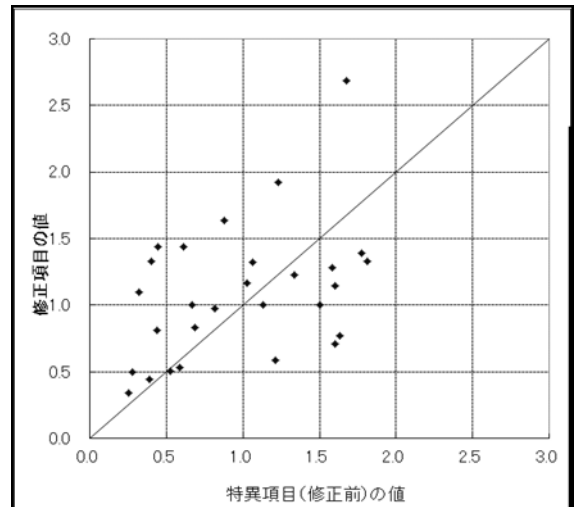


図 5-6 特異項目と修正項目の
IRT 識別力パラメタ

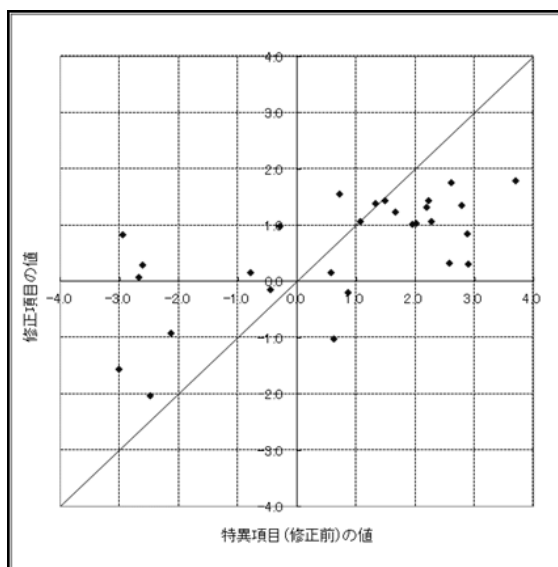


図 5-7 特異項目と修正項目の
IRT 困難度パラメタ

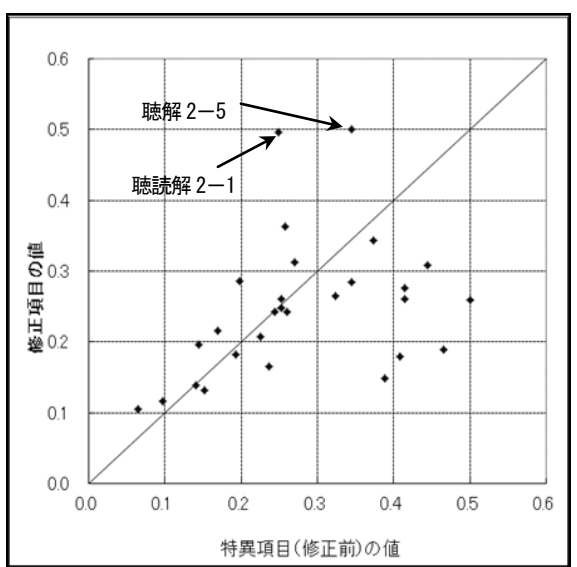


図 5-8 特異項目と修正項目の
IRT 当て推量パラメタ

5.2.3 項目特性曲線の比較

3PLM においては、識別力パラメタ、困難度パラメタ、当て推量パラメタの 3 つのパラメタによって項目特性が表されるが (2.2.6.1 節を参照)、これらのパラメタは互いに交絡した形で項目特性曲線に影響を与えている。したがって、項目特性の改善を吟味するためには、単にパラメタ値を検討するだけではなく、項目特性曲線を描くことが重要である。

特異項目について、修正前と修正後の項目特性曲線を比較したものが図 5-9 から図 5-36 である。縦軸が正答確率、横軸が潜在特性値 θ を表す。これらの図から、多くの項目で潜在特性値 θ の有効範囲で正答確率の値が大きく変化するように改善されていることがわかる。

たとえば、図 5-11 の聴解 1-4 や図 5-18 の聴解 3-19 の項目特性曲線を見ると、修正後（修正項目）の項目特性曲線は修正前（特異項目）のとは比べ、傾きが大きくなっているが、これは項目識別力の向上を意味している。図 5-14 の聴解 2-6 や図 5-32 の読解 2-6 の項目特性曲線は、傾きには変化はないものの、修正後に位置が左側に移動しているが、これは項目困難度が適正の範囲に修正されたことを示している。図 5-20 の聴読解 1-3 や図 5-25 の聴読解 2-6 は、項目特性曲線の負の側の漸近線の位置が、修正後、下がっていることがわかるが、これは当て推量パラメタが適正に修正されたことを示している。一方で、前節で述べたように、図 5-13 の聴解 2-5 や図 5-27 の読解 1-1 などについては、項目特性曲線の傾きや位置、また漸近線の位置に変化がなく、修正によって項目特性値が改善されていないことがわかる。5.2.4 節に、具体的な分析とともに、いくつかの項目の内容について詳述する。なお、項目特性の変化について、本章の調査では、聴解、聴読解、読解の部門別では主だった特徴は見られなかった。

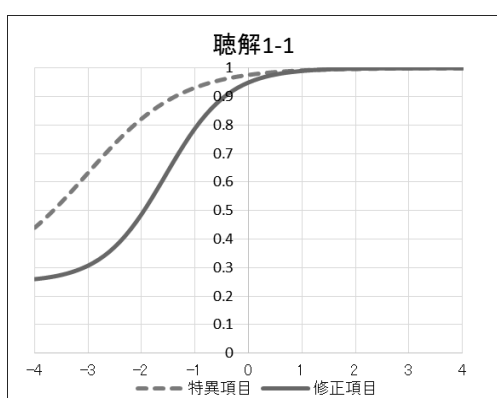


図5-9

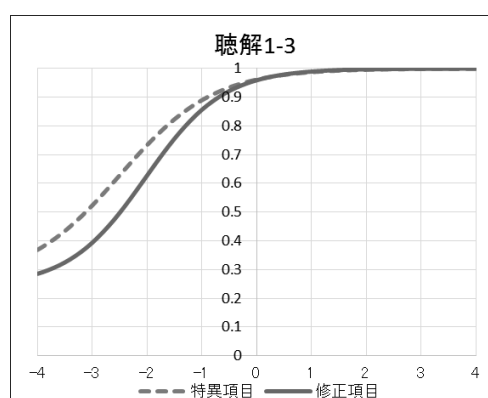


図5-10

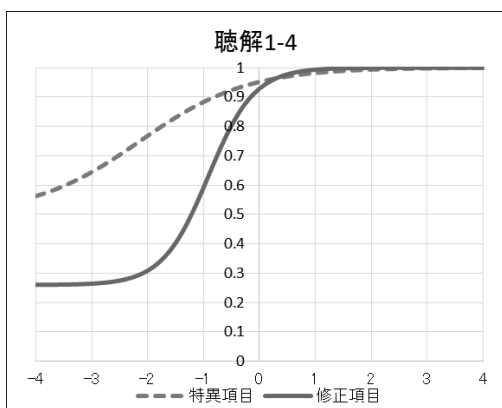


図5-11

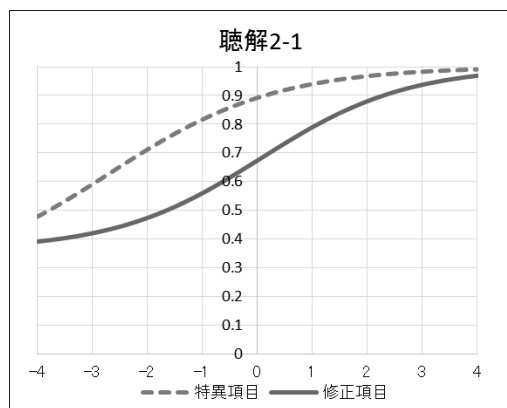


図5-12

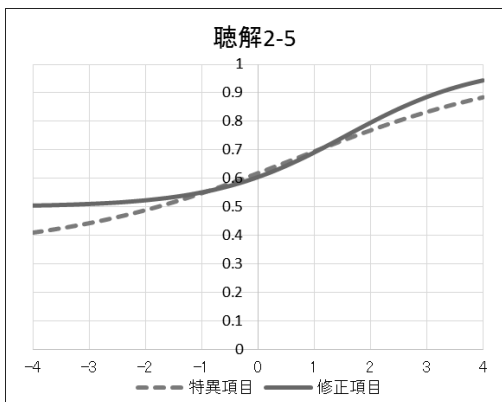


図5-13

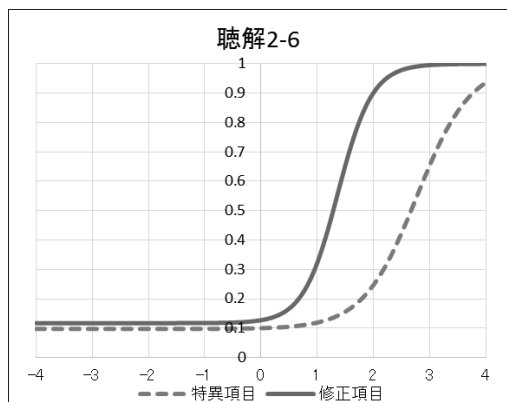


図5-14

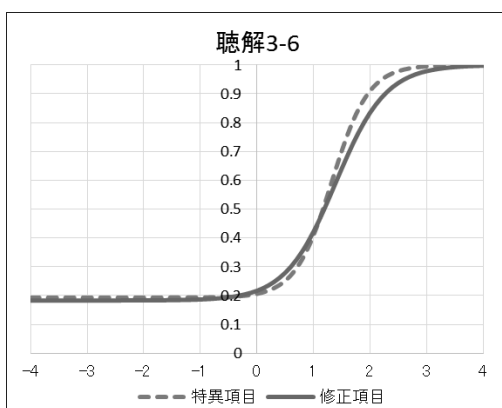


図5-15

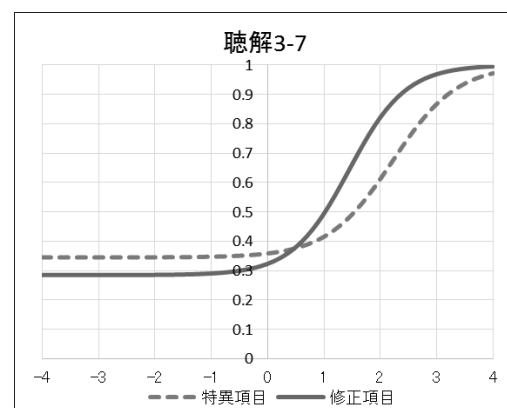


図5-16

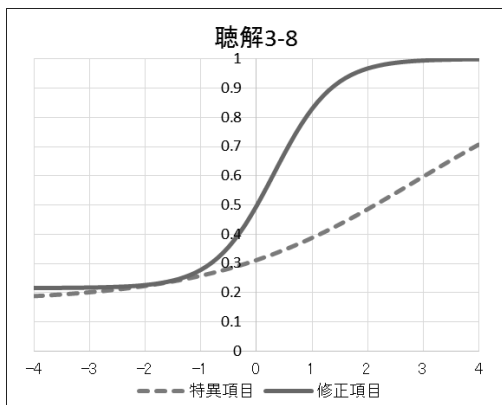


図5-17

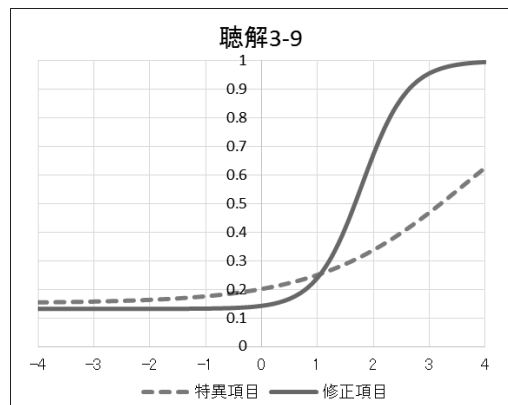


図5-18

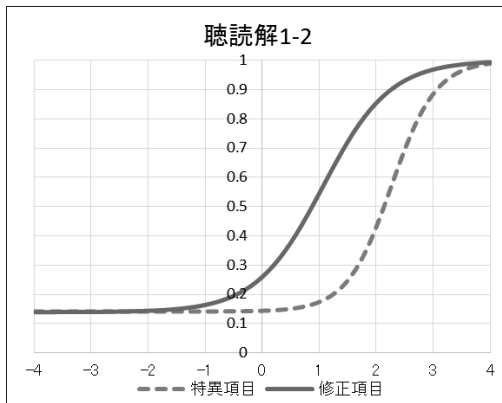


図5-19

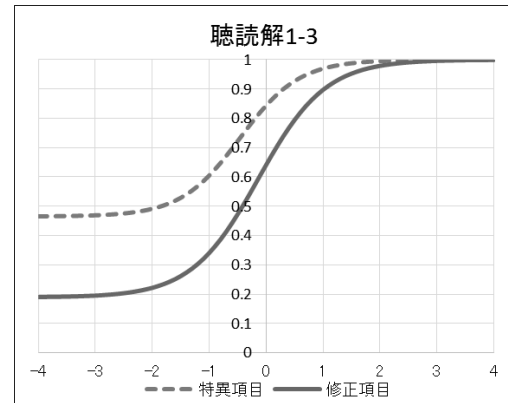


図5-20

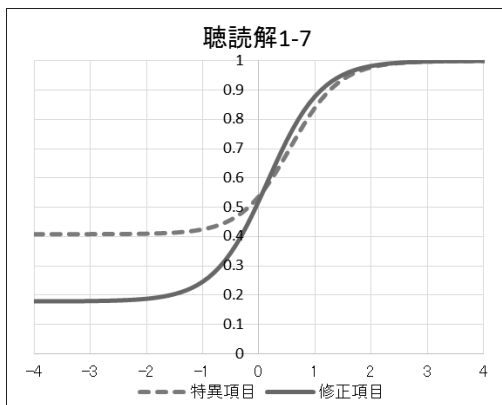


図5-21

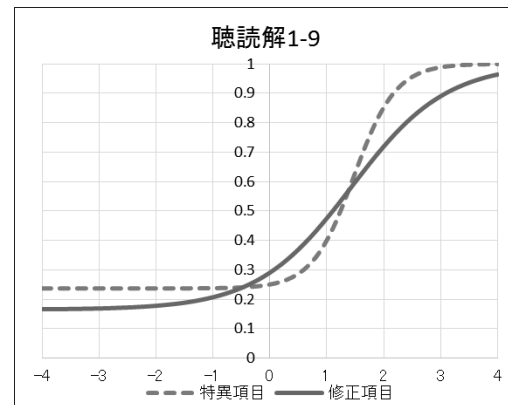


図5-22

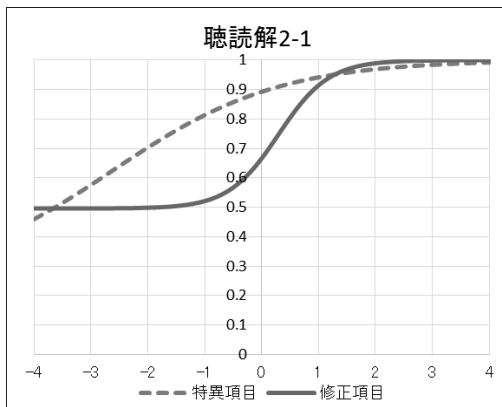


図5-23

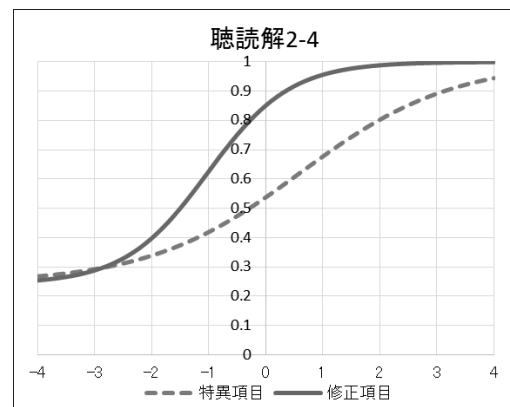


図5-24

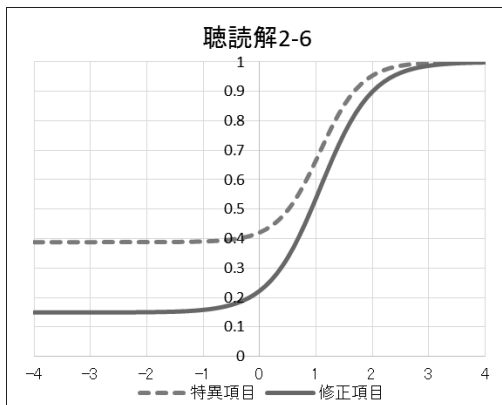


図5-25

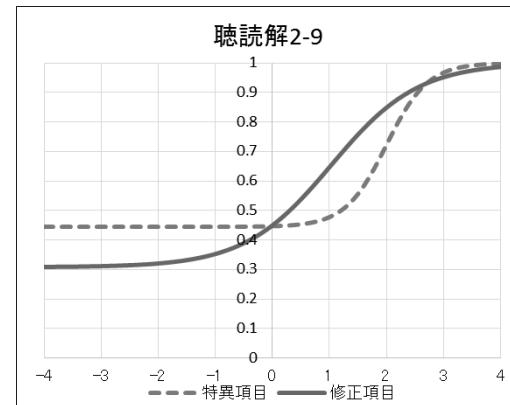


図5-26

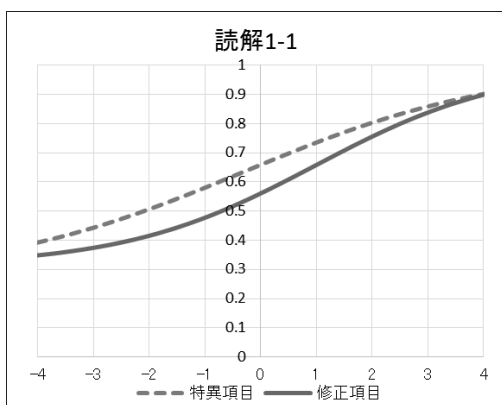


図5-27

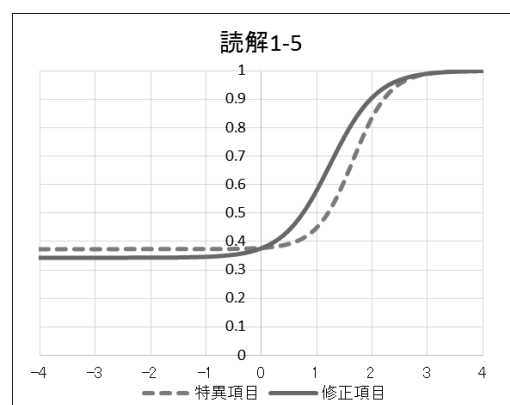


図5-28

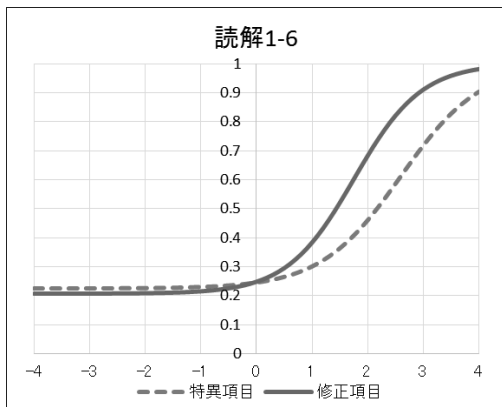


図5-29

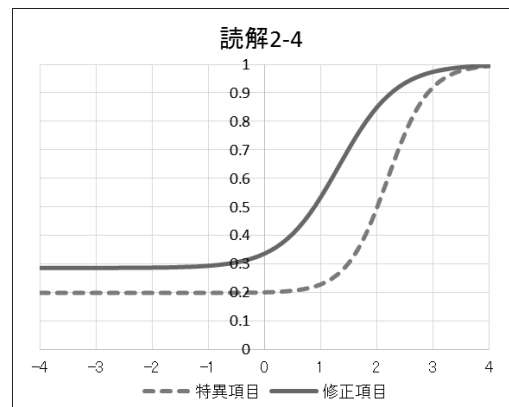


図5-30

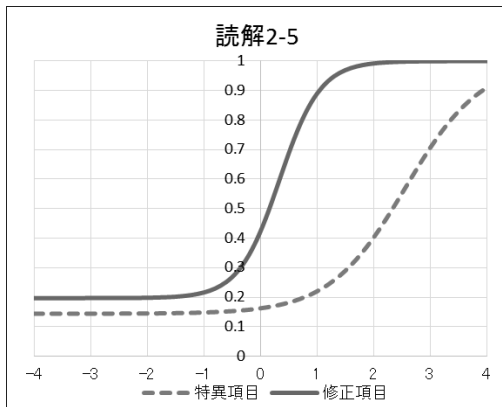


図5-31

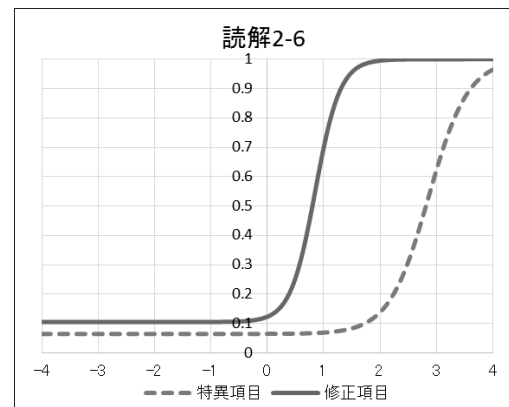


図5-32

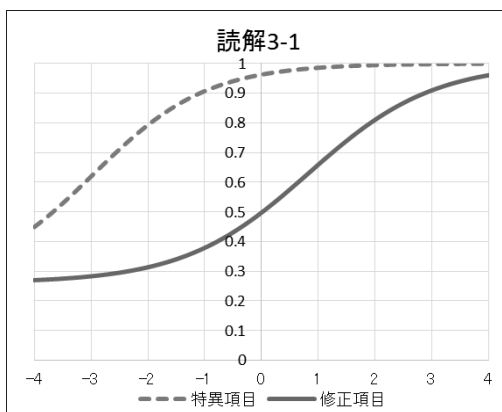


図5-33

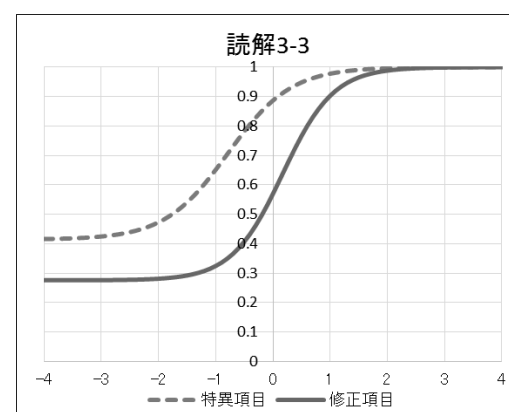


図5-34

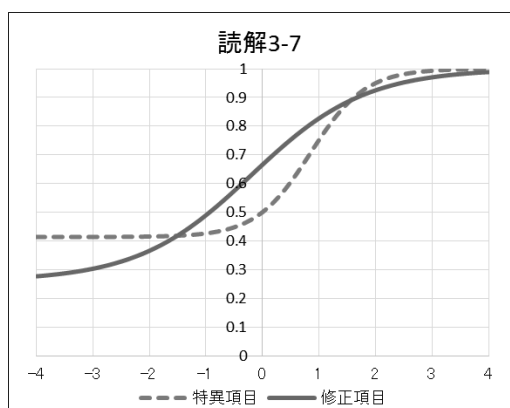


図5-35

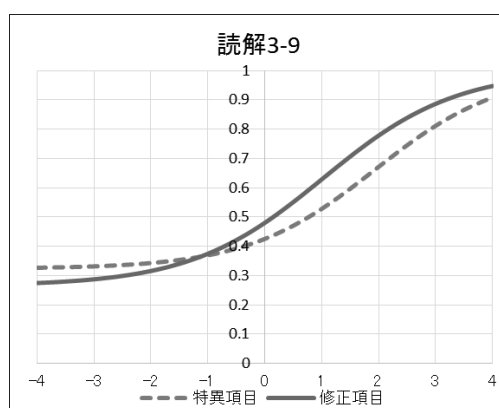


図5-36

5.2.4 修正の実際例

本節では、28 項目の特異項目のうち、項目統計量が特異である要因の分析内容および修正の内容と、修正後の項目統計量の例を挙げる。また、受験者の解答の傾向からテスト項目を分析するために、修正前と修正後の設問回答率分析図（2.2.7.1.2 節を参照）を比較する。本章で用いた設問回答率分析図は、受験者を潜在特性値 θ にしたがって最も低い群（LL）から最も高い群（HH）までの 5 群に分け、群ごとに各選択肢の選択率をプロットしたものである。なお、正答選択肢のプロットは太い実線で示した。

5.2.3 節で示した項目特性曲線の比較を再掲し、修正前と修正後の項目特性の変化を示す。

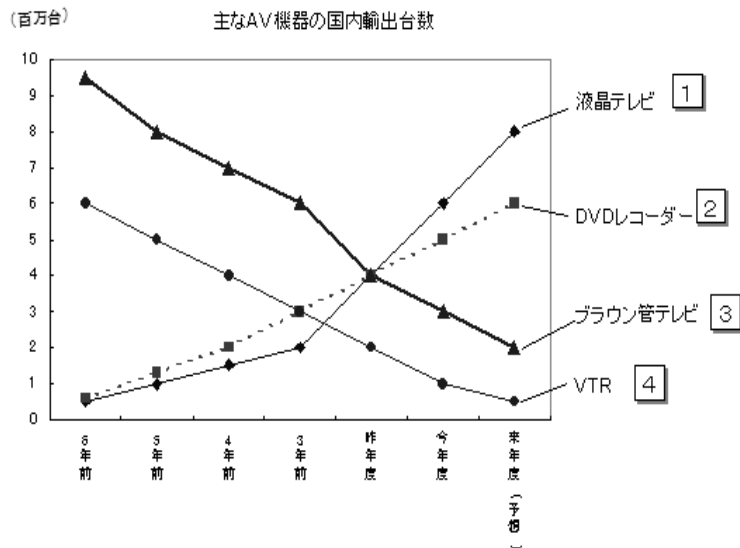
(1) 項目番号：聴読解 2-9

<修正前（特異項目）>

【設問】（音声情報）

セミナーで、男の人が国内の AV 機器の売上動向について話しています。男の人はこのあと、どの製品について話しますか。

【視覚情報】



【会話文】(音声情報)

男1: このグラフは、ここ数年の主な音響・映像機器の日本国内出荷台数の推移を示すものです。昨今、AV機器は急速なデジタル化、そして高度情報化の流れの渦の中にあります。まず、DVDレコーダーを例にとって見ますが、その国内出荷台数は3年前、ついにVTRを抜き、その地位が逆転しました。で、その次の年は、液晶テレビがブラウン管テレビを上回った年となりました。それでは、昨年の逆転劇の主役を例にとって、もう少し詳しく、これからの生き残り戦略についてお話したいと思います。

(正答: 1)

<統計情報>

修正前(特異項目)の統計データ

項目—特性値相関0.147 IRT困難度パラメタ2.000 IRT当て推量パラメタ0.445

修正項目の統計データ

項目—特性値相関0.237 IRT困難度パラメタ1.032 IRT当て推量パラメタ0.308

設問回答率分析図

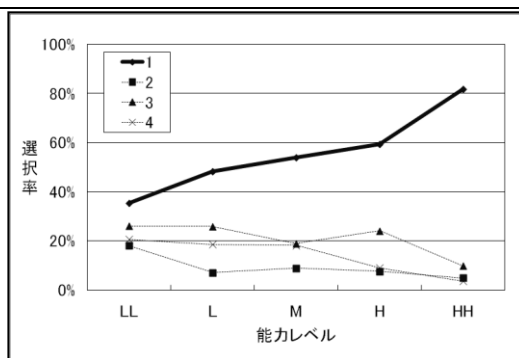


図 5-37a 聴読解 2-9 特異項目

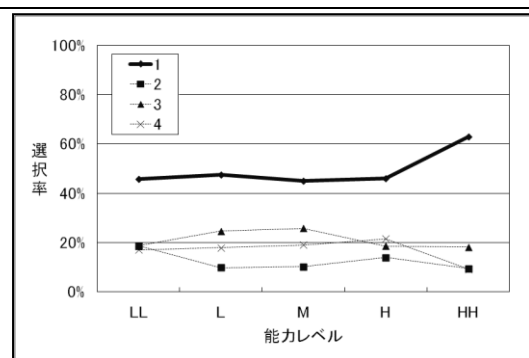


図 5-37b 聴読解 2-9 修正項目

項目特性曲線

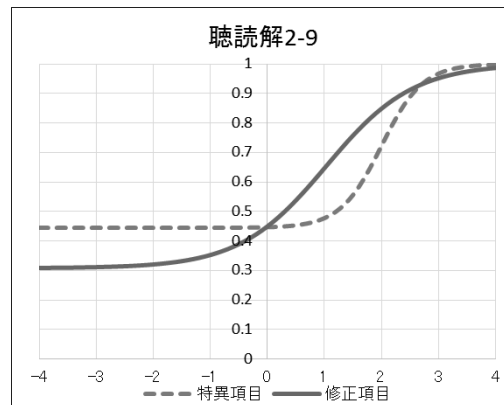


図 5-26 (再掲)

図 5-37a を見ると、正答選択肢が最も多く選ばれているものの、潜在特性値の最も低い受験者群 (LL) から中上位の受験者群 (H) までの正答選択肢の選択率が上下しており、項目-特性値相関が低く、受験者の能力を識別する力が弱い項目となっている。IRT 困難度パラメタは 2.0 と難しい項目で、IRT 当て推量パラメタが 0.4 を上回っており、潜在特性値の低い受験者も偶然に正答を得る確率が高くなっている。特異項目の内容を見ると、この項目は、発話の最後にある「逆転劇の主役」という情報から、まず「液晶テレビ」を選び出し、その後、AV 機器業界の生き残り戦略を聞くという設定になっているが、「生き残り戦略」という部分のみ聞き取った受験者は、「逆転劇」で敗者となった「ブラウン管テレビ (選択肢 3)」を選んだ可能性がある。キーワードである「逆転劇の主役」というビジネス場面で使われる抽象的な表現が受験者には理解しにくかったことが考えられる。また、勝者である「液晶テレビ」を例にとり、「生き残り戦略」を語るという設定には無理がある。そこで、5.1.2.2 節で挙げた修正方針 1) に沿って、発話の最後にある「逆転劇の主役」

「生き残り戦略」を削除し、「その次の年は、薄型テレビがブラウン管テレビを追い抜く年となり、これからのAV機器の主役に踊り出しました。それでは、競争に勝ってAV機器の主役となったこちらの機器について、各メーカーの生産と販売の現状を詳しくお話したいと思います」というように、キーワードを複数の表現でパラフレーズするように修正した。

発話の一部を変更したことにより、修正項目の項目－特性値相関は0.237、IRT当て推量パラメタは0.308となり、改善が見られた。潜在特性値の高い受験者群（MからHH）の正答率が上昇したことから（図5-37b）、無意味に正答を得られにくくしていた要因を取り除くことができ、結果的に項目統計量が改善されたと考えられる。

(2) 項目番号：聴読解 2-6

<修正前（特異項目）>

【設問】（音声情報）

上司が外出先から電話をかけて来ました。部下はこのあと、書類のどの部分を訂正しますか。

【視覚情報】

関係各位

〇〇〇〇年〇月〇日

ABC株式会社
営業本部長 松野太郎

販売者会議のご案内

下記の通り、販売者会議を開催いたしますので、ご案内いたします。

記

日時:〇〇〇〇年×月×日 10:00~17:00
場所:ABC株式会社 大会議室
次第:

1. 開会の辞
2. 挨拶:
3. 講演:「販売戦略」
4. 報告:「販売実績について」

(昼食休憩)

5. 〇〇:××××
6. ××:〇〇〇〇
7. 閉会の辞

ABC株式会社
副社長 山田一郎

JFC銀行頭取 佐藤浩氏

ABC株式会社
営業課長 大山和男

【会話文】（音声情報）

男 2：はい，営業三課です。

男 1：あ，松野だけど。

男 2：あ，部長，お疲れ様です。

男 1：うん，ちょっと頼みたいことがあるんだけど。

男 2：はい。

男 1：あの，例の販売者会議の件なんだけど。

男 2：はい。

男 1：うん，今，JFC 銀行の頭取から連絡があって。

男 2：ああ，佐藤頭取ですね。

男 1：佐藤さんは午前中しか時間が取れないんで，あとは副頭取の山下さんが引き継ぐとのことなんだ。

男 2：あ，はい。わかりました。

男 1：それと，うちの挨拶なんだけど，やっぱり，社長にお願いすることになったから。

男 2：ああ，よかったですね。

男 1：それで，案内状を，至急，変更しておいてほしいんだけど。

男 2：はい，わかりました。じゃ，すぐに変更します。

男 1：うん，よろしく。

男 2：あ，あの，二日の午前中の講演も，山下さんがされるってことでいいですか。

男 1：いや，講演のほうはそのまま。そのあと抜けるらしいから。

男 2：わかりました。じゃ，30 分以内にお送りしますんで。

（正答：2）

＜統計情報＞

修正前（特異項目）の統計データ

項目—特性値相関0.159 IRT困難度パラメタ1.065 IRT当て推量パラメタ0.388

修正項目の統計データ

項目—特性値相関0.367 IRT困難度パラメタ1.077 IRT当て推量パラメタ0.149

設問回答率分析図

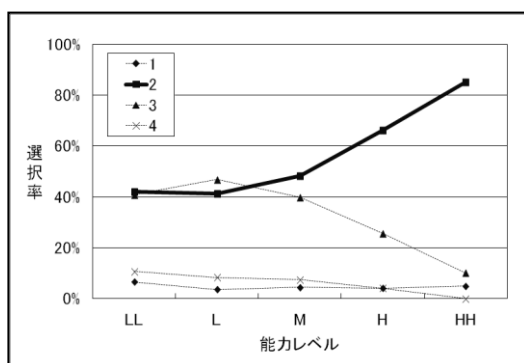


図 5-38a 聴読解 2-6 特異項目

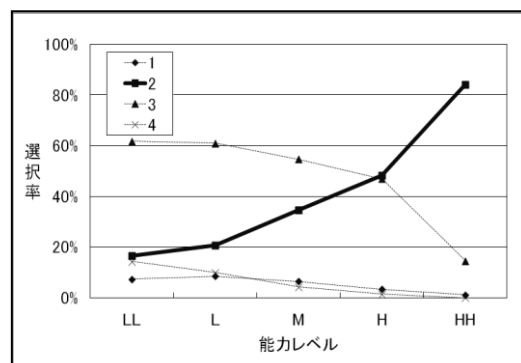


図 5-38b 聴読解 2-6 修正項目

項目特性曲線

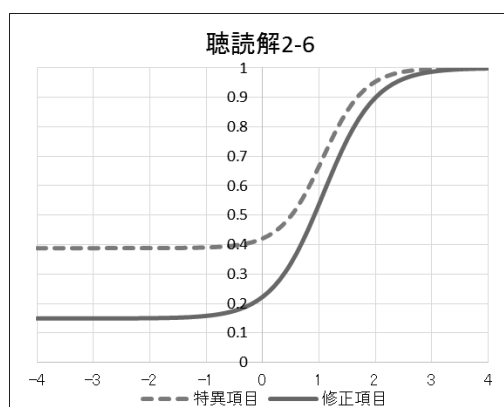


図 5-25 (再掲)

図 5-38a を見ると、最も潜在特性値の低い受験者群 (LL) の正答選択肢の選択率が、次の特性値レベル (L) の選択率を上回っており、このことが原因で項目—特性値相関が 0.159 と低くなっていると考えられる。特異項目の内容を見ると、会話の終盤に「二日の午前中」「山下さん」という発話があるが、書類の文字情報には「二日」「山下」という情報はなく、聴解情報と文字情報の整合性に欠けていることがわかる。そのため、これらの整合性を保つ修正が必要である。そこで、5.1.2.2 節で挙げた修正方針 2) に沿って、最後の「男 2 : あ、あの、二日の～」からの発話を削除し、「男 1 : あと、大山さんは予定通り。男 2 : はい、わかりました」に変更し、誤答選択肢 4 について言及するように修正した。

会話を変更したことにより、修正項目の IRT 困難度パラメタは 1.077 とほぼ変わらないまま、項目—特性値相関が 0.367 と改善された。潜在特性値の低い受験者の解答傾向が選択肢

3にも集まり（図5-38b），結果的に識別力が改善されたと考えられる。

（3）項目番号：読解 3-3

<修正前（特異項目）>

【設問】（文字情報）

取引先から次のような手紙が来ました。商品について何をしてほしいと言っていますか。

【視覚情報】

		〇〇〇〇年〇月〇日	
□□商事株式会社			
営業部長	〇〇〇〇殿		
		△△株式会社	
		販売部部长	〇〇〇〇
<p>拝啓 貴社ますます御清栄のこととお喜び申し上げます。</p> <p>さて、この度は、弊社製品『炊飯器・たけるくん』をお買い上げいただきまして、心より御礼申し上げます。</p> <p>先日も直接お話を致しました通り、おかげ様で多くの皆様にご愛好いただき、大変、ご好評をいただいております。品質に関しましても、絶対の自信を持っておりまして、弊社では、日本のみならず海外にもお勧めしております。</p> <p>今後も、同製品の品質改良には、たゆまぬ努力を続けて参る所存でございますが、同製品につきまして、使い勝手等、ユーザー各位に評価いただくと共に、ご要望等、お伺いしたいと思います。</p> <p>つきましては、別添のアンケート用紙にご記入の上、ご返送いただけましたら幸いです。</p> <p>今後とも変わらぬご支援、ご愛顧を賜りますよう宜しくお願い致します。</p> <p>取り急ぎ、お買い上げの御礼かたがたご挨拶申し上げます。</p>			
			敬具

【選択肢】（文字情報）

- | | |
|------------------|----------------|
| 1. 不良品を送り返してほしい。 | 2. 追加注文してほしい。 |
| 3. 意見を聞かせてほしい。 | 4. 顧客を紹介してほしい。 |

（正答：3）

<統計情報>

修正前（特異項目）の統計データ

項目—特性値相関 0.311 IRT 困難度パラメタ-0.789 IRT 当て推量パラメタ 0.414

修正項目の統計データ

項目—特性値相関0.445 IRT困難度パラメタ0.170 IRT当て推量パラメタ0.276

設問回答率分析図

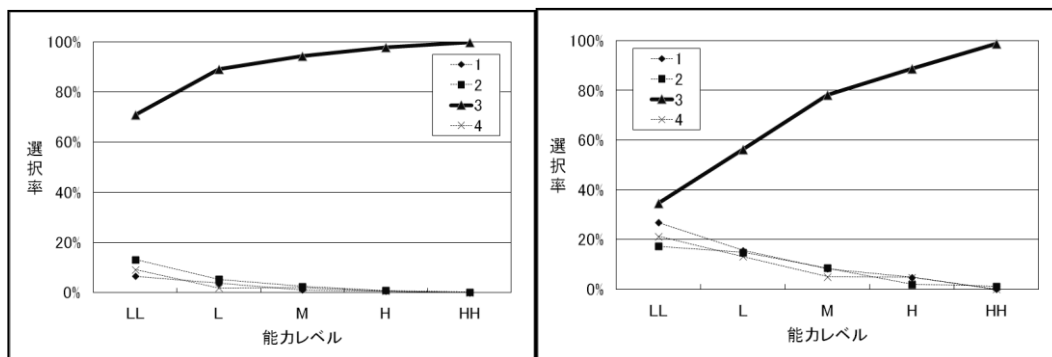


図 5-39a 読解 3-3 特異項目

図 5-39b 読解 3-3 修正項目

項目特性曲線

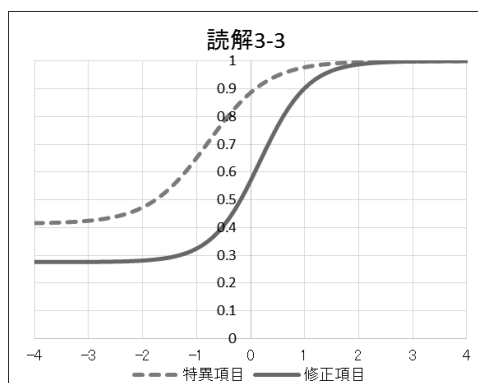


図 5-34（再掲）

図 5-39a を見ると、潜在特性値が低い受験者群でも正答率が高いため、IRT 困難度パラメタも低く、IRT 当て推量パラメタが高くなっていると考えられる。特異項目の内容を見ると、誤答選択肢 1, 2, 4 につながる語句（不良品、追加、注文、顧客等）が文中になく、

「評価」「ご要望」「アンケート」などの語からの類推で、必然的に、正答選択肢3が選ばれやすくなったと考えられる。読解テキストの文書については、消費者向けのアンケートを、販売部長の個人名で、取引先の営業部長に出すということは一般的ではないため、受験者は、文書をどのような状況で、どのような立場で読むのか当惑した可能性が考えられる。そこで、5.1.2.2節で挙げた修正方針3)に沿って、文書の発信者を「販売部販売促進課顧客サービス係」、受信者を「営業部業務課調査係」と修正して、文書を読む状況や立場を明確にした。また、選択肢をすべて調査の依頼に関するものとし、「1.調査を海外で実施してほしい 2.調査の報告をしてほしい 3.調査に協力してほしい 4.調査結果を集計してほしい」に変更した。

本文の情報の一部を変更および選択肢を文書の内容に近いものにそろえたことにより、図5-39bに示すように潜在特性値の低い受験者群(LLからM)の正答率が下がり、修正項目のIRT困難度パラメタは0.170と適度に高くなり、IRT当て推量パラメタは0.276に改善された。また、項目—特性値相関は0.445に上昇している。

(4) 項目番号：読解 2-4

<p><修正前（特異項目）></p> <p>【設問】（文字情報）</p> <p>会議の結果がわかり次第，早急に措置を（ ）必要がある。</p> <p>【選択肢】（文字情報）</p> <p>1. かかる 2. 使う 3. 操る 4. 講じる （正答：4）</p>
<p><統計情報></p> <p>修正前（特異項目）の統計データ</p> <p>項目—特性値相関 0.190 IRT困難度パラメタ 2.188 IRT当て推量パラメタ 0.198</p> <p>修正項目の統計データ</p> <p>項目—特性値相関0.233 IRT困難度パラメタ1.319 IRT当て推量パラメタ0.286</p> <p>設問回答率分析図</p>

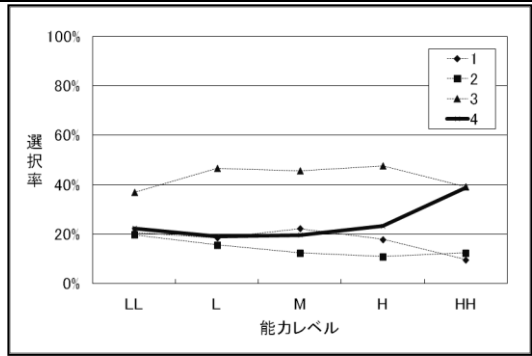


図 5-40a 読解 2-4 特異項目

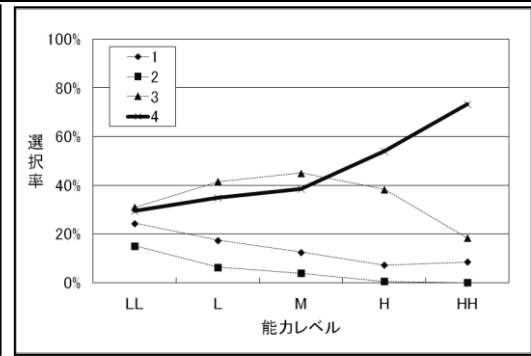


図 5-40b 読解 2-4 修正項目

項目特性曲線

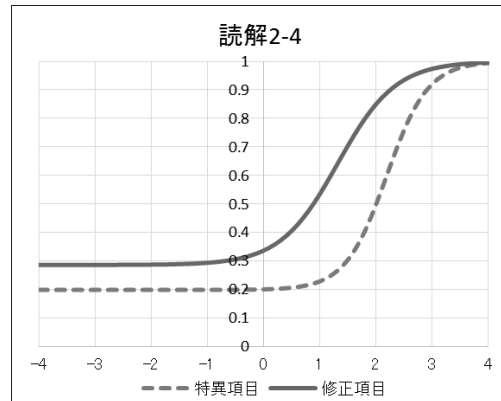


図 5-30 (再掲)

項目—特性値相関が 0.2 を下回り、受験者の日本語能力レベルを識別しにくい項目である。また、IRT 困難度パラメタが高く、難しい項目となっている。図 5-40a を見ると、潜在特性値の高い受験者は正答を得ているが、すべてのレベルにおいて誤答選択肢 3 の「操る」の選択率が高い。この語は、中国語においても、「世論」「物価」など、抽象的な語と共起することがあることから、こうした母語の干渉などが選択率の高さの要因と考えられる。また、誤答選択肢 1 の「かかる」、2 の「使う」については、初級レベルで学習する基本的な語彙のため、学習済みだと考えられるが、「措置」と「講じる」のコロケーションについては、潜在特性値の高い受験者群でも未習者が多く、したがって能力を識別する力が低かったと考えられる。そこで、5.1.2.2. で挙げた修正方針 4) に沿って、設問文の「措置」を「対策」に改変し、最も選択率の高かった誤答選択肢 3 の「操る」をひらがな表記に修正した。「操る」は実際のビジネス場面では漢字のまま使われることもあるが、ここでは中国語母語話者が中国語の「操」の意（手に持つ、従事する）の影響を受けやすい状況を配

慮して、ひらがな表記に修正した。このような対応は、個々のテスト項目の設問文や正答選択肢の内容などに即して適切に行う必要がある。

本文の表現および選択肢の表記を修正したことにより、修正項目のIRT困難度パラメタは1.319と低くなり、やや易しくなったものの、項目—特性値相関は0.233となり、項目識別力は改善された。図5-40bを見ると、誤答選択肢3は修正後も多く選ばれているが、能力レベルの上昇とともに選択率は下がる傾向にあることから、誤答選択肢も迷わし効果としては有効に機能していることがわかる。

5.3 第5章のまとめと今後の課題

本章においては、統計的に好ましくない項目特性を持つことが判明したテスト項目（特異項目）を、5.1.2.2 節に示した測定の質を高めるための方針に従って修正し、その修正された項目を実験的に実施することにより、項目特性が改善されることを示した。具体的には、特異項目 28 項目のうち 23 項目について、修正によって項目—特性値相関等の項目統計量の改善が観察され、特異項目の枠から外れた。

なお、5.1.2.2 節に示した方針 1) ～5) のうち、本章における特異項目の修正に最も多く反映させたのは、

- 1) ビジネス場面で使われる抽象的な表現が、困難度を上げ、識別力を下げる可能性が考えられるため、テスト項目のテキスト部分に複数の情報を配置し、その複数の表現から手掛かりを得られるようにパラフレーズすること。

であり、修正項目 28 項目中 7 項目で反映させた。また、

- 2) 聴読解テスト項目では、視覚情報（文字情報）と聴覚情報との整合性を保つようにすること。

は、聴読解の修正項目 7 項目中 5 項目で、

- 3) 読解テスト項目の設問文では、いつ、だれが、どこで、だれに対して、どのような状況で当該文書を書き、それを読む者は、いつ、どこで、どのような状況で読むのかなど、状況や文脈を適正に設定すること。

は、読解の修正項目 10 項目中 3 項目で、

- 4) 読解テスト項目においては、受験者の母語の干渉が識別力を低くする要因となりうるため、選択肢の作成では、たとえば表記や発音などの類似・非類似性といった対照言

語的観点から、受験者の諸母語（主に中国語、韓国語、英語を想定）を意識すること。
は、読解の修正項目 10 項目中 2 項目で、

- 5) キーワードのマッチングで正答を選べるテスト項目は、日本語能力の低い受験者も正答を得やすく、識別力が低くなるため、できるだけ避けること。

は、修正項目 28 項目中 3 項目でそれぞれ反映させた。

この修正方針の適用数に関する結果の解釈については、修正項目が 28 項目という、非常に限られた範囲内であり一般化はできないが、修正方針 1) の適用数が最も多かったことから、ビジネス場面で使われる抽象的な表現の扱いが、ビジネス日本語テストの測定精度を確保する上で重要なポイントであることが示唆される。池田（1996）は、ビジネス現場で使われる話し言葉（社内および社外での会議や打ち合わせ、電話での会話）は、ビジネス専門語彙の割合が低く、その大部分がビジネスの場以外での日常的に用いられている語彙であると述べているが、日常的に用いられる語彙であっても、使い方次第ではビジネス特有の表現になりうる。たとえば本章の例では「逆転劇の主演」「生き残り戦略」などがあるが、経験的にもこのようなビジネス表現は比較的多いと言えよう。修正方針 1) を適用した項目が多かったということは、日本語能力が高い受験者でもこのような抽象的な表現になじみがなく、意味や状況を理解しにくいということを示しているとも言える。テストにおいては、ビジネス場面で使われる抽象的な表現が聴解や読解のテキストの中で単独で用いられると、項目識別力が低くなる可能性があり、測定の質の確保という意味では好ましくないが、実際のビジネス場面では使われることが多い。したがって、外国人材に対するビジネス日本語教育においては、このような表現を意識して取り上げていくべきと言えるのではないだろうか。

修正方針 2) ～5) については、ビジネス日本語能力を測定するテストにかかわらず、日本語のテスト全般に適用可能であると考えられる。ただし、4.2.1.3 で述べたように、特に BJT においては、テストの場面や設定が現実のビジネス場面を反映しているかどうかという真正性を重視しており、受験者が課題達成のための視点をフォーカスしにくい設定の出題はしていない。すなわち、BJT の出題項目は、受験者がそのテスト項目で設定されたビジネス場面の当事者のひとりという立場で、音声を聞いたり文字情報を読んだりするため、修正方針 3) のような視点が、他の一般的な日本語能力を測定するテストと比べ、重要になると言える。これらの修正方針を、BJT 以外のビジネス日本語テストにおいてどの程度注視すべきかについては、テストが測る構成概念や仕様との兼ね合いによるが、3.3 節で

述べたビジネスコミュニケーション能力の構成要素のうち、(1)言語能力の測定が中心のテストであれば、本章と同様の方針が有効であり、かつ採用すべきであると考ええる。

ところで、表 5-1 に示した特異項目の基準は 5.1.1.1 節でも述べたように主に項目の信頼性に関わる基準であった。なぜなら、もしも 2PLM が用いられているとしたならば項目－総合点相関や項目－特性値相関は、Lord (1980) の 3 章が示すように、IRT 識別力パラメタとほぼ単調増加の関係にあること、IRT における項目情報量関数は識別力パラメタに比例すること、そして、項目情報量関数の和としてのテスト情報関数が測定の精度を表すことを総合すると、項目－総合点相関が高くなることは測定の標準誤差の減少を意味し、その意味で、改善されたテスト項目を用いて作成されるテストフォームの信頼性は向上すると考えられるからである (2.2.3 節を参照)。そしてこのことは、例えば Lord (1980) や Alderson et al. (1995) が述べているように、テストの信頼性の高さがテストの妥当性を高めるための必要条件であることを考慮すれば、今回の修正は単にテストの信頼性の改善にとどまらず、テストの妥当性をも改善させた可能性がある。しかし、その点に関しては、より妥当性に焦点を当てた今後の研究で明らかにされるべきである。本章に続く第 6 章でその一端を明らかにしたい。

本章によって、テスト項目の項目統計量を改善するための具体的な方策が示されたことは、テストを実施した後の、項目統計量が特異であったテスト項目の改善という側面だけでなく、新規にテスト項目を作成する際の方針としても有効であり、より信頼性・妥当性の高いテストの開発に資するものである。また、従来、項目統計量が不整備であったために、予備試験段階もしくは本試験実施後に、項目バンクへと追加補充されなかったテスト項目も、適切な修正・改変を加えれば、新たにテスト項目として利用できることを、定量的にも定性的にも示したもので、大規模試験におけるテスト項目作成の合理化・精緻化の一助になると考えられる。

しかし、実験テストの結果、項目統計量が改善されなかったテスト項目も 5 項目あり、これらについては別の要因が影響していると考えられるが、本章においてはそれを特定するには至らなかったため、検証の範囲を広げる必要があるということが問題点として残っている。

大規模試験において、テスト開発者は、知識や経験に基づいて入念に検討を重ねながらテスト項目を作成して出題するが、結果的に受験者の能力を適切に識別する力が低かったと判明し、いくつかのテスト項目が採点処理から除外されるという事態になることがある。

我が国の大規模試験では通常、テスト項目の事前テストを行わないため、それは本来あってはならない事態である。しかしながら、そのような状態の中では、テスト項目の統計的性質が低下する原因を突き止め、測定の質の高いテストを開発することが、テスト開発者のせめてもの責務である。池田央（2008）は、「テストが表す数値を正しく読み取り解釈するには、作るときの計画およびその事実から得られる結果の積み重ね（証拠）と、その経験をもとにした納得できる説明（構成概念妥当性）が大事」であるとし、「そのための情報収集と結果を結びつける実地訓練がまだまだ不足している」と述べている。テスト全体の信頼性、妥当性を向上させるためには、実証データを蓄積し、また、テスト開発者の知識と経験に基づく主観的な判断の両面からの検証が必要であると考ええる。次章では、ビジネス日本語テストにおける DIF について、統計的データと主観的判断の両側面からの検証を試みる。

第6章 ビジネス日本語テストにおける DIF の分析

第5章では、ビジネス日本語テストにおいて、どのような要因がテストの質に影響し、テスト項目にどのような修正を加えれば測定の質が改善されるかということを明らかにした。第5章は主にテストの信頼性に関わる側面について検証したのに対し、本章では、テストの妥当性を脅かす要因のひとつである、特異項目機能（DIF）の分析を行った。

6.1 研究目的と研究方法

6.1.1 研究課題の設定

6.1.1.1 研究背景と DIF 研究の意義

多くの人に大きな影響を与える利害関係の大きいテスト (high-stakes test) においては、すべての受験者に対して公平性が保たれているかどうかの検証が重要であることは論を俟たない。テスト開発の現場では、1.4 節で示したテストの開発過程のステップのうち、「4. テスト項目の作成」の段階で、テスト項目作成の専門家（以下、専門家）が公平性の判断を主観的に行って公平性の確保に対応している。

2.2.7 節のはじめに述べたように、ETS では、本テストを実施する前に事前テストを行い、項目分析を行うことによって DIF 項目を特定したり修正したりする作業を行っているという。我が国のテストでは事前テストを行わないため、専門家による主観的な判断がテストの公平性を保つ上で重要な役割を果たしていると考えられるが、専門家は具体的にどのようなビリーフ（信念や確信）を持っており、それによってどのような基準で判断しているのだろうか。またその判断は、実際に受験者に出題した場合に得られる統計的な結果と一致するのであろうか。1.3 節で概要を述べたように、ETS の TOEIC プログラムや、我が国におけるテスト・スタンダード（日本テスト学会（編），2007）でも、特定の受験者に不利にならないように配慮するべきであると指摘されてきたにもかかわらず、また、テスト開発現場でも常に議論されてきたにもかかわらず、これらの疑問を解決するための実証的な研究は不十分であった。

特異項目機能、すなわち DIF の研究は、1960 年代にアメリカで本格的に始まった（孫ほか，1995）。学力テストや選抜テストの設問の一部がエスニック・マイノリティーに不利になることが社会問題として認識されるようになり、テストバイアスや項目バイアスの

研究が活発に行われるようになった。そして、単に特定の集団に不利益をもたらすということ以外に、より広い意味で項目機能の違いを見出すということから、“バイアス”よりも中立的な“DIF”という用語が用いられるようになってきた (Camilli & Shepard, 1987 ; Holland & Thayer, 1988)。

ここで特に、テスト項目の成績に現象として集団差が認められる場合を DIF 項目、その原因が本来のバイアスの定義に当てはまる場合、すなわち測定の目的と対象とする概念にずれがあり、それが受験者にとって好ましくない場合をバイアス項目と定義する (田崎, 2008)。DIF がバイアスであると判断される場合、テストが測ろうとしている構成概念とは異なるものを測っていることになるため、DIF は、すなわち、テストの妥当性を脅かす要因のひとつである。

ただし、DIF がバイアスであるかどうかの判断については、Cole & Moss (1993) は「テスト利用の文脈における意図した解釈に依存する」と述べている。これは、たとえば、車のエンジンの構造について日本語で書かれた文章を読んで解答するテスト項目が、日本語の読解能力を測定することを目的として出題され、車のエンジンに関する知識の多い男性のほうが女性よりも正答を得やすいとしたら、そのテスト項目にはバイアスがあると言えるが、これがもし車の整備士の資格を得るために車のエンジンについての知識があるかどうかを測ることを目的として出題されたものであれば、そこに男女差が生じていてもバイアスとは言えない。つまり、そのテストがどのような目的で行われ、結果がどのように利用されるかによって、DIF 項目であるかバイアス項目であるかが解釈されるのである。

Angoff (1982) が「項目がバイアスを持つか否かを決めるのは人間の主観であって統計的分析はそれを補助する道具に過ぎない」と述べているように、DIF 分析では統計的手法に加えて、非統計的な論理的考察がセットで行われる必要がある (田崎, 2008)。また孫ほか (1995) は、「統計的な DIF 分析の循環的欠陥を補う上で重要な役割を果たすのが『教科分野やテスト作成の専門家の主観』である」と述べている。しかし一方で、専門家が予想する DIF 項目は、統計的な DIF 分析で検出される DIF 項目とあまり一致しないことを示す研究や、専門家は統計的 DIF よりも DIF の程度を大きく判断する傾向があるといった報告もある (Engelhard, 1990 ; Ross & Okabe, 2006)。前述したように、テスト開発の現場では、専門家が主観的にバイアスを判断するが、これらの研究結果からも専門家の経験的知見からされるその判断は、非経済的である可能性も考えられる。したがって、言語テストにおける DIF 研究の最大の意義は、バイアス項目を検出し採点対象から除外する

ことよりも、統計的 DIF 分析と主観的 DIF 分析の両面からの検証が行われることによって、専門家がテストの妥当性を高めるための実践的な示唆を得ることであると考える。

6.1.1.2 先行研究と本章の意義

米国 ETS では、1989 年以降、大学進学適性試験 SAT (Scholastic Assessment Test) のすべてのテスト項目において DIF の検証を行っているという (Burton & Burton, 1993)。日本語能力を測定する大規模試験において DIF 分析を行っているものには、酒井ほか (2007)、井上ほか (2007)、三枝 (2008) などがあり、日本語能力試験の旧 2 級、3 級レベルのテスト項目において受験者の母語別の集団を比較した DIF 分析を行っている。三枝 (2008) では、「文法」「文字」「聴解」で構成される日本語テストについて、中国語母語話者と韓国語母語話者のグループを取り上げて DIF 分析を行い、日本語能力を測定するテストにおいては「漢字に関わる項目が大きな DIF を生じさせている」ことが明らかになったが、一方で「漢字の影響に対して、試験作成者が試験において意図的に DIF を生じさせないようにコントロールすることは難しいだろう」と述べている。

日本語のテストで母語以外の下位集団における DIF 分析の報告は管見の限りでは見当たらない。また、ビジネス日本語能力を測定対象とするテストのレベルでは、DIF 分析はまだ行われていない。統計的・主観的両側面から DIF の要因を検証した研究も十分とは言えない。そこで本章では、BJT のテスト項目において DIF 分析を行う。分析は、受験者の解答データによる統計的な結果とあわせて、専門家による主観的な分析も行い、比較検証する。

これまで述べてきたように、日本企業による外国人材の採用が活発に行われている昨今において、ビジネス日本語テストの重要性はますます高まっている。孫ほか (1995) によると、DIF の検証によって、テストの利用目的や測定対象をテスト開発者に意識させ、それが厳しい妥当性検証を促す機会になっているという。本章で得られる結果がビジネス日本語テストをはじめとする日本語大規模試験開発の現場に与える示唆は多い。また、DIF の検証によって特定の下位集団におけるビジネス日本語学習の特徴を明らかにすることができ、ビジネス日本語教育の実践にも資すると考えられる。

6.1.2 研究目的

前節までに述べたように、テスト項目の難易度は、受験者をその属性により分類して得られる下位集団ごとに変化することがある。一般に、下位集団ごとに観測される項目特性の差異は DIF と呼ばれるが、ビジネス日本語テストにおいては性別（男・女）や居住地（日本国内・国外）による DIF は、一部の受験者にとって不利益をこうむる DIF すなわちバイアスである可能性が高い。本章は、このような下位集団に着目して DIF 分析を行う。具体的には、以下の 3 つの課題を設定し、研究を行った。

- (1) どのような項目で性別、居住地の DIF が検出されるか。(統計的 DIF 分析: Study1)
- (2) テスト項目作成の専門家は、性別、居住地の DIF およびバイアスに対し、どのようなビリーフを持っているか。(主観的 DIF 分析: Study2)
- (3) 統計的な DIF と主観的な DIF は一致するか。一致するとすれば、どのような点が一致しているか。(統計的 DIF 分析と主観的 DIF 分析の比較: Study3)

6.1.3 研究方法

6.1.3.1 テストデザインと受験者の概要

本章の調査に使用したビジネス日本語テストの項目は、第 4 章で使用した、BJT 模擬テスト (4.1.2.2 節を参照)、そして第 5 章で使用した実験テストフォーム (共通項目および修正項目で構成される) のテスト項目 (5.1.2.2 節を参照) である。ここでは、後者を test1、前者を test2 とし、表 6-1 にそれぞれ聴解、聴読解、読解の合計項目数や信頼性係数 α を示した。分析対象項目は合計 160 項目である。テストの有効回答数¹³は、test1 が 1127 名、test2 が 1317 名である。2 つのテストそれぞれの受験者の属性 (年齢や国籍等) はほぼ同じであり、そのほとんどがビジネス経験の浅い学生である (表 6-2)。test1 および test2 の尺度得点の分布を図 6-1、6-2 に示した。test1 と test2 の共通受験者 64 名について尺度得点の相関係数を算出したところ、0.795 であり、2 つのテストの困難度は似通っているものと考えられる。なお、test1 と test2 それぞれの尺度の一次元性 (2.2.6 節を参照) を確認するため、項目間テトラコリック相関行列の固有値を求めたところ (図 6-3、6-4)、どちらも第一固有値が第二固有値以下と比べて際立って高くなっていることから、尺度に一次元を仮定することに問題はないと考え、分析を行った。

¹³ 両テストとも、聴解、聴読解、読解の順で出題されるが、読解の途中で解答を中止した受験者のデータは分析に含めなかった。

表 6-1 テストデザイン

	test1 (item01-60) $\alpha=.849$	test2 (item61-160) $\alpha=.932$
受験者群 1(1127)	60 項目 (聴解 21, 聴読解 18, 読解 21)	
受験者群 2(1317)		100 項目 (聴解 35, 聴読解 30, 読解 35)

表 6-2 受験者の内訳

総数	2,444 (test1 : 1,127, test2 : 1,317)	
性別	test1:男性 276 名, 女性 684 名, その他(無回答) 167 名 test2:男性 460 名, 女性 699 名, その他(無回答) 158 名	
居住地	test1:日本国内 440 名, 国外 687 名 test2:日本国内 894 名, 国外 423 名	
所属	国内外の大学および日本語学校の日本語学習者	
国籍	男性	中国(台湾含む) (457 名) 韓国 (148 名) タイ (15 名) アメリカ, フランス, その他 (116 名)
	女性	中国(台湾含む) (883 名) 韓国 (299 名) タイ (70 名) アメリカ, フランス, その他 (131 名)
	国内	中国(台湾含む) (936 名) 韓国 (113 名) タイ (10 名) アメリカ, フランス, その他 (275 名)
	国外	中国(台湾含む) (522 名) 韓国 (367 名) タイ (80 名) アメリカ, フランス, その他 (141 名)

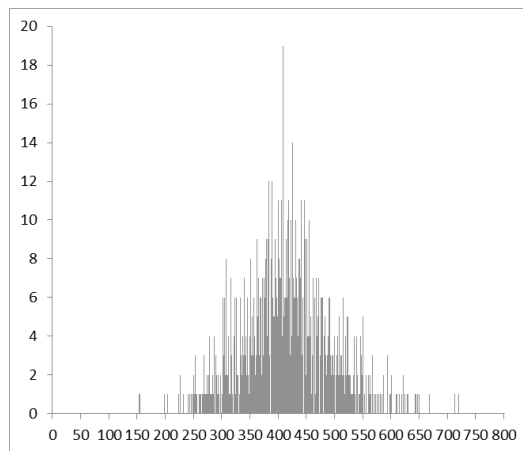


図 6-1 test1 の尺度得点分布

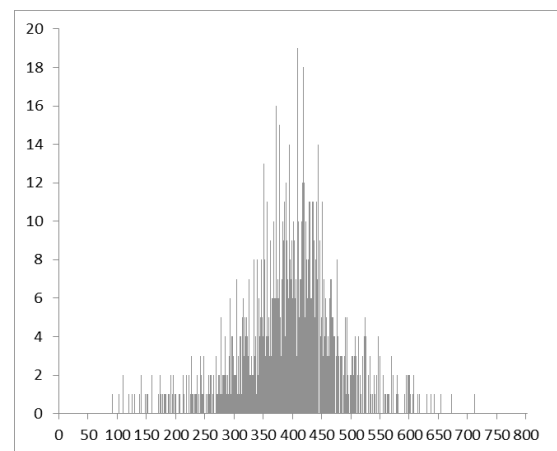


図 6-2 test2 の尺度得点分布

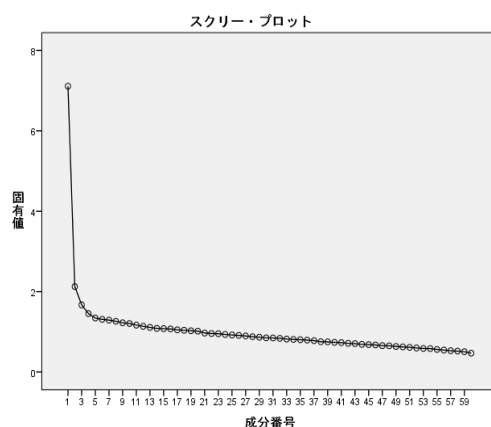


図 6-3 test1 のスクリープロット

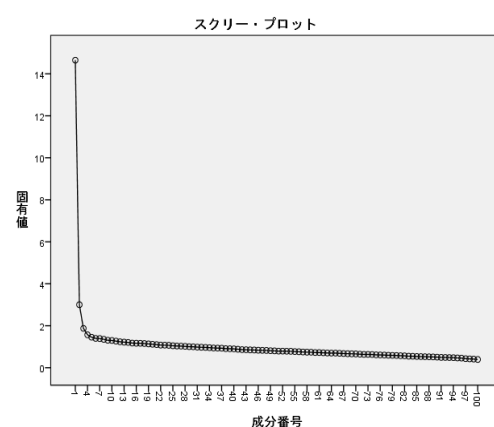


図 6-4 test2 のスクリープロット

6.1.3.2 統計的 DIF 分析 (Study1)

統計的 DIF 分析の方法はいくつかある（熊谷ほか，2003；熊谷，2012）。各方法・指標はそれぞれに異なる特徴を持っているものの，DIF を検出する方法として決定的なものではなく，複数の方法・指標を併用することが望ましいとされている（渡辺ほか，1999）。本章では，1) Mantel-Haenszel 検定法 2) 分散分析法 3) IRT 法の 3 つの手法を用い，どれかひとつの手法で DIF が検出されれば，それを DIF 項目とした。性別による DIF 分析では，女性の受験者を焦点集団，男性の受験者を参照集団とし，居住地による DIF 分析では，日本国内の受験者を焦点集団，国外の受験者を参照集団としてそれぞれ検定を行った。

1) Mantel-Haenszel 検定法

Mantel-Haenszel 検定法はもともと医学分野で用いられていた統計的手法を Holland & Thayer (1988) が DIF 分析に応用したもので、母集団の分布を仮定しないノン・パラメトリック法に分類される方法である。検定の手続きについては野口ほか (2014) に詳しいが、本研究では、まず受験者群を BJT の尺度得点のレベルによって 4 段階の得点グループ (J1 以上, J2, J3, J4 以下) に分けた。そして、テスト項目ごとに、得点グループ別の焦点集団と参照集団それぞれの正答数と誤答数をカウントし、下位集団と正誤の 2×2 の分割表を作成した。これらの分割表が全体として同じ傾向を示すか否かを、Mantel-Haenszel 統計量を用いて検定した。具体的には、4 段階の得点グループに共通のオッズ比 (M-H Odds Ratio) の推定値を求め、各得点グループで下位集団と正誤が独立であるという帰無仮説が支持されるかを、 χ^2 値を用いて検定した。有意水準は 5% とした。分析には、SPSS Statistics 22.0 for Windows を使用した。また本章は DIF の有無を示すための検定のみを行い、ETS で用いられている、DIF の程度・大きさを表す統計量 ΔMH は示していない。

この方法は、分析結果を分割表にまとめることから、分割表アプローチとも呼ばれる。分析の実行が容易であることやサンプルサイズの小さいデータ、歪みの強いデータにも対応可能であるものの、項目識別力が均質でない場合には誤った判断がされることがあるため、項目困難度の検証が中心になるという特徴がある (田崎, 2008)。

2) 分散分析法

分散分析 (Analysis of variance) は、データに正規分布を仮定してグループ間の平均値の差を検証する統計手法である。van de Vijver & Leung (1997) が示した分散分析法による DIF 分析は、2 要因分散分析を応用したものである。本章では、1) と同様に、まず受験者群を BJT の尺度得点のレベルによって 4 段階の得点グループ (J1 以上, J2, J3, J4 以下) に分けた。そして、下位集団 (性別または居住地) および得点グループを独立変数、項目反応を従属変数とし、分散分析を行った。有意水準は 5% とした。分析には、SPSS Statistics 22.0 for Windows を使用した。

下位集団と得点グループに交互作用があれば、それは属する下位集団によって項目の識別力が異なることを意味し、「不均一 DIF (non-uniform DIF)」を検出する。また、交互

作用は無いが下位集団に主効果があれば、それは属する下位集団によって項目の困難度が異なることを意味し、「均一 DIF (uniform DIF)」を検出する。2 つの下位集団を対象にする場合、均一 DIF は、ある能力において一方の集団の項目正答率がもう一方の集団より常に一定量だけ高い（または低い）場合に示される。それに対し不均一 DIF は、たとえば得点の低いグループでは項目正答率の差が大きいが、得点が高くなるにつれて差が縮まったり、あるいはある点で逆転したりする場合に示される。本章の分散分析法では、2 つの集団それぞれで同じ能力水準の受験者の正答率を比較した際に、一貫して一方の集団に有利または不利となっているテスト項目を対象としていることから、均一 DIF が検出された項目を DIF 項目としている。

分散分析法は、分析の実行が容易であることや均一 DIF と不均一 DIF の検証が可能であるという利点がある一方で、グループの分け方によって分析結果が異なること、また、潜在変数に対応しておらず、あくまで分析対象が観測変数になることなどが欠点である(田崎, 2008)。

3) IRT 法

田崎 (2008) によれば、IRT を用いた DIF 分析は、下位集団間で項目特性曲線がどのように異なるかを検討するものである。項目特性曲線の傾きを表す「識別力パラメタ」と位置を表す「困難度パラメタ」の集団間の差異を検証することによって、それぞれ不均一および均一 DIF の有意性を確認する。具体的には、Muraki (1999 : 222-223) が示すように、識別力パラメタと困難度パラメタそれぞれについて、二つの集団間の差を推定値の漸近分散の和の平方根を用いて標準化した値の 2 乗が、自由度 1 の χ^2 分布に従う特性を利用して検定を行う。本研究においても「識別力パラメタ」と「困難度パラメタ」の 2 つの項目パラメタを用いて分析を行った。有意水準は 5% とした。分析には、PARSCALE4.1 を用い、周辺最尤推定法によって行った (DIF 分析用プログラムコードの詳細は、巻末の資料 3 を参照されたい)。

IRT 法による DIF 分析が Mantel-Haenszel 検定法や分散分析法と異なるのは、受験者の能力値を観測変数でとらえるのではなく、潜在変数を仮定し、潜在特性と項目反応とが非線形関係で記述されることである。したがって、必ずしも両者の結果が一致するとは限らない。なお本章においては、以上の 3 つの手法のどれかひとつで DIF が検出された場合を DIF 項目とするが、IRT 法で不均一 DIF が存在する項目では、属する集団によって潜

在特性値（ここではビジネス日本語能力のレベル）と正答率の関係が変わるため、単純に集団間での困難度の違いを比べることが難しくなる（田崎，2008）ことから、IRT 法において、性別もしくは居住地のそれぞれの下位集団で不均一 DIF が検出された項目は、Mantel-Haenszel 法や分散分析法で DIF が検出された場合でも DIF 項目としなかった¹⁴。

統計的 DIF 分析を行った後、DIF が検出された要因の一端を探るため、日本国内に居住した経験のない、男性の日本語学習者 2 名および女性の日本語学習者 7 名に対し、DIF が検出された項目のみで構成したテストを新たに実施した。具体的な居住地は、中国、ミャンマー、タイである。そして、それぞれの項目で、どのようなプロセスで解答に至ったかを探るため、各人に対し 30～40 分程度のインタビュー調査を行った。なお、調査協力者である日本語学習者には、調査の趣旨が DIF の要因の特定であることや、実施したテスト項目が DIF 項目である旨は伝えず、音声情報や視覚情報（写真、イラスト、文字情報）のどのような点を読み取って（聞き取って）自身の解答に至ったか、どのような点で解答に迷ったかについて質問した。テスト項目を分析する際、受験者が当該の項目に解答するときどこからヒントを得たり、逆に誤解したりしているかという情報を、受験者自身から得ながら問題点を明らかにしたり、改善策を見出していったりすることは、実際のテスト運用で実施することは難しいが、テスト項目作成者には気が付かない重要な視点を与えてくれる場合もあり非常に有益である。

この調査のためのテストは、DIF 項目のみを選び出して構成した問題用紙と音声情報の電子ファイルを E メール等で調査協力者に送り、各自が自宅や学校等で実施した。その後のインタビュー調査はインターネット電話を通して行った。インタビュー調査で得られたコメントは調査協力者の承諾を得て文字化資料とした。なおこれ以降、この調査を「受験者調査」と呼ぶ。

6.1.3.3 主観的 DIF 分析（Study2）

主観的 DIF 分析は、以下の手順で行った。

- 1) 統計的 DIF 分析で DIF が検出されなかった項目と検出された項目で構成された実験テ

¹⁴ IRT 法において、性別の DIF では 160 項目中 11 項目、居住地の DIF では 32 項目で（それぞれ 5%水準で）不均一 DIF が検出されている。そのうち、性別の DIF では 1 項目、居住地の DIF では 4 項目が、他の 2 つの手法どちらかで DIF が検出されているが、これらを DIF 項目とはしなかった。

- ストフォームを作成する。
- 2) 専門家9名に、実験テストフォームの各項目について内容を精査した上で、性別、居住地による DIF があるかどうか、あるとすればその理由は何かを予想してもらう（「項目審査」と呼ぶ）。なお、専門家には各項目の統計的 DIF 分析の結果は伝えない。
 - 3) 2) の項目審査の内容を見ながら個別にインタビュー調査を行い、予想の理由の詳細や補足的な質問を行う。
 - 4) 2) 3) の調査の結果から、性別、居住地の DIF およびバイアスに対する専門家のビリーフを明らかにする。

なお、居住地による DIF について、主観的分析では「国内受験者のほうが正答率が高い」という予想のみを扱い、「国外受験者のほうが正答率が高い」という予想は調査の対象外としている。それは、BJT の開発はすべて日本国内で行われており、また、BJT の受験者の半数以上が国内受験者であること（調査時点）から、テスト項目作成の担当者は、少数派である国外受験者に不利にならないように意識してテスト項目を作成しており、国外受験者に有利となる可能性を想定していないためである。

6.1.3.3.1 実験テストフォーム

実験テストフォームは、統計的 DIF 分析を行った 160 項目の中から、性別または居住地による DIF が検出された項目（DIF 項目）と検出されなかった項目（DIF フリー項目）を抽出して構成した。DIF 項目は、主に IRT 法によって検出されたものであるが、Mantel-Haenszel 検定法や分散分析法のみで検出された項目も含まれている。実験テストフォームの項目数は表 6-3 の通りである。なお、DIF フリー項目については、トピックや形式、話者の性別、難易度等のバランスを考えて選んだ（巻末の資料 4 を参照）。

表 6-3 実験テストフォームの項目数

部門	セクション	DIF フリー項目	DIF 項目 (性別)	DIF 項目 (居住地)	合計
聴解	写真描写問題	3	0	3	6
	表現力問題	3	1	2	6
	聴解問題	5	3	1	9
聴読解	表現力問題	4	1	4	9
	聴読解問題	5	2	2	9

読解	語彙・文法問題	4	1	1	6
	表現力問題	3	2	1	6
	読解問題	5	1	3	9
		32	11	17	60

6.1.3.3.2 専門家

本章において主観的 DIF 分析を行う専門家とは、過去に BJT のテスト項目作成、編集、出題までのとりまとめに携わっていた方で、そのプロセスにおいてバイアスの検討を行い、出題の可否を決定する権限を持っていた。現在もほとんどの方が BJT や BJT 以外のビジネス日本語テストの開発に関わっており、ビジネス日本語テストの作成経験年数は最も短い方で約 2 年、最も長い方で約 17 年であり、平均は約 12 年であった。また、すべての専門家が過去あるいは現在までに国内外の日本語教育機関での日本語教育の経験を持っている。専門家の性別およびビジネス日本語テスト項目作成の経験年数の内訳を表 6-4 に示す。

表 6-4 専門家の内訳

専門家	性別	ビジネス日本語テスト 項目作成経験年数
A	女	10 年
B	女	2 年
C	女	10 年
D	女	11 年
E	女	16 年
F	男	12 年
G	男	14 年
H	男	4 年
I	男	17 年

6.1.3.3.3 項目審査とフォローアップ調査

専門家による主観的 DIF 分析では、内容を精査しながら、ひとつの項目につき性別および居住地の 2 つの観点での審査を依頼した。その際、「日本語能力が同じであっても、性別／居住地によって正答率に違いがあると予想されるか。」という質問に対し、性別については「男性のほうが正答率が高い（と予想する）」から「女性のほうが正答率が高い（と予想する）」の 5 段階、居住地については、「正答率に違いはない（と予想する）」から「国

内受験者のほうが正答率が高い（と予想する）」の 3 段階で、それぞれの下位集団における正答率の違いの程度を予想し、審査表に記入してもらった（図 6-5、巻末の資料 5 を参照）。

GID	部門		UID	問題名称	性別によって正答率に違いがあるかどうか 男性のほうが正答率が高い ← 正答率に違いはない → 女性のほうが正答率が高い				
					-2	-1	0	1	2
例	聴解	§ 1	例	書き方の例			○		

GID	部門		UID	問題名称	居住地によって正答率に違いがあるかどうか 正答率に違いはない ← 国内受験者のほうが正答率が高い →		
					0	1	2
例	聴解	§ 1	例	書き方の例			○

図 6-5 審査表の一部

また、予想の理由も自由記述式で回答してもらった。その後の個別インタビュー調査では、予想の理由の詳細や補足的な質問を行った。

6.2 結果

6.2.1 統計的 DIF 分析（Study1）

6.2.1.1 性別による DIF

性別による DIF が統計的に検出された項目は全 160 項目中 20 項目であった（表 6-5）。そのうち、女性に有利な項目は 11 項目、男性に有利な項目は 9 項目であった。また、6.1.3.2 で述べた 3 つの統計手法すべてで DIF が検出された項目は 3 項目である（表中に網掛けで表示）。性別による DIF が検出された項目の例は、6.2.4 節の（1）～（4）で詳述する。

表 6-5 性別による DIF 項目

#	部門	M-H 法	分散分析法	IRT 法
7	聴解	女性に有利	女性に有利	女性に有利
13	聴解	×	×	女性に有利
15	聴解	男性に有利	×	×
20	聴解	×	×	女性に有利

25	聴読解	×	×	女性に有利
26	聴読解	×	×	女性に有利
28	聴読解	×	×	女性に有利
34	聴読解	×	男性に有利	×
38	聴読解	女性に有利	×	女性に有利
52	読解	×	男性に有利	×
54	読解	男性に有利	男性に有利	男性に有利
78	聴解	女性に有利	×	×
80	聴解	×	女性に有利	×
97	聴読解	女性に有利	×	女性に有利
98	聴読解	×	×	男性に有利
120	聴読解	×	(不均一)	男性に有利
121	聴読解	男性に有利	×	男性に有利
134	読解	男性に有利	男性に有利	男性に有利
138	読解	女性に有利	×	女性に有利
142	読解	×	×	男性に有利

注：×は DIF が検出されなかったもの。

6.2.1.2 居住地による DIF

居住地による DIF が統計的に検出された項目は全 160 項目中 67 項目であった(表 6-6)。そのうち、国内受験者に有利な項目は 38 項目で、6.1.3.2 で述べた 3 つの統計手法すべてで DIF が検出された項目は国内 14 項目、国外 7 項目の計 21 項目である（表中に網掛けで表示）。なお、6.1.3.3 節のはじめに、本章で国外受験者に有利になった項目は対象外とする理由について、テスト項目作成時に国外受験者に有利になることを想定していないためと述べたが、現実的には国外受験者に有利な項目も 29 項目検出されている。これは決して少なくない数であるが、項目の内容から、国外受験者に有利になった原因あるいは国内受験者に不利になった原因を探るのは困難である。この点に関しては後で述べるように、国外受験者の解答傾向に焦点を当てた今後の研究の課題とし、本章では、国内受験者に有

利となった項目に着目し、その内実について分析を行った。国内受験者に有利な DIF が検出された項目の例は、6.2.4 節の (5) ～ (12) で詳述する。

表 6-6 居住地による DIF 項目

#	部門	M-H 法	分散分析法	IRT 法
10	聴解	国内に有利	国内に有利	×
26	聴読解	×	×	国外に有利
29	聴読解	国内に有利	×	×
30	聴読解	国内に有利	国内に有利	国内に有利
33	聴読解	×	×	国内に有利
34	聴読解	国内に有利	国内に有利	×
35	聴読解	×	×	国外に有利
38	聴読解	国外に有利	国外に有利	国外に有利
40	読解	国内に有利	×	×
41	読解	国外に有利	国外に有利	国外に有利
42	読解	国外に有利	×	国外に有利
44	読解	国外に有利	(不均一)	国外に有利
45	読解	×	国外に有利	×
46	読解	×	×	国外に有利
48	読解	×	×	国外に有利
49	読解	国外に有利	国外に有利	国外に有利
50	読解	国外に有利	×	国外に有利
52	読解	×	(不均一)	国外に有利
62	聴解	国内に有利	国内に有利	国内に有利
63	聴解	国内に有利	国内に有利	国内に有利
65	聴解	国外に有利	国外に有利	×
68	聴解	国内に有利	×	×
69	聴解	国内に有利	(不均一)	国内に有利
71	聴解	国内に有利	国内に有利	国内に有利

72	聴解	国外に有利	国外に有利	国外に有利
74	聴解	国内に有利	国内に有利	×
76	聴解	国外に有利	×	×
77	聴解	×	×	国外に有利
78	聴解	国外に有利	×	国外に有利
79	聴解	×	×	国外に有利
83	聴解	×	国内に有利	×
85	聴解	×	×	国外に有利
89	聴解	国内に有利	×	×
91	聴解	国内に有利	×	×
92	聴解	×	×	国外に有利
95	聴解	国内に有利	国内に有利	国内に有利
96	聴読解	×	×	国外に有利
97	聴読解	国外に有利	×	国外に有利
98	聴読解	国内に有利	×	国内に有利
99	聴読解	国内に有利	国内に有利	国内に有利
101	聴読解	国内に有利	×	×
104	聴読解	国内に有利	国内に有利	国内に有利
106	聴読解	国内に有利	国内に有利	国内に有利
107	聴読解	国内に有利	×	×
110	聴読解	国外に有利	×	×
111	聴読解	国内に有利	国内に有利	国内に有利
112	聴読解	国内に有利	×	×
114	聴読解	国内に有利	×	国内に有利
115	聴読解	国外に有利	国外に有利	国外に有利
116	聴読解	国内に有利	×	国内に有利
117	聴読解	国内に有利	国内に有利	国内に有利
118	聴読解	国内に有利	国内に有利	国内に有利
121	聴読解	国内に有利	×	×

123	聴読解	×	×	国外に有利
130	読解	×	(不均一)	国外に有利
134	読解	国外に有利	国外に有利	国外に有利
137	読解	国内に有利	(不均一)	国内に有利
139	読解	×	×	国外に有利
141	読解	国外に有利	国外に有利	国外に有利
146	読解	国内に有利	×	国内に有利
147	読解	国内に有利	国内に有利	国内に有利
150	読解	×	×	国内に有利
151	読解	国内に有利	×	×
154	読解	国内に有利	国内に有利	国内に有利
156	読解	国内に有利	国内に有利	国内に有利
158	読解	国内に有利	×	×
160	読解	国内に有利	国内に有利	×

6.2.2 主観的 DIF 分析 (Study2)

6.2.2.1 性別による DIF

審査表へのコメントとインタビュー調査の結果から、ビジネス日本語テストにおける性別の DIF およびバイアスに対する専門家のビリーフは以下のようにまとめられる。

- 1) BJT はビジネス場面における課題遂行能力を測定するテストであるため、性別によってテスト項目の難易度に差異は生じるとは考えにくい。
- 2) 一般的に男性のほうが遭遇する可能性の高い状況・場面に関するテスト項目では、男性受験者のほうが正答率が高いと予想される。
- 3) 一般的に女性のほうが目にしたり耳にしたりする機会が多い語彙や表現を問うテスト項目では、女性受験者のほうが正答率が高いと予想される。
- 4) ビジネス場面からやや外れた雑談などの場面で、女性および男性に特有の話題や談話形式に関する前提知識を必要とするタスクの設定されたテスト項目では、難易度に差異が生じる可能性がある。ただしそのような話題に関する知識量は個人差が大きい。

- 5) 性別によって著しく難易度に差異が生じるテスト項目の出題は基本的には不適切である。ただし、軽微な差異に関しては許容されるべきである。

実験テストフォームのテスト項目においては、専門家はほとんどすべての項目で「性別によって正答率に違いはない」と予想していた。全項目で「違いはない」と予想した専門家も2名いた。差異が予想されたのは計12項目であるが、いずれも「(正答率に)多少差があるかもしれないという程度」と述べており、明確な差異があると示した(審査表の-2あるいは2と判断した)専門家はいなかった。したがって、専門家の性別とテスト項目の性別によるDIFの予想との間に何らかの関係性を見いだすことはできなかった。

6.2.4節で挙げる項目の例の中では、(4)、(6)、(7)、(12)、(13)の5項目で、専門家は性別によるDIFを予想していたが、このうち、統計的にも性別のDIFが検出されたのは(4)のみであった。

6.2.2.2 居住地によるDIF

審査表へのコメントとインタビュー調査の結果から、ビジネス日本語テストにおける居住地のDIFおよびバイアスに対する専門家のビリーフは以下のようにまとめられる。

- 1) 日本国内に居住していたほうが遭遇することの多い状況・場面に関するテスト項目では、国内受験者のほうが正答率が高いと予想される。
- 2) カタカナ語や敬意表現をはじめ、日本国内に居住していたほうが目にしたり耳にしたりの機会が多い語彙や表現を問うテスト項目では、国内受験者のほうが正答率が高いと予想される。
- 3) 日本国内に居住していればなじみのある書式や談話形式に関するテスト項目では、国内受験者のほうが正答率が高いと予想される。
- 4) 日本では一般的だが、国外では一般的でない制度や習慣に関する知識を前提とするテスト項目では、国内受験者のほうが正答率が高いと予想される。
- 5) BJTは日本の商習慣や文化的な背景に関わる知識を前提とするテストであるため、居住地によって難易度に差異があっても直ちに不適切であるとは言えない。テスト項目で問う内容が日本語でビジネスをするために必要な知識であるなら、出題に問題はない。ただし、「日本国内で働く」ことを前提としたテストではないので日本語能力やビ

ビジネス知識に関係のない要因によって差異が生じているテスト項目の出題は避けるべきである。

居住地による正答率の差異について、すべての専門家が60項目中6～26項目(SD=8.2)で「国内受験者のほうが正答率が高い」という予想をしており、居住地による差異が予想されたのは計54項目であった。6.2.4節で挙げる項目の例の中では、(1)～(13)のすべての項目で、専門家は居住地によるDIFを予想していたが、このうち、(1)～(4)および(13)については、統計的にはDIFが検出されていない。

6.2.3 統計的DIFと主観的DIFの比較(Study3)

実験テストフォームにおいて、統計的に検出されたDIFと、専門家の予想によって見出された主観的なDIFを比較したものが表6-7である。表内の「統計的DIF」のカラムに「女性」「男性」「国内」とあるものは、その集団の統計的DIFが検出されたことを表す。また、「性別」「居住地」のカラムの数字は、各段階を予想した専門家の人数を表し、その中で太枠で囲ったセルは統計的DIFが検出されていることを表している。たとえば実験テスト番号12は、性別による統計的DIFが検出されており女性が有利な項目であるが、その差異を予想した専門家はいなかった。

これを見ると、性別による差異について、統計的DIFと主観的DIFはほとんど一致せず、主観的に見出されるDIFはごくわずかであることがわかる。6.2.2.1節にあるように、専門家はビジネス日本語テストにおいては性別による難易度の差異は基本的にはないというビリーフを持っており、ビリーフの2)～4)に挙げた、差異が生じる要因となる点は実験テストフォームの項目にはほとんど該当しないと考えていた。ただし、6.2.4節で挙げる項目例(12)のように、専門家の判断が男性に有利と考える場合と女性に有利と考える場合とに割れるものもあり、性別によるDIFの判断の方向性が専門家によって異なる可能性も示唆された。

居住地による差異については、専門家による主観的DIFは統計的DIFより多く見出されていることがわかる。それは、統計的にはDIFがない項目についても専門家は居住地による差異があると予想していることがあるためである。6.2.2.2節にあるように、居住地による差異についての専門家のビリーフは、状況・場面、ビジネス場面で使われる語彙・表現へのなじみ、習慣の違いというように、その判断の根拠は広範にわたっている。これら

のピリーフによって専門家は実際よりも敏感に DIF を見出している可能性が考えられる。

表 6-7 実験テストの統計的 DIF と主観的 DIF の比較

統計的 DIF 分 析#	実験 テス ト#	統計的 DIF	部門	性別					居住地		
				男性のほうが 正答率が高い		正答率 に違い はない	女性のほう が正答率が 高い		正答率 に違い はない	国内のほう が正答率が 高い	
				-2	-1	0	1	2	0	1	2
02	01		聴 解	0	0	9	0	0	8	1	0
62	02	国内		0	0	9	0	0	2	6	1
04	03			0	2	7	0	0	5	4	0
63	04	国内		0	1	6	2	0	2	5	2
66	05			0	1	8	0	0	5	4	0
69	06	国内		0	0	8	1	0	6	3	0
08	07			0	0	9	0	0	7	1	1
71	08	国内		0	1	8	0	0	4	5	0
73	09			0	0	9	0	0	8	1	0
74	10	国内		0	0	9	0	0	5	3	1
12	11			0	0	9	0	0	7	2	0
07	12	女性		0	0	9	0	0	8	0	1
84	13			0	0	9	0	0	8	0	1
15	14	男性		0	0	9	0	0	9	0	0
16	15			0	0	9	0	0	9	0	0
13	16	女性		0	1	8	0	0	7	2	0
21	17			0	0	9	0	0	4	4	1
95	18	国内		0	1	8	0	0	6	3	0
93	19			0	0	9	0	0	6	3	0
20	20	女性		0	0	8	1	0	7	2	0
86	21			0	0	9	0	0	3	5	1
96	22		聴 読 解	0	0	9	0	0	5	4	0
99	23	国内		0	1	8	0	0	8	1	0
98	24	国内		0	0	9	0	0	8	1	0
30	25	国内		0	0	9	0	0	5	4	0
106	26	国内		0	0	9	0	0	5	3	1
100	27			0	0	9	0	0	9	0	0
26	28	女性		0	0	8	1	0	7	2	0
105	29			0	0	9	0	0	7	1	1
28	30			0	0	9	0	0	3	5	1
36	31			0	0	9	0	0	7	2	0

121	32	男性	読 解	0	1	8	0	0	4	4	1
123	33			0	0	9	0	0	3	5	1
38	34	女性		0	0	9	0	0	6	2	1
122	35			0	0	9	0	0	4	5	0
111	36	国内		0	0	9	0	0	9	0	0
124	37			0	0	9	0	0	8	1	0
118	38	国内		0	0	9	0	0	6	3	0
119	39			0	0	9	0	0	7	2	0
126	40			0	0	9	0	0	8	1	0
40	41	国内		0	0	9	0	0	7	2	0
131	42			0	0	9	0	0	6	2	1
42	43			0	0	9	0	0	7	2	0
43	44			0	0	9	0	0	8	1	0
134	45	男性		0	0	9	0	0	8	1	0
136	46			0	1	8	0	0	2	4	3
138	47	女性		0	0	9	0	0	7	2	0
47	48			0	0	9	0	0	3	4	2
137	49	国内		0	0	9	0	0	9	0	0
51	50			0	0	9	0	0	8	1	0
142	51	男性		0	0	9	0	0	7	2	0
153	52			0	0	9	0	0	7	1	1
146	53	国内		0	0	9	0	0	7	2	0
159	54			0	0	9	0	0	7	2	0
54	55	男性		0	0	9	0	0	8	1	0
58	56			0	0	9	0	0	9	0	0
156	57	国内		0	0	9	0	0	7	2	0
59	58			0	0	9	0	0	6	3	0
147	59	国内		0	0	9	0	0	8	1	0
60	60			0	0	9	0	0	7	2	0

統計的 DIF と主観的 DIF があまり一致しないという結果は、先行研究の結果を支持するものであったが、ビジネス日本語テストにおいては、専門家は性別による DIF を見出すことはほとんどなく、居住地による DIF は実際以上に多く見出す傾向があるということが明らかになった。ただし、統計的 DIF 分析は検定に依拠したものであり、サンプルサイズ（標本サイズ）が大きくなればなるほど、実質的な差がなかった場合でも p 値は小さくなり、統計的に有意であるという結果になりやすいという問題がある（水本ほか，2011）。本章においても、受験者数が多かったために、検定の結果が有意となった可能性が考えら

れる。統計的 DIF と主観的 DIF の分析の結果が一致しないという結論を導くためには、そのような統計的検定の効果量と検定力の問題について十分に考慮する必要がある。

なお、6.1.3.2 節で述べた通り、本章では 3 つの手法を用いて統計的 DIF 分析を行ったが、どの手法でも検定の有意水準を 5%としている（表 6-7 の「統計的 DIF」も 5%水準としたときに検出されたものである）。この有意水準を 3%, 1%とすることによって、検出される DIF 項目の数は少なくなるが、5%水準では IRT 法が最も多く DIF を検出し、次に Mantel-Haenszel 法、分散分析法と続いた（表 6-8）。3%水準でも同様であるが、1%水準とすると、IRT 法と Mantel-Haenszel 法の検出数が同じか逆転する傾向があった。

表 6-8 統計的 DIF（分散分析法, IRT 法については均一 DIF）の検出項目数¹⁵

	Mantel-Haenszel 法		分散分析法		IRT 法	
	性別	居住地	性別	居住地	性別	居住地
5%水準	12	60	8	32	18	62
3%水準	8	51	3	23	12	54
1%水準	4	44	1	16	4	37

サンプルサイズが大きく統計的に有意な差を検出しやすいということを前提に、専門家が検出す主観的な DIF との一致という観点で考えれば、下位集団ごとに有意水準を決定することも可能ではあるが、実際には主観的 DIF 分析の結果が正しいとは限らない。今後は、DIF の影響度合いを示す指標（たとえば、Mantel-Haenszel 法における ΔMH や IRT 法における困難度パラメタの差など）を効果量とし、その効果量を用いて統計的分析と主観的分析の結果の一致不一致について検討する必要があるだろう。

6.2.3.1 統計的 DIF と主観的 DIF が一致する項目

表 6-7 で、統計的 DIF と主観的 DIF が一致している項目については、具体的に専門家はどのような部分で DIF の要因を正しく指摘できているのであろうか。そこでまず、統計的に DIF が検出された要因の一端を探るため、6.1.3.2 節で述べたように、統計的 DIF 分析を行った後に、日本国内に居住した経験のない、男性の日本語学習者 2 名および女性の

¹⁵ IRT 法で不均一 DIF が検出された項目も数に含むため、表 6-5 および表 6-6 の数とは一致しない。

日本語学習者 7 名に対してインタビュー調査を実施した（受験者調査）。そして、この受験者調査で日本語学習者から得られたコメントと、専門家による主観的 DIF 分析から得られたコメントとを比較し、日本語学習者が解答に至ったプロセスと、専門家の予想がどのような点で一致しているかについて探った。6.2.4 節で詳細な項目分析の例を挙げるが、専門家のコメントと日本語学習者のコメントが一致している項目の特徴については、全体として次の 3 点にまとめられる。

- 1) 性別の DIF については、専門家による予想もごくわずかであったが、日本語学習者への調査でも、性差が影響していると思われる内容に学習者自身が言及したのは、次節に挙げる項目例の（4）の 1 項目のみであった。そのため性別の DIF については、両者がどのような点で一致しているのかを見極めることは難しい。
- 2) 居住地の DIF については、語彙レベルでの一致がいくつか見られた。次節に挙げる項目例の（9）の「ミリ」のほか、（8）では、「前売」という語彙について、また、次節では取り上げていないが、実験テスト番号 18 聴解 セクション 3 聴解問題 6（項目内容は巻末資料 4 を参照）では、「カスタマイズ」という語彙について、国内受験者に有利となるとした専門家のコメントが、日本語学習者のコメントと一致していた。これらの項目では、正答を得る上で重要なキーワードであったり、全体理解のための助けになるような語彙について、専門家は「国内にいたほうが目にしたり耳にしたりしている可能性が高い」として DIF を予想しているが、国外居住者である日本語学習者がその語彙を「聞いたことがない」「見たことがない」と述べ、正答に至っていない状況からも、専門家は語彙の知識の有無の差による DIF の要因について、適切に指摘している可能性が示唆される。これは、6.2.2.2 節で挙げた居住地の DIF およびバイアスに対する専門家のビリーフのうち、2) に該当するものであると考えられる。
- 3) 居住地の DIF について、一部の項目では、書式や制度に関するなじみがあるかどうかの影響するという専門家の予想が、日本語学習者のコメントと一致している傾向があった。たとえば次節に挙げる項目例の（10）や（11）のように、特に、短い時間で瞬時に状況を読み取り、正答を判断するテスト項目においては、専門家は「前提知識があるほうが有利」と予想しているが、国外居住者である日本語学習者が「自分の国には（この制度が）ない」「（このような書式を）自分の国では見たことがない」と述べ、正答に至っていない状況からも、専門家は書式へのなじみ、日本では一般的な制度の知識の差

による DIF の要因について、適切に指摘している可能性が示唆される。これは、6.2.2.2 節で挙げた居住地の DIF およびバイアスに対する専門家のビリーフのうち、3) や 4) に該当するものであると考えられる。

ただし、表 6-7 を見るとわかるように、実験テストフォームにおいて、統計的に DIF が検出されており、さらに専門家によって DIF が予想された項目は、性別の DIF が 11 項目中 4 項目、居住地の DIF が 17 項目中 15 項目と少ない。そのため、本章の受験者調査によって、日本語学習者から得られたコメントと、専門家による主観的 DIF 分析から得られたコメントが一致する部分の内容を類型化するのは困難であり、あくまで傾向としてとらえるべきである。

6.2.4 項目分析

6.2.1 節で統計的に DIF が検出された項目、6.2.2 節で示したビリーフによって専門家が主観的に DIF を見出した項目、および 6.2.3.1 節で統計的 DIF と主観的 DIF が一致した項目の中から、ここでまとめて項目例を挙げ、分析する。統計的 DIF が検出された項目については、女性受験者と男性受験者、または国内受験者と国外受験者の ICC、設問回答率分析図、そして受験者調査で得られたコメントを示す。また、主観的 DIF 分析の結果と専門家のコメントを記述する。

(1) 統計的DIF 分析番号 7／実験テスト番号 12 聴解 セクション2 表現力問題 6

＜女性に有利＞

【設問】（音声情報：女声）

取引先から見積もりを送ってほしいと電話がありました。送り先の担当者名を尋ねます。
何と言いますか。

【視覚情報】



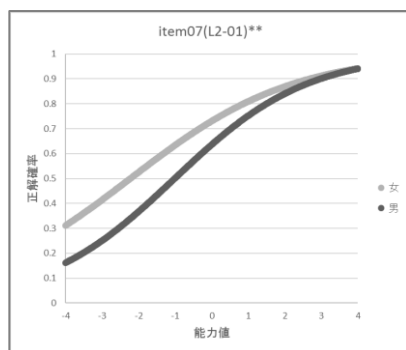
【選択肢】（音声情報：女声）

1. どこにお送りいたしましょうか。
2. だれにお送りいたしましょうか。
3. どなたにお送りいたしましょうか。
4. どっちにお送りいたしましょうか。

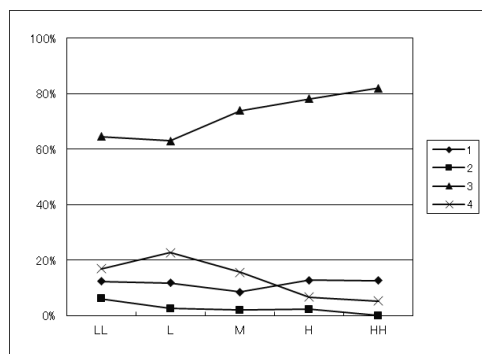
(正答：3)

統計的DIF 分析

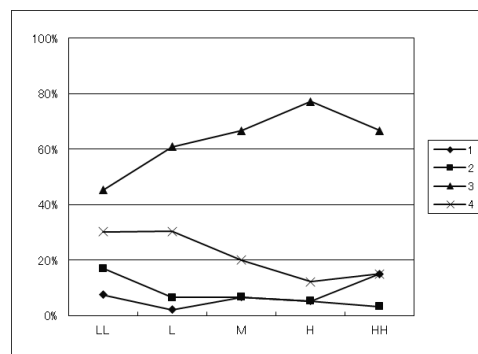
【ICC】



【設問回答率分析図】



（女性）



（男性）

受験者調査で得られたコメント（抜粋）

- ・特に難しくありません。“どなたですか” というように使うことがあります。（選択肢 3 と解答，中国・女性）
- ・“どなた” は勉強しましたが，使ったことはない。（選択肢 4 と解答，タイ・男性）
- ・見積もりを送ってほしいと言っているので，送る場所を聞きます。（選択肢 1 と解答，タイ・男性）

主観的 DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0 名	0 名	9 名	0 名	0 名

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
8 名	0 名	1 名

コメント（一部抜粋）

- ・「どなた」が国内のほうが耳にする機会が多そうなので。「どなた」が「だれ」よりも丁寧な言い方だということを日常的に知っている国内のほうが有利。（2 と予想）

このテスト項目は、写真を見ながら音声聴いて質問に答える形式で、解答時間が 5 秒と短いため、瞬時に状況を判断して適切な表現を選ぶ必要がある。ICC を見ると、潜在特性値の低い位置で項目特性曲線の乖離が見られ、女性受験者のほうが正答率が高いことがわかる。受験者調査で得られたコメントから、男性受験者は女性受験者に比べて「どなた」という不定称の人称代名詞の使い方に慣れていない可能性が示唆された。設問回答率分析図（男性）を見ると、誤答選択肢 2「だれに」や 4「どっち」の選択率が女性と比べて高いことから、男性受験者は「どなたにお送りしましょうか」というのが取引先の相手に対する使い方として適切であると瞬時に判断できなかったことが考えられる。また、潜在特性値の低い男性受験者が特に、設問にある「送り先の担当者名」を音声で聞く際に瞬時に個人宛であると理解できず、正答にたどり着かなかった可能性も考えられる。

主観的 DIF 分析では、女性のほうが正答率が高いと予想した専門家はいなかった。正答を得る上でのキーとなる「どなた」について、専門家の一人は性別による差異ではなく、居住地による差異として現れると予想していた。

(2) 統計的 DIF 分析番号 54／実験テスト番号 55 読解 セクション 3 読解問題 4¹⁶

<男性に有利>

【設問】(文字情報)

取引先から次のような文書が届きました。商品について、何をしてほしいと言っていますか。

【視覚情報】

〇〇〇〇年〇月〇日	
□□商事株式会社 営業部業務課調査係御中	△△株式会社 販売部販売促進課顧客サービス係
拝啓 貴社ますます御清栄のこととお喜び申し上げます。	
さて、この度は、弊社製品『炊飯器・たけるくん』をお買い上げいただきまして、心より御礼申し上げます。	
先日も直接お話を致しました通り、おかげ様で多くの皆様にご愛好いただき、大変、ご好評をいただいております。品質に関しましても、絶対の自信を持っておりまして、弊社では、日本のみならず海外にもお勧めしております。	
今後も、同製品の品質改良には、たゆまぬ努力を続けて参る所存でございますが、同製品につきまして、使い勝手等、ユーザー各位に評価いただくと共に、ご要望等、お伺いしたいと思います。	
つきましては、別添のアンケート用紙にご記入の上、ご返送いただけましたら幸いです。	
今後とも変わらぬご支援、ご愛顧を賜りますよう宜しくお願い致します。	
取り急ぎ、お買い上げの御礼かたがたご挨拶申し上げます。	
敬具	

【選択肢】(文字情報)

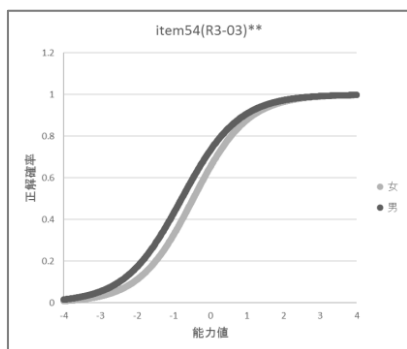
1. 調査を海外で実施してほしい
2. 調査の報告をしてほしい
3. 調査に協力してほしい
4. 調査結果を集計してほしい

(正答：3)

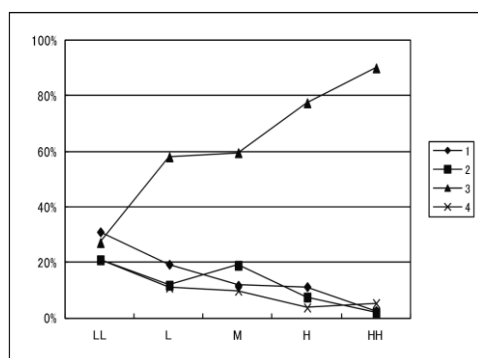
統計的 DIF 分析

¹⁶ このテスト項目は、5.2.4 節の(3)に挙げた項目の、修正項目である。

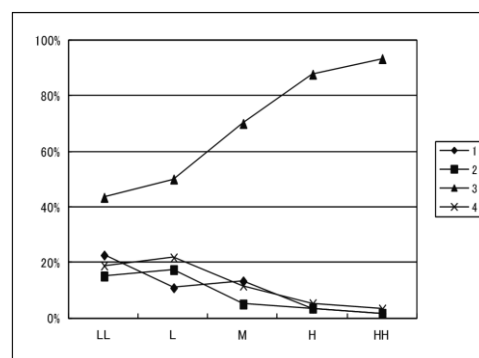
【ICC】



【設問回答率分析図】



(女性)



(男性)

受験者調査で得られたコメント（抜粋）

- ・この問題はとても難しいです。内容は長くて漢字が難しいと感じています。でも最後の「ご記入」と「ご返送」の言葉で答えが推測できました。(選択肢3と解答、タイ・女性)
- ・海外にもお勧めしていますので(選択肢1と解答、ミャンマー・女性)。
- ・アンケートをして返送します。“ユーザー”という言葉は知っています。「ユーザー登録」などと使いますね。(選択肢3と解答、ミャンマー・男性)

主観的 DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0 名	0 名	9 名	0 名	0 名

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
8 名	1 名	0 名

コメント（一部抜粋）

- ・国内居住者にはこのような、アンケートにご協力くださいという文書を目にしたことのある人もいると思う。ネット注文したりすると必ず届くので。（1 と予想）

このテスト項目は、文書を読んで、文書の発信者が何をしてほしいと言っているかを読み取る形式であり、男性受験者のほうが正答率が高い。ICC を見ると、潜在特性値の低～中程度の位置で、項目特性曲線にやや乖離が見られる。設問回答率分析図を比べると、潜在特性値が低～中程度の女性受験者は、同程度の男性受験者と比べ、誤答選択肢 1 や 2 を選択している割合が高いことがわかる。

文書中の「つきましては～」の一文から正答選択肢「調査に協力してほしい」を選ぶものだが、その対象が誰であるかは、その前のパラグラフにある「ユーザー各位に評価していただく」が指す内容を特定しなければならない。つまり、「ユーザー」という語が重要なキーワードになるわけだが、受験者調査で得られたコメントから、コンピュータやゲーム等で日常的にこの語を目にすることの多い男性受験者と比べ、女性受験者はなじみが薄いため、常識的な判断から「ユーザー」が文書の受信者（〇〇商事株式会社 営業部業務化調査係）でなくその顧客であると考えた可能性が示唆される。

主観的 DIF 分析では、男性のほうが正答率が高いと予想した専門家はいなかった。居住地の DIF を予想した専門家の一人は、視覚情報の文書形式から、6.2.2.2 節で挙げたビリーフのうち、3)「日本国内に居住していればなじみのある書式や談話形式に関するテスト項目では、国内受験者のほうが正答率が高い」を適用したものと考えられるが、実際には居住地による統計的 DIF は検出されていない。

(3) 統計的 DIF 分析番号 134／実験テスト番号 45 読解 セクション 1 語彙問題 6

<男性に有利>

【設問】(文字情報)

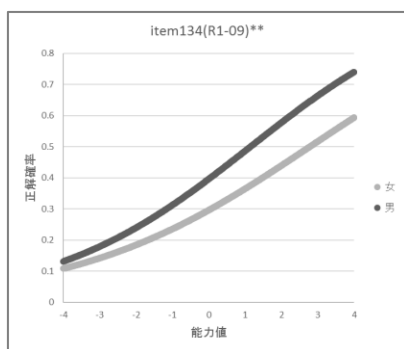
受注工事のかなりの部分が架空であり, 売り上げや利益を水増しする____ものであった。

【選択肢】(文字情報)

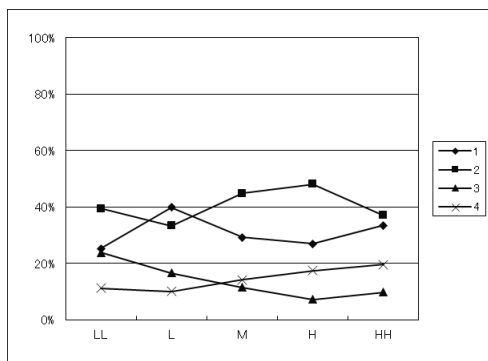
1. ための 2. ほどの 3. ことの 4. あまりの (正答: 1)

統計的 DIF 分析

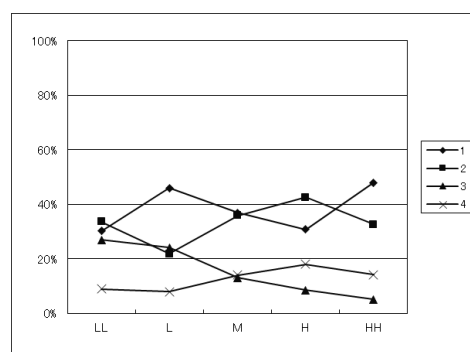
【ICC】



【設問回答率分析図】



(女性)



(男性)

受験者調査で得られたコメント (抜粋)

- ・言葉が難しくてわかりません。(選択肢 3 と解答, タイ・男性)
- ・漢字が多いので文の意味がわかりません。(選択肢 2 と解答, ミャンマー・女性)

主観的 DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0 名	0 名	9 名	0 名	0 名

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
8 名	1 名	0 名

コメント（一部抜粋）

・「水増し」は国内ではネット、新聞等で目に触れることのある表現だから、どう使われるかわかれば答えやすそうだと感じた。(1 と予想)

このテスト項目は、「受注工事」「架空」「水増し」と、比較的難易度の高い語彙が使われている文章を読んで、文章の前件と後件の関係を考えながら理由・原因の「ための」を選択する形式であり、男性受験者のほうが正答率が高い。しかし、ICC の傾きが小さいことや、設問回答率分析図の比較で、女性受験者、男性受験者ともに項目特性値が高くなるにつれて正答選択肢の選択率が高くなるという状況が見られないことから、このテスト項目は項目識別力が極めて低いことがわかる。そのため、設問回答率分析図から女性受験者が不利になった要因を探るのが困難である。また、受験者調査で得られたコメントからも、性別の DIF 検出の要因に迫ることはできなかった。

主観的 DIF 分析では、男性受験者のほうが正答率が高いと予想した専門家はいなかった。居住地の DIF を予想した専門家の一人は、6.2.2.2 節で挙げたビリーフのうち、2)「日本国内に居住していたほうが目にしたり耳にしたりする機会が多い語彙や表現を問うテスト項目では、国内受験者のほうが正答率が高い」を適用したものと考えられるが、実際には居住地による統計的 DIF は検出されていない。

(4) 統計的DIF分析番号 20／実験テスト番号 20 聴解 セクション 3 聴解問題 8¹⁷

＜女性に有利＞

【設問】（音声情報：女声）

デパートの店員が話しています。部下はこのあと、取引先に何を問い合わせますか。

【会話文】（音声情報）

女2：あ、主任。納品されたカーテン、チェック、終わりました。

男1：新作のカーテンのことね。

女2：はい。

男1：今度のカーテンも人気が出そうだね。

女2：はい。前回は約3週間で完売してますから、今回の新作も同じくらいでの完売を見越しています。

男1：新作はパステルカラーの3色模様だったよね。

女2：はい、そうです。前は1色のグラデーション模様で、それはそれで好評でした。今回の新作に関しては、すでにお客様からの問い合わせも来てます。

男1：うん、ゴージャスな感じだからな。じゃ、追加注文も考えたほうがいいかな。ボーナスの時期だし、店頭に並べる数を増やすか。

女2：じゃ、追加注文しておきますか。まだ余裕ありますかね…。

男1：じゃ、とりあえず先方に、ファックスであとどのくらいこちらにまわせるか、確認しておいてくれる？

女2：はい、わかりました。

【視覚情報】



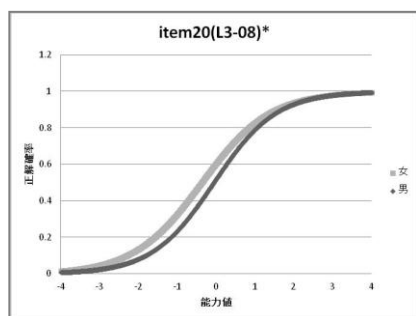
¹⁷ このテスト項目は、2.2.7.1.3 節の(2)に挙げた項目を、5.1.2.2 節の方針によって修正したものである。

【選択肢】（音声情報：女声）

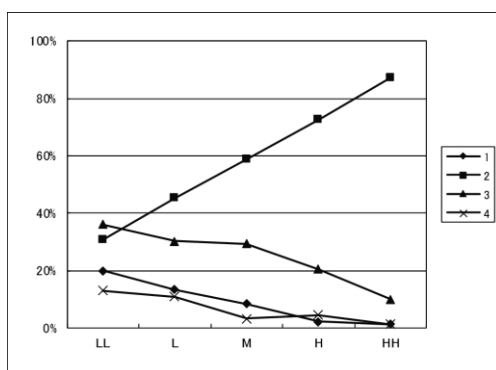
1. 商品の模様 2. 商品の在庫 3. 商品の納期 4. 商品の色 （正答：2）

統計的 DIF 分析

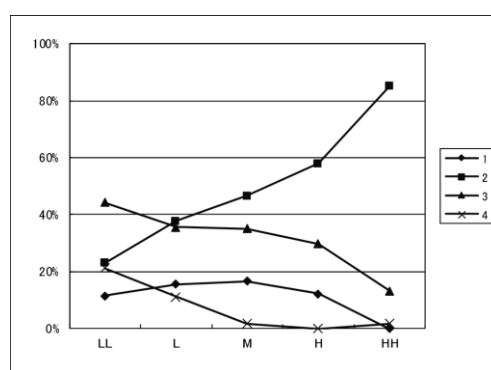
【ICC】



【設問回答率分析図】



（女性）



（男性）

受験者調査で得られたコメント（抜粋）

- ・カーテンの話はわかりましたが，色のこと？言葉がよくわかりません。カラーとか，3色とかで。（選択肢4と解答，タイ・男性）
- ・（パステルカラーという語は）はっきりとはわからないけど聞いたことがあります。うすいピンクとか？たぶんインターネットで服を見たときに。グラデーション，ゴージャスはわかります。（選択肢2と解答，中国・女性）

主観的 DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0名	0名	8名	1名	0名

コメント（一部抜粋）

- ・パステルカラー，グラデーション，ゴージャスなど，キーワードではないが話の流れから知っていたほうが有利だと思うので。女性のアパレル用語として使われているため。（1と予想）

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
7名	2名	0名

コメント（一部抜粋）

- ・カタカナ語の多用。国内のほうが日本語的な発音で耳にしている可能性が高いので。（1と予想）
- ・正答に大きな影響はないと思われるが、ボーナスの支給は日本と海外で状況が異なるため、国内受験者のほうがその知識があり、状況を理解しやすいと考えられる。ボーナス商戦などを経験的に知っているだろう。購買意欲を喚起する時期だということなど。（1と予想）

このテスト項目は、イラストを見ながら音声を聴いて質問に答える形式であり、女性受験者のほうが正答率が高い。ICCを見ると、潜在特性値の中程度の位置で、項目特性曲線に若干の乖離が見られる。設問回答率分析図を比べると、男性受験者のほうが全体的に誤答選択肢の選択率が高い。受験者調査で得られたコメントからは、男性受験者が、会話文で使われている語彙がわからなかったと話しているのに対し、女性受験者は耳にしたり目にしたりする経験がある語彙であることが示唆される。主観的 DIF 分析においても、専門家の一人がこの点を指摘しており、これは 6.2.3.1 節で述べた、性別の差異における統計的 DIF と主観的 DIF の一致が見られる項目である。なお、居住地の DIF を予想した専門家 2 名が、6.2.2.2 節で挙げたビリーフの 2)「日本国内に居住していたほうが目にしたり

耳にしたりする機会が多い語彙や表現を問うテスト項目では、国内受験者のほうが正答率が高い」および4)「日本では一般的だが、国外では一般的でない制度や習慣に関する知識を前提とするテスト項目では、国内受験者のほうが正答率が高い」を適用させ、居住地による差異を指摘しているが、実際には居住地による統計的 DIF は検出されていない。

(5) 統計的DIF分析番号62／実験テスト番号2 聴解 セクション1 写真描写問題2

<国内に有利>

【視覚情報】

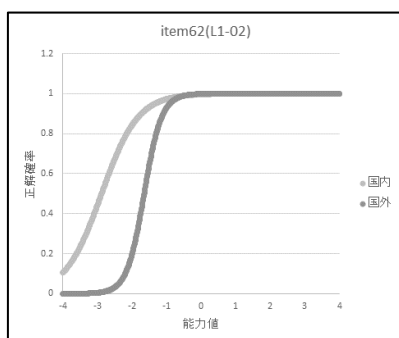


【選択肢】（音声情報：女声）

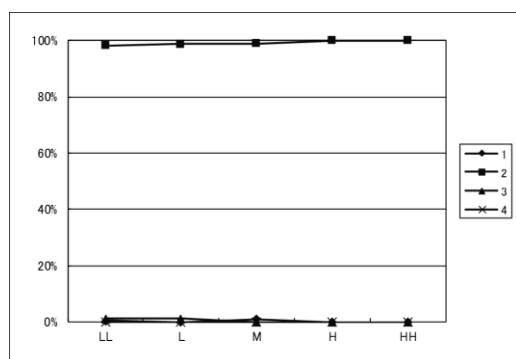
- | | |
|----------------|------------------|
| 1. 商品を回収しています。 | 2. 名刺を交換しています。 |
| 3. 品物を入荷しています。 | 4. 契約書にサインしています。 |
- （正答：2）

統計的DIF分析

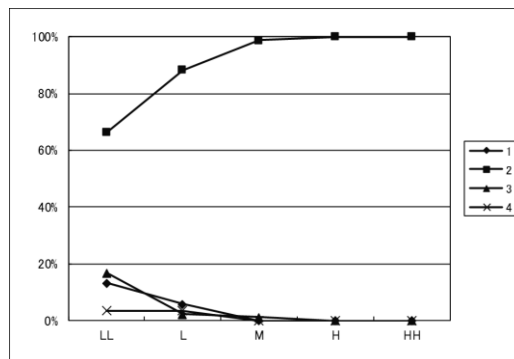
【ICC】



【設問回答率分析図】



(国内)



(国外)

受験者調査で得られたコメント（抜粋）

- ・中国でも見たことがあるのでわかりやすい。職場でよくやる。（選択肢 2 と解答，中国・女性）
- ・ミャンマーにはこのような名刺交換はないが，映画やアニメなどで見たことがある。やったこともある。（選択肢 2 と解答，ミャンマー・女性）
- ・この問題はわかりやすいです。タイ人も名刺を交換しますから。このようなスタイルではないが，近い。日本語を勉強する学生なら知っているはずです。（選択肢 2 と解答，タイ・女性）

主観的 DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0 名	0 名	9 名	0 名	0 名

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
2 名	6 名	1 名

コメント（一部抜粋）

- ・名刺交換という習慣がない国もあり，国外では目にする機会が限られる。（1 と予想，同意見他 4 名）
- ・名刺交換が一般的でない国では，写真の状況を一瞬で把握できない可能性がある。（1 と予想）
- ・国外では名刺交換の方法は特にマナーとして確立しているわけではないので，問題の写真を見ても他の場面と混同することもあるだろう。（2 と予想）

このテスト項目は，5 秒間で写真の内容を最も正しく表している文を選ぶ形式である。ICC を見ると，潜在特性値の低い位置で項目特性曲線の乖離が見られ，国内受験者のほうが正答率が高いことがわかる。設問回答率分析図では，国内受験者はどの能力水準でも 100%に近い受験者が正答を得ているが，国外受験者は潜在特性値の低い群でやや正答率

が低い。受験者調査では、全員がこの項目に正答したため、国内受験者に有利となった要因を探ることはできなかったが、主観的 DIF 分析で数名の専門家が居住地による差異を指摘しているように、このような形式で名刺交換をしない国で、かつ、日本語能力があまり高くなくドラマや映画、アニメなどでも名刺交換の場面を見たことがない受験者は、ほかの場面と勘違いする可能性がある。これは 6.2.2.2 節で挙げた専門家のビリーフのうち、1) 「日本国内に居住していたほうが遭遇することの多い状況・場面に関するテスト項目では、国内受験者のほうが正答率が高い」を適用したものと考えられる。なお、性別の DIF を予想した専門家はいなかった。

(6) 統計的DIF 分析番号 71／実験テスト番号 8 聴解 セクション2 表現力問題2

<国内に有利>

【設問】（音声情報：女声）

同僚の奥さんから、電話がかかってきました。同僚は、病気で休むと言っています。何と答えますか。

【視覚情報】



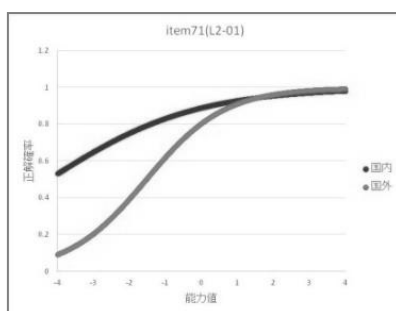
【選択肢】（音声情報：女声）

1. わかりました。お困りでしょうね。
2. わかりました。しょうがありませんね。
3. わかりました。お大事になさってください。
4. わかりました。お気をつけてください。

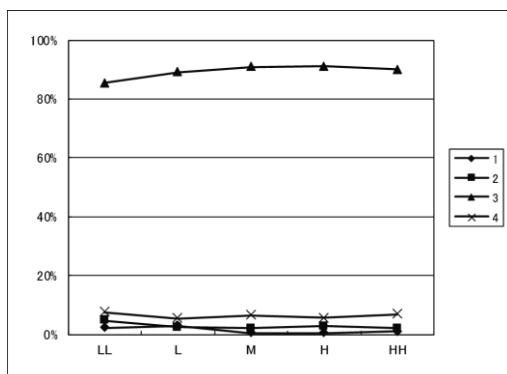
（正答：3）

統計的DIF 分析

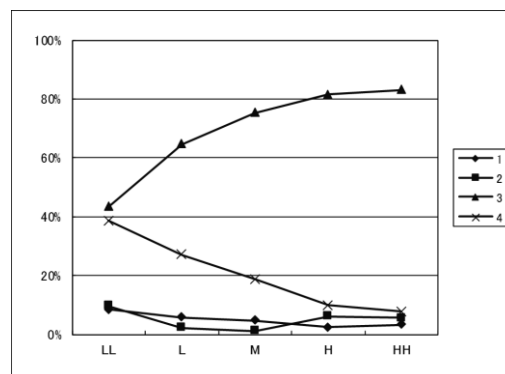
【ICC】



【設問回答率分析図】



(国内)



(国外)

受験者調査で得られたコメント（抜粋）

- ・中国でも同じような言い方がある。使ったことも言われたこともある。初級段階で勉強した。（選択肢 3 と解答，中国・女性）
- ・中国でもあるのでわかりやすい。大学のときに勉強した。初級。ただし使ったことも言われたこともない。（選択肢 3 と解答，中国・女性）
- ・同僚だから、「しょうがありません」と。丁寧に言う必要がないと思った。「お大事に」という表現は映画で見たことがある。（選択肢 2 と解答，ミャンマー・女性）

主観的 DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0 名	1 名	8 名	0 名	0 名

コメント（一部抜粋）

- ・女性の方が短い時間でいろいろと気を回すとすれば逆に悩む？（-1 と予想）

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
4 名	5 名	0 名

コメント（一部抜粋）

- ・初級の人には「お大事に」などを知らないもので、「気を付けて」を選ぶ可能性が高い。日

本にいれば初級でも「お大事に」と聞く可能性がある。日本人とコミュニケーションをとる機会が多ければ聞くことが多い。奥さんが電話をするというシチュエーションも珍しい。それも影響するかもしれない。(1 と予想)

- ・会社を欠勤することに対する意識が日本と違う国に住んでいる受験者にはわかりにくい可能性もあり。同僚の奥さんに「それはしかたありませんね」と言う国もあると思う。奥さんから電話がかかってくる状況を想像しにくいかもしれない。(1 と予想)
- ・挨拶の中では、若干一般的ではないもの。生活の中で使ってこそ覚えられる。そういう意味で国外の学習者は国内よりも定着していない可能性がある。(1 と予想)

このテスト項目は、写真を見ながら音声を聴いて質問に答える形式で、解答時間が 5 秒と短いため、瞬時に状況を判断して適切な表現を選ぶ必要がある。設問回答率分析図を見ると、国内受験者はどの能力水準でも正答率が 80%を超えているが、国外受験者の潜在特性値の低い群で正答選択肢と誤答選択肢 4 の選択率が拮抗しており、相対的に国外のほうが正答率が低くなっている。ICC を見ても、潜在特性値が低い位置で正答確率に大きな乖離があることがわかる。日本国内の日本語学習者は、学校や会社を休む連絡をすると「お大事に」と声をかけられる経験があり、学習の早い段階から表現を聞いて知っている可能性が考えられる。

受験者調査では、9 名中 8 名が正答を得ていたが、誤答選択肢を選んだ一人のコメントからは、設問文にある「同僚の奥さんから電話がかかってきました」という状況設定を聞き逃していることがわかる。このことから、国外受験者は国内受験者と比べて「お大事に」という表現を使う頻度が少ないということとは別に、病気の社員に代わってその家族が会社で電話で連絡をしてくるという状況をイメージできていない可能性も示唆される。

主観的 DIF 分析では、複数の専門家が、6.2.2.2 節で挙げたビリーフの 2)「日本国内に居住していたほうが目にしたり耳にしたりする機会が多い語彙や表現を問うテスト項目では、国内受験者のほうが正答率が高い」を適用させ、国内受験者の「お大事に」という表現に対するなじみを指摘しているが、一部には欠勤する本人以外による連絡についても言及しているものもあり、6.2.3.1 節で述べた、居住地の差異における統計的 DIF と主観的 DIF の一致が見られる項目である。なお、性別の DIF を予想した専門家の一人は、丁寧な表現を使う機会の性差についてコメントしているが、実際には性別による統計的 DIF は検出されていない。

(7) 統計的 DIF 分析番号 99／実験テスト番号 23 聴読解 セクション1 表現力問題 2

<国内に有利>

【設問】（音声情報：女声）

上司と部下が話しています。何について話していますか。

【会話文】（音声情報）

男1：よお！

男2：あ、部長。

男1：どう？今月の生産ライン，調子はどう？

男2：第1ラインの生産性が上がりました。

男1：それじゃ，今月の目標台数の達成は確実だね。

男2：はい。

【視覚情報】



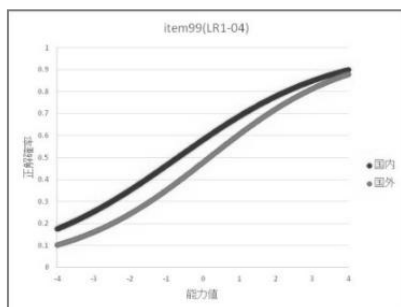
【選択肢】（文字情報）

- | | |
|------------|----------|
| 1. 体の調子 | 2. 機械の調子 |
| 3. 売り上げの実績 | 4. 生産の効率 |

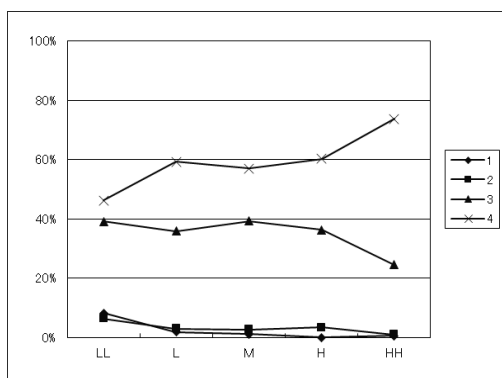
(正答：4)

統計的 DIF 分析

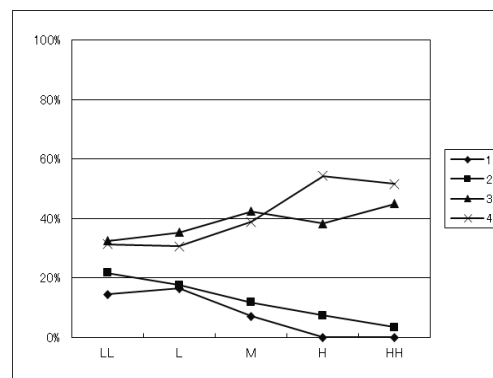
【ICC】



【設問回答率分析図】



(国内)



(国外)

受験者調査で得られたコメント (抜粋)

- ・部長，生産，目標のことを話している。だから売り上げのことと思った。（選択肢 3 と解答，ミャンマー・女性）
- ・男性同士の，上司と部下の会話は普段あまり聞かない。でも，ドラマなどで見ることもある。（選択肢 4 と解答，中国・女性）

主観的 DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0 名	1 名	8 名	0 名	0 名

コメント (一部抜粋)

- ・「調子はどう？」という表現は，やや男性的な表現という印象を受ける。自分で使ったことがなければ，肢 2 や肢 3 と迷うかもしれない。男性のほうが聞く機会が多いと思うので。（-1 と予想）

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
8 名	1 名	0 名

コメント (一部抜粋)

- ・「調子」が何を指すのかが理解できない場合は，肢 2 や肢 3 と迷うかもしれない。文脈

依存でわかっている同士の会話は、状況理解が正答確率に影響するかもしれない。(1と予想)

このテスト項目は、短い会話を聞いて、何について話しているかを文字で書かれた選択肢から選ぶ形式である。ICCを見ると全体的に国内受験者のほうが正答率が高い。設問回答率分析図を比べると、国外の潜在能力値の低い群から中程度の群で誤答選択肢 1～3 の選択率が国内の同水準より高くなっている。受験者調査で誤答選択肢を選んだ一人のコメントからは、正答を選ぶための「生産性が上がりました」という表現を特定するよりも、体調や仕事の調子を聞いているような会話全体のイメージや、「目標の達成」という語から「売り上げの実績」をイメージし、誤答選択肢を選んでいる可能性が示唆される。また、受験者のコメントからはうかがいがい知れないが、男性の上司と男性の部下による会話で、この上司の「よお！」や「調子はどう？」といったくだけた表現が、国内の学習者には身近であるが、国外の学習者にはあまり親しみがないかもしれない。

主観的 DIF 分析では、一人の専門家が居住地による差異を指摘しており、6.2.2.2 節で挙げたビリーフの 1)「日本国内に居住していたほうが遭遇することの多い状況・場面に関するテスト項目では、国内受験者のほうが正答率が高い」を適用させ、だれかが会話しているのをそばで聞くという状況は国内受験者のほうが多く遭遇するため、状況理解をしやすい可能性について述べている。性別の DIF を予想した専門家の一人は、「調子はどう？」という表現を男性のほうがよく耳にするという指摘をしているが、受験者調査で中国の女性がコメントしているように、男性同士の会話もドラマなどで耳にしていることがあるので、あまり影響を受けないのかもしれない。なお、性別による統計的 DIF は検出されていない。

(8) 統計的 DIF 分析番号 106／実験テスト番号 26 聴読解 セクション1 表現力問題 5

<国内に有利>

【設問】（音声情報：女声）

同僚と 2 人で、来月開催の見本市に行くことになりました。事前に入場券を買っておきます。2 人分でいくら必要ですか。

【視覚情報】

第19回 ゲームソフト見本市TOKYO

テーマ 「近未来」

会 期 ○○○○年○月○日（土）ー○月○日（水）

会 場 ○○○ ○○○○

開催時間 平日10:00-19:00 土・日9:30-19:00

出品者数 15カ国から175社が出品

入場料 一般（高校生以上）：1,200円（前売1,000円）
（消費税込） 小中学生：600円（前売 500円）

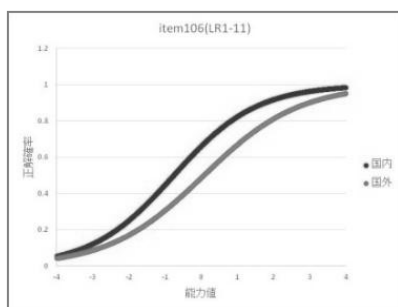
※ 当日売16時以降入場料
一般（高校生以上）：1,000円
小中学生：500円

【選択肢】（文字情報）

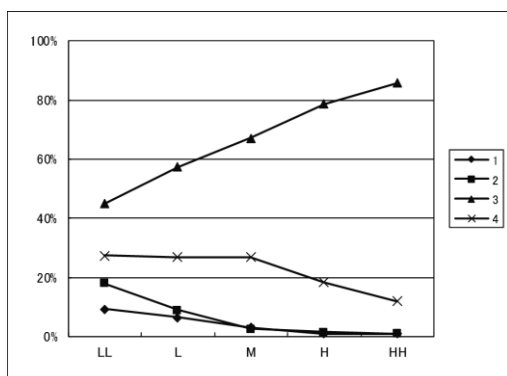
1. 1,000 円 2. 1,200 円 3. 2,000 円 4. 2,400 円 （正答：3）

統計的 DIF 分析

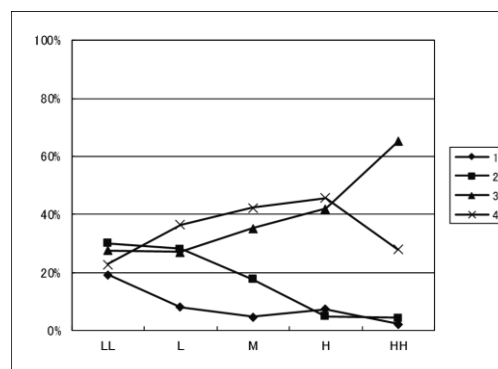
【ICC】



【設問回答率分析図】



(国内)



(国外)

受験者調査で得られたコメント（抜粋）

- ・ 同僚と行くので、小中学生（の料金）ではない。（選択肢 4 と解答，中国・女性）
- ・ 同僚と二人だから。（選択肢 4 と解答，ミャンマー・女性）
- ・ “前売” は聞いたことがない。いつ予約したら前売ですか。（タイ・男性）

主観的 DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0 名	0 名	9 名	0 名	0 名

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
5 名	3 名	1 名

コメント（一部抜粋）

- ・ 「前売」というシステムは見たことがないかもしれない。（1 と予想）
- ・ 短時間に読むためには書式を知っているほうが有利。国内居住者のほうが見慣れている可能性のある書式・表現なので。「前売」「当日売」のシステムは国によってはない可能性も。（1 と予想）
- ・ 「前売」という仕組みがない国の人には不利。知らなければ何のことかまったくわから

ないだろう。国内に居ればかなり目にする語彙。(2 と予想)

このテスト項目は、視覚情報の文書を読みながら音声を聴いて質問に答える形式で、解答時間 16 秒間で音声情報と文字情報を処理するスキミングの能力を測るものである。ICC を見ると、全体的に項目特性曲線の乖離が見られ、国内受験者のほうが正答率が高いことがわかる。受験者調査で得られたコメントから、正答を得るためのキーワードとなる「前売」という語彙について、前売りチケットや前売り券の設定が一般的でない国もあり、普段目にする事の少ない国外受験者の中には設問文の「事前に入場券を買っておく」という情報と瞬時に結び付けられなかった可能性が考えられる。設問解答分析図（国外）を見ると、誤答選択肢 4 の選択率が国内受験者と比べて高いことも、「事前を買っておく」という状況が理解できなかったことをうかがわせる。受験者の中には、「漢字を見れば何となく意味はわかるが、短い時間ですぐに判断できなかった」と回答した日本語学習者も数名いた。短い時間で音声と視覚情報を処理する形式では、前提知識があるほうが有利かもしれない。主観的 DIF 分析でも、専門家がこの 2 点を指摘しており、6.2.3.1 節で述べた、居住地の差異における統計的 DIF と主観的 DIF の一致が見られる項目である。なお、性別の DIF を予想した専門家はいなかった。

(9) 統計的 DIF 分析番号 98／実験テスト番号 24 聴読解 セクション 1 表現力問題 3

<国内に有利>

【設問】（音声情報：女声）

男の社員が、取引先と機械の部品のことについて、電話で話しています。取引先の人に何をお願いしていますか。

【会話文】（音声情報）

男 2：……ええ、ええ、そうです。なんとか、あと 1 ミリ、ええ、できれば縮めてもらって。……そうなんです。……ええ。いいですよ、延ばしても、納期は。……はい。じゃ。なんとかその線をお願いします。

【視覚情報】

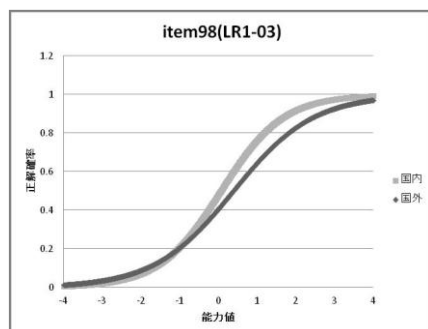


【選択肢】（文字情報）

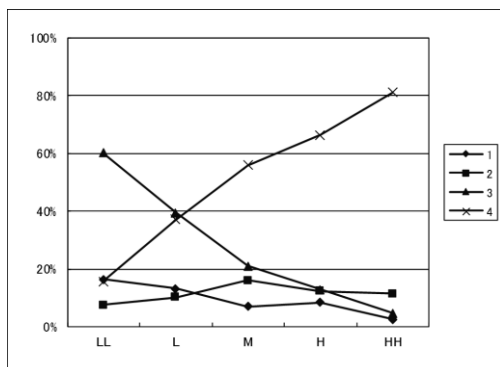
- | | |
|---------------|-----------------|
| 1. 部品の数を少なくする | 2. 部品の金額を安くする |
| 3. 部品の納期を早くする | 4. 部品の大きさを小さくする |
- （正答：4）

統計的 DIF 分析

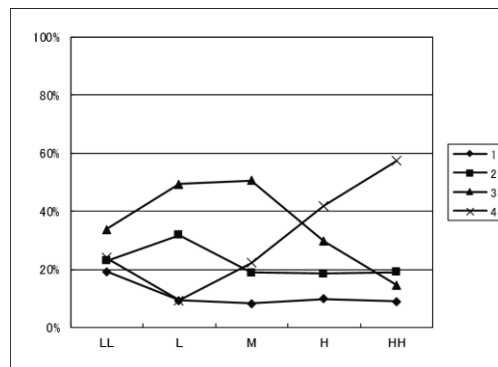
【ICC】



【設問回答率分析図】



(国内)



(国外)

受験者調査で得られたコメント（抜粋）

- ・わからない言葉たくさんありますのでわかりにくいと思います。「イチミリ」「チヂメテ」がわからなかった。ああ、ミリメートルのことですか？最初聞いたとき、わかりませんでした。（選択肢3と解答，タイ・男性）
- ・単語がわかればOK。「1ミリ」で大きさとわかった。ミリという語は難しくない。（選択肢4と解答，中国・女性）

主観的 DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0名	0名	9名	0名	0名

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
8名	1名	0名

コメント（一部抜粋）

- ・「縮める」という語がわからないと正答を選べない可能性。国内のほうがよく耳にする語彙だと思う。ミリが長さの単位であるとわかるかどうか。ミリがわからなければ，納期を縮める，つまり3を選ぶ可能性もある。（1と予想）

このテスト項目は、短い会話を聞いて、何について話しているかを文字で書かれた選択肢から選ぶ形式である。ICC を見ると潜在特性値の中～高の位置で項目特性曲線の乖離が見られ、国内受験者のほうが正答率が高いことがわかる。設問回答率分析図を比べると、国外の潜在特性値の中程度の群から高い群で、誤答選択肢 3 の選択率が国内の同水準より高くなっている。受験者調査で得られた一人のコメントから、会話文中の「ミリ」や「縮める」という語彙が聞き取れなかったために、部品の大きさに関わる正答選択肢を得ていない可能性が示唆される。正答を得ている受験者から「単語がわかれば OK」というコメントが得られていることから、「ミリ」や「縮める」が正答を得る上でのキーとなっていることがわかる。

主観的 DIF 分析でも、一人の専門家が 6.2.2.2 節で挙げたビリーフの 2)「日本国内に居住していたほうが目にしたり耳にしたりする機会が多い語彙や表現を問うテスト項目では、国内受験者のほうが正答率が高い」を適用させ、「ミリ」という語彙について、居住地による差異を指摘しており、6.2.3.1 節で述べた、居住地の差異における統計的 DIF と主観的 DIF の一致が見られる項目である。なお、性別の DIF を予想した専門家はいなかった。

(10) 統計的DIF 分析番号 30／実験テスト番号 25 聴読解 セクション1 表現力問題 4

<国内に有利>

【設問】（音声情報：女声）

次の広告は、ある会社の社員募集要項です。応募する時は、まず最初に何をどのように言っていますか。

【視覚情報】

〇〇〇〇年度 社員募集要項

募集人員：正社員 5名
年 齢：大卒・35歳位まで
給 与：委細面談
休日休暇：完全週休二日制・有給休暇あり

応 募：
以下の書類をご郵送ください。書類選考の上、面接のご連絡を致します。応募書類は返却いたしません。

- ・履歴書（要写真添付・当社所定用紙）＊
- ・応募理由書（形式自由）

＊当社所定用紙は下記採用担当宛に電話またはファックスにてご請求ください。

応募締切：〇〇年〇月〇日必着

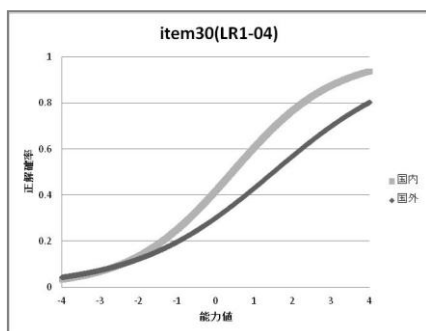
書類請求・送付先：
株式会社 ミスト
〒〇〇〇-〇〇〇〇 東京都〇〇区〇〇北〇-〇〇
管理部人材開発課 採用担当
TEL.03-〇〇〇〇-〇〇〇〇
FAX.03-△△△△-△△△△

【選択肢】（文字情報）

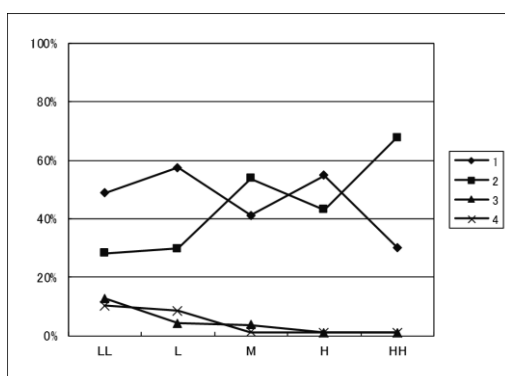
- | | | |
|--------------|----------------|--------|
| 1. 応募用紙を送る | 2. 指定の用紙を取り寄せる | |
| 3. 応募理由を知らせる | 4. 面接可能日を知らせる | (正答：2) |

統計的DIF 分析

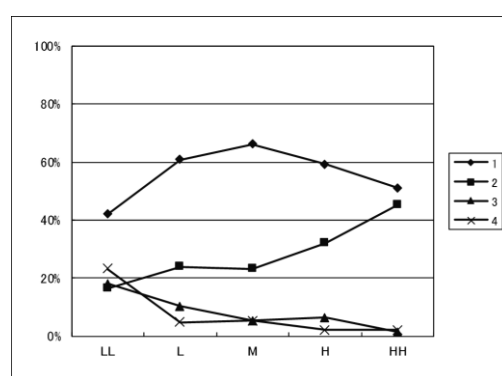
【ICC】



【設問回答率分析図】



(国内)



(国外)

受験者調査で得られたコメント（抜粋）

- ・中国の企業で指定用紙を取り寄せるところはほとんどない。自由書式で書いて送るので。日本語の募集要項を見ることはない。経験にない人は理解が難しいと思う。(選択肢 1 と解答, 中国・女性)
- ・「以下の書類をご郵送ください」で 1 を選んだ。中国では所定用紙を取り寄せることはあまりない。自分で履歴書を作って、面接のときに渡す。一次試験の時に用紙を受け取ることとはある。日本語で書かれた社員募集要項を見たことはない。見たことのない人にとって、解答は難しいかもしれない。1 と 2 の違いがわかりにくいと思う。(選択肢 1 と解答, 中国・女性)

主観的 DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0 名	0 名	9 名	0 名	0 名

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
5 名	4 名	0 名

コメント（一部抜粋）

- ・日本の従業員募集や就職に関する慣習に触れていない場合は、選択肢 2 以外を正答だと考えるかもしれない。はじめに何をするか、という理解は、よく目にするかどうかの影響しそう。(1 と予想)
- ・短い時間でかなり読まなければならないので、ある程度前提知識が必要。このような文書に遭遇している可能性の高い国内のほうがイメージしやすく有利。(1 と予想)

このテスト項目は、視覚情報の文書を読みながら音声を聴いて質問に答える形式で、解答時間 16 秒間で音声情報と文字情報を処理するスキミングの能力を測るものである。ICC を見ると潜在特性値の中～高の位置で項目特性曲線の乖離が見られ、国内受験者のほうが正答率が高いことがわかる。設問回答率分析図（国外）を見ると、国外の受験者はどの能力水準でも正答選択肢 2 よりも誤答選択肢 1 を多く選択していることがわかる。受験者調査でも誤答選択肢 1 を選んだ受験者が多く、その理由として「（自分の国では）指定用紙を取り寄せるということがない」ことを挙げている。国内受験者の中には、就職活動やアルバイトの応募などで指定用紙を取り寄せるケースを見たことや経験したことがあるのかもしれない。

主観的 DIF 分析でも、複数の専門家が 6.2.2.2 節で挙げたビリーフの 3)「日本国内に居住していればなじみのある書式や談話形式に関するテスト項目では、国内受験者のほうが正答率が高い」あるいは 4)「日本では一般的だが、国外では一般的でない制度や習慣に関する知識を前提とするテスト項目では、国内受験者のほうが正答率が高い」を適用させ、

書式へのなじみや応募書類に関する知識について居住地による差異を指摘しており、6.2.3.1 節で述べた、居住地の差異における統計的 DIF と主観的 DIF の一致が見られる項目である。なお、性別の DIF を予想した専門家はいなかった。

(11) 統計的DIF 分析番号 146／実験テスト番号 53 読解 セクション3 読解問題2

<国内に有利>

【設問】（文字情報）

次のような手紙を受け取りました。用件は何ですか。

【視覚情報】

〇〇〇〇年〇月〇日

株式会社 〇〇〇
総務部 〇〇〇〇様

株式会社 〇〇〇〇〇
営業部 〇〇〇〇

拝啓 時下ますますご盛栄のこととお慶び申し上げます。
平素は弊社発行『週刊ビジネス世界』をご購読頂き、有り難うございます。

現在、予約購読を頂いておりますが、来る4月に購読終了となります。3月31日までに更新の手続きをして頂き、一年分の購読料を前納して頂ける場合、当初の割引価格からさらに30%引きの特典がございます。この機会に是非、予約購読更新のお手続きをお薦めいたします。

『週刊ビジネス世界』は、4月より誌面リニューアルを行い、ますます充実した内容を満載し、読者の皆様にご満足いただけるビジネスマガジンとして生まれ変わります。引き続き弊誌のご利用をよろしくお願いいたします。

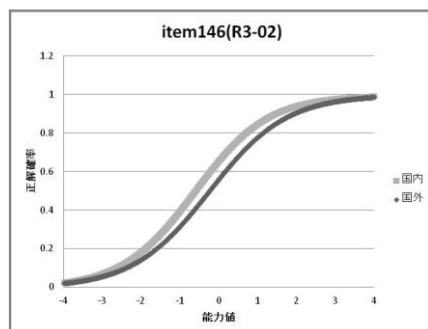
敬具

【選択肢】（文字情報）

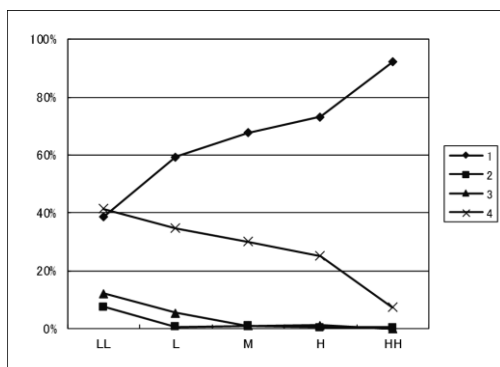
- | | |
|-------------------|-------------------|
| 1. 購読を続けてほしい | 2. 購読を終了してほしい |
| 3. 未納の購読料を支払ってほしい | 4. 新規の購読を申し込んでほしい |
- （正答：1）

統計的DIF 分析

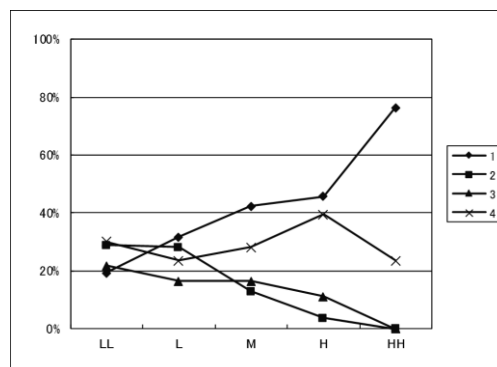
【ICC】



【設問回答率分析図】



(国内)



(国外)

受験者調査で得られたコメント（抜粋）

- ・「購読終了となります」から。雑誌の定期購読はミャンマーにはない。だからこのような知らせもない。漢字が多くてわかりにくい。（選択肢 2 と解答，ミャンマー・女性）
- ・「4 月より誌面のリニューアルを行い」「引き続き～」から選んだ。購読はしたことがない。このような文書を見たことがない。（選択肢 4 と解答，中国・女性）

主観的 DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0 名	0 名	9 名	0 名	0 名

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
7 名	2 名	0 名

コメント（一部抜粋）

- ・定期購読というシステムそのものがない国の人には不利になると思うので。国内では広告を目にする機会もあるだろう。（1 と予想）
- ・定期購読という仕組みがない国の人には不利。内容でなく，制度・システムの知識の問題。（1 と予想）

このテスト項目は、文書を読んで、文書の発信者が伝えようとしている用件を読み取る形式であり、国内受験者のほうが正答率が高い。ICC を見ると、潜在特性値の中程度の位置で、項目特性曲線に乖離が見られる。設問回答率分析図を比べると、どの能力水準でも国内受験者は国外受験者と比べ、正答選択肢を選択している割合が高いことがわかる。

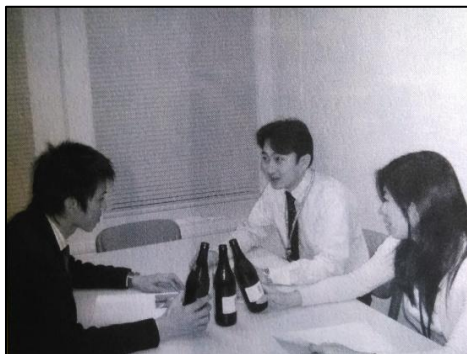
文書のあいさつ文に続く段落にある、「来る 4 月に購読終了」「この機会に是非、予約購読更新のお手続きをお薦めいたします」という文章から、正答選択肢「購読を続けてほしい」を選ぶものだが、受験者調査から得られたコメントを見ると、「購読終了」を読み取って誤答選択肢 2 を、また、最後の段落にある「誌面リニューアル」「生まれ変わります」などの表現を、誤答選択肢 4 の「新規の購読」に結び付けた受験者がいたものと考えられる。さらに、雑誌の定期購読がない国もあり、定期購読継続という文書の用件をイメージできなかった可能性も示唆される。

主観的 DIF 分析でも、複数の専門家が 6.2.2.2 節で挙げたビリーフの 4)「日本では一般的だが、国外では一般的でない制度や習慣に関する知識を前提とするテスト項目では、国内受験者のほうが正答率が高い」を適用させ、雑誌の定期購読という、国内では一般的な制度の知識について居住地による差異を指摘しており、6.2.3.1 節で述べた、居住地の差異における統計的 DIF と主観的 DIF の一致が見られる項目である。なお、性別の DIF を予想した専門家はいなかった。

(12) 統計的DIF 分析番号 63／実験テスト番号 4 聴解 セクション1 写真描写問題 4

<国内有利>

【視覚情報】



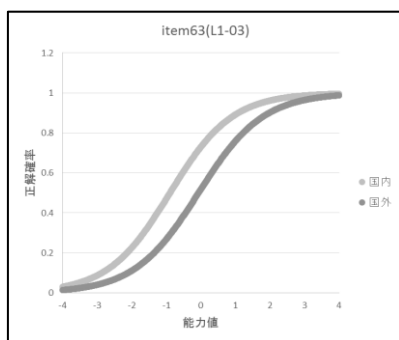
【選択肢】（音声情報：女声）

1. 社員たちが、社員食堂で食事をしています。
2. 社員たちが、レストランでパーティーを開いています。
3. 社員たちが、会議室で部内会議を開いています。
4. 社員たちが、ビールを販売しています。

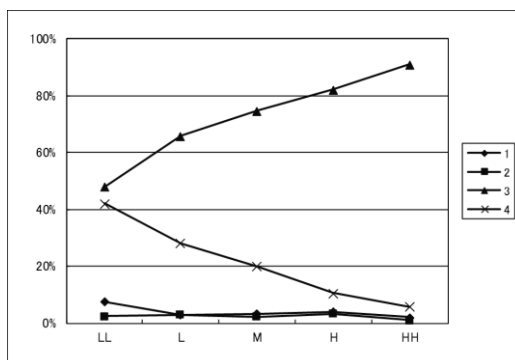
（正答：3）

統計的DIF 分析

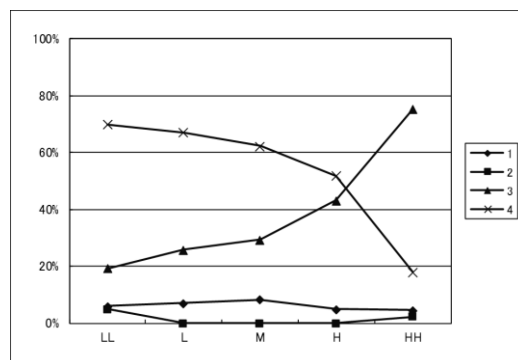
【ICC】



【設問回答率分析図】



(国内)



(国外)

受験者調査で得られたコメント（抜粋）

- ・写真は事務室のような場所で、乾杯することがわかりにくい。事務室？乾杯？（販売と聞き違い）（選択肢 4 と解答、中国・女性）
- ・スタッフが机の周りに座ってビールをほかの会社に販売していると思った。（選択肢 4 と解答、ミャンマー・女性）
- ・写真の場面はちょっとわかりにくいと思います。良く見ると、紙を見ているので、会議室かなと思った。（選択肢 3 と解答、タイ・男性）

主観的 DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0 名	1 名	6 名	2 名	0 名

コメント（一部抜粋）

- ・同僚とお酒を飲む機会が男性のほうが多いと思われることから、誤答に導かれる可能性がある。（1 と予想）
- ・男性のほうが同僚と瓶ビールで飲む機会が多そうなので勘違いする可能性が考えられる。（1 と予想）
- ・女性で、同僚との飲み会に参加する機会が少ない人だとこれは乾杯している場面だと誤解するかもしれない。（-1 と予想）

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
2 名	5 名	2 名

コメント（一部抜粋）

- ・この部屋が会議室であるかどうかはわかりにくい。このような個室の店で食事をする国も多々ある。(1 と予想)
- ・国によってはオフィスや社員食堂で飲酒する機会があるため。視覚情報の中に会議室だと判断できる情報（例えばホワイトボードがあるなど）が少ないので、「社員食堂」を選択する可能性があると思う。(1 と予想, 同意見他 3 名)
- ・飲酒の習慣がない国の人や日本の会社の社員食堂や会議室を見たことがない人だと、選択肢 3 以外を正答だと考えるかもしれない。(2 と予想)
- ・ビールを瓶のまま直で飲むことが多い国の人、パーティー会場と思う可能性がある。(1 と予想)

このテスト項目は、5 秒間で写真の内容を最も正しく表している文を選ぶ形式である。ICC を見ると全体的に国内受験者のほうが正答率が高い。設問回答率分析図を比べると、全体として国内受験者のほうが正答選択肢の選択率が高く、国外受験者は誤答選択肢 4 の選択率が高いことがわかる。受験者調査では、この写真の場所が会議室であることがわかりにくいことが多く指摘されている。また、誤答選択肢 4 の選択率が高いのは、「販売」を「ビール（で）乾杯」と聞き違えており、また場所がわかりにくいことも重なったために導かれたものと思われる。

主観的 DIF 分析ではまず居住地による差異について、6.2.2.2 節で挙げたビリーフのうち、1)「日本国内に居住していたほうが遭遇することの多い状況・場面に関するテスト項目では、国内受験者のほうが正答率が高い」を適用し、目にしたことのある国内受験者のほうが、レストランや社員食堂ではなく会議室であると判断しやすいこと、社内で飲酒する習慣、あるいは瓶のままビールを飲む習慣のある国にとっては間違いやすいことが指摘されたが、本章の受験者調査で得られたコメントとは一致していない。次に、性別による差異の予想については、「男性のほうが正答率が高い」と「女性のほうが正答率が高い」の

両方に判断が割れている。男性に不利と考える専門家は、飲酒の機会が多い男性が写真の場면을飲酒の場面と間違える可能性についてコメントしており、一方で女性に不利と考える専門家は飲酒の機会が少ない女性が逆にこれを乾杯の場面と誤解する可能性についてコメントしており、両者は逆説的なとらえ方をしていることがわかった。なお、実際には性別による統計的 DIF は検出されていない。

(13) 統計的DIF 分析番号 136／実験テスト番号 46 読解 セクション2 表現力問題 1

<DIF 非検出>

【設問】(文字情報)

A: ちょっと, 一杯, 寄っていかない? おごるよ。

B: いえ, 今日は, ぜひ私たちに_____。

【選択肢】(文字情報)

1. 払ってください 2. 払わせてください
3. 払わせてあげます 4. 払ってもらってください (正答: 2)

主観的DIF 分析

【性別】

男性のほうが正答率が高い		正答率に違いはない	女性のほうが正答率が高い	
-2	-1	0	1	2
0名	1名	8名	0名	0名

コメント (一部抜粋)

- ・A-Bは, 男-女, 男-男, 女-男, 女-女の組み合わせが考えられるが, 日本語社会では日本人女性がBの発話をする機会が男性より少なく感じられる。男性のほうがこの言葉を発している可能性が高い。(-1 と予想)

【居住地】

正答率に違いはない	国内受験者のほうが正答率が高い	
0	1	2
2名	4名	3名

コメント (一部抜粋)

- ・おごる, おごられる, に対する考え方は, 国内外で差がありそう。それが解答に影響を与える可能性がある。(1 と予想, 同意見他 3名)
- ・「ちょっと一杯」がどういう意味か理解できない場合は, どういう状況か理解しにくいかもしれない。(2 と予想)
- ・国内では「～させてください」という言い方はよく耳にすることのある表現だから。(1 と予想, 同意見他 1名)

このテスト項目は、文字で書かれた短い会話を読んで、そのあとに続く表現を選択する形式で、統計的 DIF は性別、居住地ともに検出されていない。談話形式から、会話文の A は上司や先輩、B は部下や後輩であることがわかり、B が A に対して、支払いをさせてほしいと頼む場面である。A との関係や場面から、使役形を使った「払わせてください」が正答である。

統計的には DIF が検出されていないが、主観的 DIF 分析では、専門家は性別、居住地ともに差異を予想している。まず性別の DIF について専門家の一人は、この会話文で B の立場になるのが男性のほうが多いという理由で、男性のほうが正答率が高いと予想していた。居住地の DIF では、会話文の「ちょっと一杯」などから、A がお酒を飲むことを誘っていることがわかるかどうか、あるいは上司・先輩が目下の人に対して飲食代をおごるという設定が必ずしも一般的でない国もあるという理由で、国内受験者のほうが正答率が高いと予想していた。また正答選択肢の言語形式についても、依頼の場面で使役表現がよく使われる国内にいるほうがなじみがあると予想していた。

6.3 第 6 章のまとめと今後の課題

本章では、ビジネス日本語テストにおける性別と居住地（日本国内と国外）を下位集団とする DIF の分析を行った。はじめに、ビジネス日本語テスト 160 項目を資料として統計的な DIF 分析を行った。分析には Mantel-Heanszel 法、分散分析法、IRT 法を用いた。その結果、性別の DIF は 20 項目、居住地の DIF は 67 項目が検出された。3 つの統計手法すべてで検出されたのは、性別の DIF が 3 項目、居住地の DIF が 21 項目であった。

次に、統計的に DIF が検出された項目と検出されなかった項目を含む 60 項目から構成されるテストを用いて、ビジネス日本語テストの項目作成の専門家 9 名（女性 5 名、男性 4 名）に対し、主観的 DIF 分析を行った。インタビュー調査では、どのような点で DIF やバイアスを判断しているかを尋ねた。インタビュー調査で、性別においては 5 つ、居住地においても 5 つのビリーフが見出された。

受験者の実際の解答に基づく統計的な DIF 分析と、専門家の経験的知見に基づく主観的な DIF 分析とを比較したところ、あまり一致していないことがわかった。専門家は性別による DIF をあまり見出さず、また、居住地による DIF は実際よりも多く見出す傾向があった。統計的 DIF と主観的 DIF が一致している項目について、国外に居住する男性および女性の日本語学習者の受験後のコメントと、専門家による主観的 DIF 分析から得られた

コメントとを比較したところ、居住地の DIF で語彙レベルでの一致がいくつか見られた。

以上が本章で得られた結果であるが、酒井ほか（2007）は「日本語能力を測定する試験における DIF 研究の意義として、

- 1) 特定の下位集団における日本語学習上の特徴を明らかにする、
- 2) 問題項目を作成する際に、DIF が生じ難くするためにはどのような配慮が必要かを明らかにでき、テストの妥当性を高めることに貢献できる、
- 3) 特定の下位集団に不利になる項目バイアスを除き、テストの公平性を保つ、

などがある」としている。本章では、ビジネス日本語テストにおける DIF の分析を行い、同じ能力水準であるにもかかわらず、性別、居住地によって難易度に差異がある項目を検出した。検出されたテスト項目を分析することによって、たとえば国外の日本語学習者へのビジネス日本語教育ではどのような点に配慮する必要があるかなど、教育上の留意点を把握することができる。

大規模試験において、テスト項目作成の専門家は知識や経験に基づいて入念に検討を重ねながらテスト項目を作成して出題しており、出題後には項目分析を行って測定の質を確認する作業も欠かせない。しかし、日本のテストにおいては、テストの内容が受験者の特定の集団に不利になっていないかどうかという項目バイアスの議論がほとんどされていない。そのような状態の中で本章の結果は、テスト項目のどのようなファクターが難易度の集団差に影響するのかという点を明らかにするための基礎資料となりうる。また本章は、性別、居住地という下位集団で難易度の差異が生じないようにするために、専門家がどのような経験的知見からテストの妥当性を確保しているのかを明らかにすることができた。しかし一方で、専門家の主観によって見出される DIF は統計的に検出される DIF とはあまり一致しないという、先行研究を支持する結果も得られた。このことは、専門家の持つ DIF およびバイアスに対するビリーフの一部が、実際の差異の原因を内包していないこと、あるいは専門家がこれらのビリーフを実際以上に広範に適用させ、潜在的に公平なテスト項目を削除してしまうという可能性を示唆している。

ただし、本章において行った統計的 DIF 分析は、DIF が検出されるかどうかのみを示したものであり、また、6.2.3 節で述べたように、統計的検定と効果量の問題を考慮すると、DIF が検出されたテスト項目を直ちに出題に不適切であると断定するものではない。たとえば BJT に限らず、6.2.2.2 節で挙げた居住地の DIF に対する専門家のビリーフ 5) に該当するようなテスト項目については、日本の商習慣や文化的知識を背景としたビジネス活

動は現実に行われており、そのような知識はビジネス日本語能力の一部と考えれば、出題に不適切であるとは言えないだろう。ただし、「日本国内で働く」ことを前提とするテストでないのであれば、日本語能力やビジネス知識と無関係の理由による DIF はバイアスである可能性が考えられる。すなわち、統計的に検出された DIF がバイアス項目であるかどうかは、Angoff (1982) が述べたように、人間の主観であり、統計的 DIF はバイアス項目を判断する際の必要条件に過ぎないを考える。したがって、統計的な分析と主観的な分析は、相互補完的に実施することが望ましいが、2 つの観点のどちらをどの程度重視するかについては、受験者の数や属性、また、専門家のテスト観やビジネス日本語能力のとらえ方などによって異なってくると言えよう。本章で扱った BJT 以外のビジネス日本語テストを開発・運用する際は、その点について十分議論しておくべきではないだろうか。

一方で、統計的 DIF と主観的 DIF が一致したテスト項目については、その DIF が生じた要因を専門家が正しく指摘できているかどうかを確かめるには、第 5 章に示した方法が有効であると考えられる。第 5 章では、項目分析の結果、項目統計量が許容範囲を超えていると判断されたテスト項目に関して、その原因を専門家間で議論し修正することによって、それらの項目が改善されることを示したが、DIF 項目に関しては、専門家のビリーフにしたがって DIF の要因を取り除く形で修正した項目を受験者に出題して、DIF が克服されたかを確かめる必要があるであろう。また、本章の受験者調査で得られた日本語学習者の解答プロセスに関する具体的なコメントは、テストの開発者側である専門家が気づかなかった多くの点を指摘した。これは、DIF 項目において DIF が検出された要因を特定し、改善の方策を示すための糸口となる。今後はより公平性の高いテスト項目作成の具体的な方策を得るため、実験を通じた実証的な研究が必要であると考えている。

また本章では、鈴木ほか (2008) や三枝 (2008) などが報告している、国籍 (母語) 別の下位集団による DIF 分析は行わなかった。その理由として、本章における受験者の国籍 (母語) は幅広いため、下位集団ごとのサンプルサイズが小さくなることが挙げられる。さらに、漢字圏受験者と非漢字圏受験者の比較となる場合は、DIF の原因が漢字の知識であることが予想されるが、漢字を使って書かれたビジネス文書やメールを読んだり書いたりするなど、ビジネス日本語能力に日本語の漢字の知識は含まれるものであるので、バイアスの議論にはならないと考えたためである。しかし今後は、バイアス項目の特定という視点だけでなく、一部の下位集団に対するビジネス日本語教育への貢献という側面から、国籍 (母語) やビジネス経験の有無、業種別などによる分析にも着手していきたい。

本章で行った統計的・主観的両側面からの DIF 分析は、専門家の持つビリーフが真であるならばそれを形式知として今後のテスト項目作成に生かすことができ、そうでないならば間違ったビリーフによる判断は潜在的に公平なテスト項目を削除してしまうことになり、非経済的であり、無効であるという傍証となる。後者の場合は、専門家に対してたとえば次のようなトレーニングが有効であると考ええる。

- 1) 統計的 DIF が検出された項目にはどのような特徴があり、どのように修正すれば改善され则认为られるかを専門家グループで継続的に議論し、記録する。
- 2) 主観的 DIF については専門家個人の特性にも着目し、DIF を予想する上で個人が持つ無効な観点(2つの下位集団の ICC に実質科学的に意味のある差がない項目に DIF を見出すケースや、DIF が検出された要因に対する予想が受験者の指摘とずれているケースなど)を洗い出し、他の専門家と共有する。

以上のようなフィードバックをすることによって、専門家は自身のビリーフを修正したり、新たなビリーフに対する気づきが生まれたりする可能性がある。それは 6.1.1.2 節で孫ほか(1995)の言うように、DIF を検証すること自体が専門家にテストの目的や測定対象能力を意識させる機会になり、厳しい妥当性検証を促すことにつながるであろう。本章の成果が日本のテスト開発現場に有効に活用されることを期待したい。

第7章 本論文のまとめと今後の課題

本章では、これまでに述べてきた各章の研究の結果をまとめるとともに、今後の課題について述べる。

7.1 本論文の研究課題と解答

本論文は、主に、国内外の大学等の日本語教育機関に在籍し、卒業後に日本語を使ってビジネス活動をすることを希望している日本語学習者のビジネス日本語能力を適切に評価・測定することを目指したものである。経済のグローバル化に伴い、日本企業においては、海外および日本国内の日本語教育機関で日本語や専門分野を身につけた日本語学習者が、高度なグローバル人材として求められる時代が到来している。外国人材をより積極的に採用するためには、彼らの日本語能力を客観的に評価・測定するツールが不可欠であり、その役割を担うもののひとつが、ビジネス日本語能力を客観的に測定する大規模試験であるが、これまで大規模試験の開発・運用上の諸課題を解決するための調査研究は十分に行われていなかった。そこで、本論文では、大規模試験のひとつである BJT を資料とし、大規模試験の開発過程の諸側面に着目し、各側面の運用に有用な示唆を得ることを目指して研究を行った。

第1章では、研究の背景と目的を述べ、本論文の構成を説明した。第2章では、ビジネス日本語に関する先行研究と、テストの信頼性と妥当性に関する先行研究を概観し、本論文の位置づけを明らかにした。第3章から第6章までの各章に設定した研究課題は以下の通りである。

本論文の研究課題

- | | |
|-----|--|
| 第3章 | ①外国人材にはどのような日本語能力が必要とされているか。
②ビジネスコミュニケーション能力とは何か。 |
| 第4章 | ①BJT ビジネス日本語能力テストが測定するビジネス日本語能力とは何か。
②ビジネス日本語テストが測定する能力を反映する Can-do statements をどのように開発するか。 |
| 第5章 | ①どのような要因がテストの質に影響するか。 |

②テスト項目の修正によって、測定の質は改善されるか。

第6章 ①どのような項目で DIF が検出されるか。

②専門家は DIF についてどのようなベリーフを持っているか。

③統計的な DIF と主観的な DIF は一致するか。一致するとすればどのような点が一致するか。

第3章では、日本語でビジネス活動を行おうとする外国人材に求められているビジネス日本語能力について、先行研究ならびに各種調査報告をもとに考察し、構成概念としてのビジネス日本語能力の一端を明らかにした。次に第4章では、第3章で論じたビジネス日本語能力の一部を測定する BJT をより精緻に分析した上で、BJT で測定されている能力を反映するビジネス日本語 Cds を開発し、その有効性を検証した。そして、これら第3章と第4章で示したビジネス日本語能力の具体的な内容を踏まえた上で、第5章と第6章の研究を行った。第5章では、ビジネス日本語テストにおいて、測定の質の向上を目的とした項目修正の有効性を確認した。これは主にテストの信頼性に着目した研究である。これに対し、テストの妥当性を脅かす要因のひとつである、DIF の分析を行ったのが第6章である。ここでは DIF の要因を、定量的・定性的に特定した。それぞれの章で具体的に明らかになったのは次のような点である。

第3章の研究課題①について、企業は外国人材に対し、必ずしも正しい日本語を書いたり話したりすることのみを求めているわけではなく、日本語を道具として使用し、社内や社外と円滑にコミュニケーションを図れることを望んでいるということがわかった。このことから、ビジネスの現場においては、言語的知識を用いてさまざまなやりとりの中で相手の考えを理解し、社内外での人間関係を構築するためのコミュニケーション能力が重視されていることが示唆された。研究課題②について、先行研究の定義や日本語学習者、企業関係者に対する調査からビジネスコミュニケーションの内容を探り、外国語でのコミュニケーション能力に関する理論的枠組みを用いて整理し、ビジネスコミュニケーション能力を構成する要素として、(1) 言語能力、(2) 非言語能力、(3) 方略的能力の3つを同定した。第3章の成果は、企業が外国人材を採用する際にどのような能力を評価すればいいのかといった、評価項目の設定、あるいは、ビジネス日本語教育においてはどのような能力を養成することが必要であるかといった、カリキュラム・シラバスの決定、教授法、教材の決定の際の参考資料となりうる。さらには、ビジネス日本語能力を測定するテストの改

定や新たなテストの開発の際に作成する、テスト細目を検討する上では基礎的な資料となりうると考えられる。

第4章の研究課題①について、大規模試験であるBJTが測定するビジネス日本語能力を、開発当時の資料を参照し、また、実際に出題されたテスト項目を分析して出題形式ごとにどのような能力を測定しているのかを整理した。研究課題②について、BJTが測定する能力を参考に、ビジネス日本語Cdsを開発したプロセスを報告した。予備調査を経て完成したCdsとBJTの得点を比較検証したところ、BJTで測定されているビジネス日本語能力を反映した尺度として改良された。第4章の成果は、まずはBJTの使用者が、テストが測るビジネス日本語能力の具体的な内容を知ることができ、今後のビジネス日本語学習の方法や内容を決める際の指針となることである。ただし、これまでBJTの得点からは、日本語を使用してどのようなことができるのか、あるいはどのような点が不足しているのかを推し量ることはできなかった。そのため、外国人材を採用する企業においても、BJTである得点を得た受験者が、現実のビジネス場面で何ができるのかを具体的に知ることができなかった。第4章で開発したビジネス日本語Cdsは、BJTで測定されるビジネス日本語能力を反映したスケールとしての有効性が確認できたため、外国人材、ビジネス日本語教育担当者、そして企業の採用担当者の三者が利用できる評価ツールとして役立つであろう。

第5章の研究課題①について、統計的に好ましくない項目特性を持つことが判明したテスト項目を分析し、項目統計量が特異であった要因を検証し、その検証結果を報告した。研究課題②について、検証の結果から導き出した、テスト項目修正のための6つの方針に従って修正し、その修正された項目を実験的に実施することにより、項目特性が改善されることを示した。第5章の成果は、テストの測定の質、ここでは項目統計量を改善するための具体的な方策が示されたことである。これによって、項目統計量が特異であったテスト項目に修正を加えれば再利用できるようになり、測定の質の高いテスト項目の作成にも寄与すると考えられる。また、このようにして項目バンクのテスト項目を多くしていけば、新しいテスト項目を毎回作成する人的・経済的資源を節約できることになり、IRTの主目的のひとつである実用性¹⁸を示したとも言えるのである。

¹⁸ 実用性(practicality)とは、テスト運営に必要な、作題、分析、実施に関わるリソース(人、時間、場所、資金など)を確保し続けられるかということで、テストを安定的に運営するために重要な特質である(野口ほか, 2014)。

第6章の研究課題①について、BJTの出題項目計160項目を資料として統計的DIF分析を行ったところ、性別のDIFは20項目、居住地のDIFは67項目が検出された。研究課題②について、ビジネス日本語テストの項目作成の専門家9名による主観的DIF分析の結果、性別においては5つ、居住地においても5つのビリーフが見出された。研究課題③について、受験者の実際の解答に基づく統計的なDIF分析と、専門家の経験的知見に基づく主観的なDIF分析とを比較したところ、あまり一致していないことがわかった。専門家は性別によるDIFをあまり見出さず、居住地によるDIFは実際よりも多く見出す傾向があった。また、統計的DIF分析と主観的DIF分析の結果が一致した項目については、当該のテスト項目に解答した受験者調査で得られたコメントと、主観的DIF分析で専門家から得られたコメントとでは、特に居住地のDIFで、語彙レベルでの一致がいくつか見られ、専門家は語彙の知識の有無の差によるDIFの要因について、適切に指摘している可能性が示唆された。第6章の成果は、ビジネス日本語テストのテスト項目のどのようなファクターが難易度の集団差に影響するのかという点を明らかにするための基礎資料を得たことである。また、統計的なDIF分析と専門家による主観的なDIF分析があまり一致しないという結果は、専門家の持つDIFおよびバイアスに対するビリーフの一部が、実際の難易度の差異の原因を内包していないこと、あるいは専門家がこれらのビリーフを実際以上に広範に適用させ、潜在的に公平なテスト項目を削除してしまうという可能性を示唆している。これらの結果を専門家にフィードバックすることで、妥当性の高いテストやテスト項目作成に寄与するものと考えられる。

7.2 本論文の意義

世界情勢の変化という大きな流れの中で、今後さらに経済のグローバル化が進み、日本企業が外国人材を採用するニーズも高まることが予想されている。それに伴い、外国人材のビジネス日本語能力を適切に評価・測定する必要性も高まることが目されている。そのような中で本論文が持つ意義について以下に述べたい。本論文の意義は、以下の2点に集約される。

7.2.1 ビジネス日本語能力を評価するための指標

第一に、ビジネス日本語能力を評価するための指標を、Cdsを用いて示したことである。

日本企業への就職を希望する留学生をはじめとした外国人材と、外国人材の採用に興味がある、または積極的に行っている企業とを、両者の希望に沿うようにマッチングさせるためのひとつの手段として、共通に評価できる指標を設定することが考えられる。その指標があれば、外国人材にとっては自身の能力のアピールポイントを示すことができるし、企業側にとっては求める人材像を示すことができる。

本論文では、まず、企業が外国人材に求める能力に関する調査から、ビジネス場面でのコミュニケーション能力が重視されていることを把握し、ビジネスコミュニケーション能力の構成要素について同定することで、構成概念としてのビジネス日本語能力を明らかにした。そして、ビジネスコミュニケーション能力を測定することを目的とした **BJT** の開発時の資料や出題項目を分析し、テストで測定するビジネス日本語能力の具体的な内容を探った。さらに、**BJT** が測定する能力を参考にしてビジネス日本語 **Cds** を開発し、ビジネス日本語能力を評価するシステムとしての有効性を確認した。本論文で開発したビジネス日本語 **Cds** は、現実のビジネス場面でどのような能力があるかを、能力記述文で示すことができ、外国人材にとっても企業にとっても理解しやすいものである。

ビジネス日本語 **Cds** は、ビジネス日本語教育の指標としても有用である。ビジネス日本語教育担当者は、学習者に教育目標を示すことができるし、学習者は到達目標を意識することができる。吉ほか (2016) が行った研究では、ビジネス日本語 **Cds** を用いて学習者のビジネス日本語学習に対する意識調査を行った。ビジネス日本語 **Cds** の各項目について、自己評価および学習意識 (「A : 授業で先生に教えてもらいたい」、「B : ビジネス日本語の教材やマンガ・ドラマなどを使って、自分で学習するのがいい」、「C : 実際に仕事を始めてから、周りの人を見ながら学んでいく」から選択) を回答させ、学習者がビジネス日本語の授業にどのような内容を望んでいるかを探っている。

ビジネス日本語 **Cds** はまた、教育効果を探るためにも利用できる。葦原ほか (2016) の研究では、外国人留学生を対象としたインターンシッププログラムの効果を測る手段としてビジネス日本語 **Cds** を用いている。R 大学では、外国人留学生の日本国内での就職を促進させるため、外国人留学生を対象としたインターンシップを実施しており、2 週間のビジネス日本語教育と、その後 4 週間のインターンシップ研修を通して、ビジネス日本語の自己評価がどのように変わったかを探っている。このように、本論文で示したビジネス日本語能力を評価するための指標は、外国人材、ビジネス日本語教育担当者、そして企業の採用担当者の三者が共通で利用できるものとなり、学習、教育、就職のそれぞれとの接

続を実現させ、日本語学習者の効率的なビジネス日本語学習や高度なビジネス日本語能力を持つ外国人材の就職支援に貢献するものである。これは、日本語教育の社会貢献のひとつとして、ビジネス日本語能力の評価システムの開発を前進させたと言えるだろう。

なお、本論文では、ビジネス日本語能力の評価を考える際に **BJT** を資料として検証した。前述したように、**BJT** は、日本語の知識やビジネスに関する専門的知識を前提として、ビジネス場面での日本語によるコミュニケーションの能力があるかどうかを測ることに最大の意義を求めている。しかしながら、現実のビジネス場面で必要となる「書く」技能や「話す」技能については直接測定することはできない。また、ビジネス活動は、業種や職種、役職によっても異なり、膨大で多様な領域であり、日々刻々と変化するものである。音声と問題用紙によって、限られた環境と時間と項目数で、ビジネス日本語の事項を網羅的に測定することは不可能である。つまり、**BJT** が測定しているビジネス日本語能力は、ビジネス活動全体のごく一部分であり、そのすべてを測れるテストではなく、**BJT** がビジネス日本語能力を測定するテストとして万能であるという立場に立ったものではない。しかし、黄（2016）が、日本の人とビジネスコミュニケーションを実際に行う経験があまりない日本語学習者にとって **BJT** の出題内容は学習の指針となるものであると述べているように、**BJT** の理念と測定する能力の概念が、ビジネス日本語能力を理解する上でのひとつの知見を提示することとなり、ビジネス日本語教育関係者の教育の指針ならびに日本語でビジネス活動を行う外国人材の日本語学習の一助になることが期待される。

7.2.2 テスト開発・運用

第二に、受験者の解答データからテスト項目の問題点を発見し、精度や質の高いテストやテスト項目の開発・改良のための方策を実証的に示したことである。

測定に不可欠な性質が信頼性と妥当性である（Bachman, 1990）。言語テストにおいても、測定の安定性を示す信頼性や、テストで測ろうとしているものと実際に測っているものがどれほど合致しているかという妥当性はもっとも重視される。大規模試験のような利害関係の大きいテストでは、多くの人に大きな影響を与える可能性のあるテストであるため、試行を繰り返し、修正を重ねなければならない。そして、どのような状況であろうと、どのような目的であろうと、テストは能力の測定に役立つものでなければならない（Bachman & Palmer, 1996）。

本論文では、**BJT** のテスト項目について、信頼性の側面と妥当性の側面から分析を行っ

た。まず、項目識別力を中心とした項目特性が好ましくないテスト項目について、その要因を検証し、その結果から導き出したテスト項目修正の方針に従って修正し、再実施したところ、項目特性が改善された。次に、測定の妥当性を脅かす要因のひとつである DIF の分析を行い、特定の下位集団にとって難易度が高くなる要因を探った。これらの研究から、テストの信頼性・妥当性を高めるための方策を示した。本論文でテスト項目の項目統計量を改善するための方策が示されたということは、テストを実施した後の、項目統計量が特異であったテスト項目の改善という側面だけでなく、新規にテスト項目を作成する際の方針としても有効であり、BJT に限らず、信頼性・妥当性の高いテストの開発・運用に資するものである。

また、本論文のテストの信頼性と妥当性に関する研究では、統計的な分析と合わせてテスト項目作成の専門家による主観的な分析を合わせて行った。研究で得られた成果から、統計的分析と主観的分析のそれぞれの観点から得られる知見が有機的に結びつくことによって、精度や測定の質の高いテスト開発だけでなく、テスト開発にかかる人的・経済的資源の節約にもつながることが証明されたと言えよう。このことは、テストの実用性を確保するための、継続的なテスト開発者の育成にも貢献することが期待される。

7.3 本論文が与えるビジネス日本語能力の評価・測定と日本語教育への示唆

本論文は、外国人材のビジネス日本語能力を適切に評価・測定することを目的として行った研究である。ここで再び、本論文がビジネス日本語能力をどのようにとらえているかについて整理したい。その上で、本論文がビジネス日本語能力の評価・測定、そして日本語教育に与える示唆について述べ、本論文の意義をより広い視野で示したい。

第3章で、日本語教育および企業関係者に対する調査の結果から、いわば「社会的要請」としてのビジネス日本語能力(=ビジネスコミュニケーション能力)について考察した(3.3節を参照)。また、昨今のビジネス日本語教育の目標が、「日本語や日本文化に精通している人材の育成」から、一社会人としての「包括的能力の育成」と変化している背景を踏まえ、現実にグローバル人材として働く外国人ビジネスパーソンに求められる「理想」としてのビジネス日本語能力について論じた。「理想」としてのビジネス日本語能力は、「社会的要請」としてのビジネス日本語能力を包含するものと考えられるが、ではテストで測るビジネス日本語能力はどうであろうか。第4章では、BJTの開発資料や実際に出題された

テスト項目を分析し、BJT が測定する構成概念について考察し、BJT がとらえるビジネス日本語能力は、「社会的要請」としてのビジネス日本語能力の一部であると述べた（4.2.1.4 節を参照）。以上、本論文の第 3 章および第 4 章で述べたビジネス日本語能力の関係を表したものが図 7-1 である。図で示したように、本論文では、テストが測定するビジネス日本語能力は、「理想」としてのビジネス日本語能力のうちのごく一部分である、ととらえている。

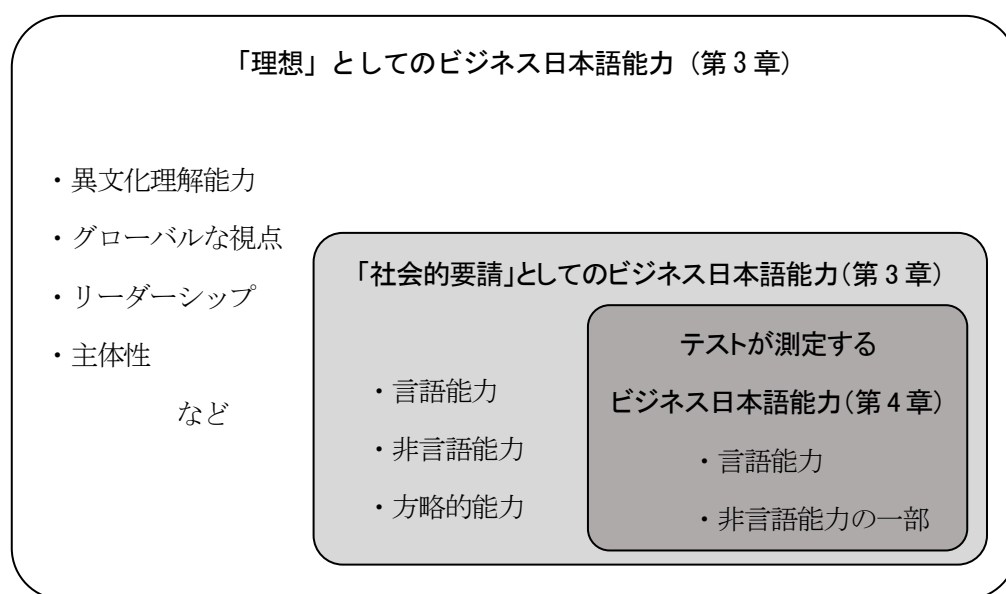


図 7-1 本論文のビジネス日本語能力のとらえ方

図 7-1 の「テストが測定するビジネス日本語能力」は、本論文では BJT が測定する能力の領域を示しているが、前述したように BJT は日本企業への就職を希望する外国人材が増加していることを背景に開発されたこと、そして現在も受験者の半数以上が就労前の学生であることから、BJT が測定するビジネス日本語能力は、いわば「ビジネスの世界の入り口」としての日本語能力であると言える。だからこそ、測定の対象が多様なビジネス場面で共通に見られる部分になり、現実のビジネス活動で使われている専門的、個別的な領域を網羅することができないとも考えられる。実際に日本語を使ってビジネス活動を行っているビジネスパーソンが受験対象のテストであれば、業種別、職種別、役職別、組織形態別など、測定の対象がより個別的になり、それによって、テストの真正性は高くなり、「社会的要請」あるいは「理想」としてのビジネス日本語能力の測定に近づくことになる

かもしれない。ビジネス日本語能力の評価・測定を考える際は、そのような現実的で高度な能力を測るテストについても検討すべきであろう。すなわち、テストが測定するビジネス日本語能力の領域をどこまで拡張できるのかというのは、ビジネス場面の真正性をテストにどこまで求めるのか、あるいは求めることができるのかを追究することと同じことを意味するのではないかと考える。当該のテストが測ろうとする構成概念やテスト利用者の要請に合わせて、適切な妥協点を探るべきであろう。

本論文の第4章でその開発過程を報告したビジネス日本語 Cds は、テストが測定するビジネス日本語能力の領域を把握した上で、BJT が直接測定できないビジネス日本語能力、「書く」「話す」「聴いて書く」などの産出機能やビジネス場面での日本語の運用能力を評価するツールとなることを目指した。検証の結果、未だ問題点は残っており、BJT がそうであるようにビジネス日本語 Cds もまた、評価できるビジネス日本語能力には限界があることが示唆されたものの、ビジネス日本語 Cds は、テストが測定するビジネス日本語能力と「社会的要請」としてのビジネス日本語能力をつなぐ役割を担うことができるだろう。本論文では BJT が測定するビジネス日本語能力を参考にビジネス日本語 Cds を開発したが、これを皮切りに、今後はテストの領域を超え、図 7-1 の「社会的要請」「理想」としてのビジネス日本語能力の領域にあるスキルを評価できるような評価ツールの開発が望まれる。

7.2.1 節で、BJT が測定する能力を反映するビジネス日本語 Cds は、学習、教育、就職のそれぞれとの接続を実現させるものだと言ったが、ビジネス日本語 Cds 以外にも、本論文で行った BJT のテスト項目の分析もまた、ビジネス日本語教育に示唆を与えている。第5章で行ったテスト項目修正に関する分析では、ビジネス場面で使われる抽象的な表現や受験者の母語の干渉などがテストの測定の質を低くする可能性について述べたが、裏を返せば、外国人材に対するビジネス日本語教育ではこの点の指導、配慮が十分でなかったとすることができる。同じように、第6章で行った DIF の分析では、特定の下位集団にとって不得意なことを把握することができる。たとえば、6.2.4 節の項目分析(6)で挙げた、病気で欠勤する同僚の家族からの連絡に対し、正答の「お大事になさってください」ではなく、誤答の「お困りでしょね」や「しょうがありませんね」の選択率が高かった国外のビジネス日本語学習者に対しては、ビジネス上で良好な対人関係を築くには、このような場面でどのような表現が望ましいかといった観点からの指導が必要かもしれない。また、日本では一般的だが、国外では一般的でない制度や習慣に関する知識を前提とするテスト

項目で国外受験者に不利となった可能性があるものが散見されたが、この結果は、必ずしもそのような項目を不適切として削除または修正をするということを要求するだけではなく、むしろ、日本特有の制度、習慣に関する内容もビジネス日本語教育の指導項目のひとつとすべきということを示していると考えられることも可能である。このように、項目分析の結果を具体的なビジネス日本語教育の指導へ応用することが可能である。2.1.2.1 節で、敬語の指導を中心とする授業や新聞読解とディスカッション等のビジネス日本語の授業は、ビジネス現場で求められている日本語能力を養成するという目的からずれているという、清（1995）の指摘を挙げたが、テストの項目分析で得られた結果からヒントを得て指導内容を決定することは、外国人材が潜在的に必要な知識やスキルを日本語教育で補完することにつながるのではないだろうか。本論文の第 5 章、第 6 章で質的な分析を行ったビジネス日本語テストの専門家は、「以前自分のクラスにこのように考えた学習者がいたから、間違えやすいかもしれない」、「ある国の学習者は、日本のこのような制度を知らなかったので、国外の出題には不向きだ」というように、各自のビジネス日本語教育の実践現場で、日本語学習者が興味を示した点や誤解していた点をテスト項目の作成に活かしている。一方で、作成したテスト項目の出題後、分析した結果から得た知見を、今度は授業で扱ったり教材に取り入れたりして、ビジネス日本語教育の現場に還元しているという。このことは教育実践とテスト開発のそれぞれから得られる知見が相互に好循環していることを示している。つまり、テストは教育実践にヒントを与え、教育実践はまた、より良いテストの開発・運用へのヒントを与えるのである。

7.4 今後の課題と展望

本論文をさらに発展させるために、今後の課題として次の点を挙げたい。

第一に、本論文では、外国人材に求められるビジネス日本語能力に着目し、その適切な評価・測定の方法について述べてきた。本論文の調査対象はこれから日本語を使ってビジネス活動をすることを目指している外国人留学生が主であり、実際にビジネス現場で働くビジネスパーソンの評価・測定に資するものであるかどうかは今後の検証が必要である。

第二に、第 4 章において量的調査として行ったビジネス日本語 Cds の開発については、産出機能である「話す」「書く」「聴いて話す」といった項目が能力を評価する上で十分に有効であるかどうか未だ解明されていない。今後さらにデータを増やしたり、ビジネス場

面でのパフォーマンスの能力を測定するテストを行うための質的な調査を行ったりするなどして研究を進めていく必要がある。Cds の項目の記述についても、「業務上未体験のタスクについてどのように評価すべきかはっきりしない」等の指摘について、その影響がどの程度あるのか、十分なデータサイズを収集しての検証が必要である。

第三に、第 5 章で行ったテスト項目の修正に関する研究では、統計的性質が特異な項目計 28 項目を選び出して検証を行ったが、BJT の過去の出題項目にはほかにも、本論文で分析した項目とは別の要因で項目統計量が不整備になっているものも含まれている可能性がある。今後はより多くのテスト項目について検証することで、ビジネス日本語テストの測定の質を高める新たな方策を示すことを課題としたい。

第四に、第 6 章で行ったビジネス日本語テストにおける DIF 分析では、性別、居住地の違いによる下位集団に着目して分析を行ったが、それ以外の下位集団、たとえばビジネス経験の有無や業種別などによる分析も行っていきたい。DIF が検出された要因が特定できなかった項目、あるいは今回は分析の対象外とした国外受験者に有利な DIF が検出された項目については、たとえば特定の文化圏の受験者で、かつ日本国内に居住した経験がある集団とない集団を比較するなど、下位集団を細分化することによって、別の結果が得られる可能性もある。今後は、統計的に検出された DIF がバイアスであるかどうかの判断を行い、バイアスと判断された項目についてはその要因を探り、第 5 章と同様に、修正した上で実験的に実施して改善されたかどうかの検証を行うことが必要である。すなわち、バイアスの要因を特定するための実証的な研究を行うことを今後の課題としたい。これについては、国外の最新のビジネス日本語教育事情を、教材や指導法を通して分析するなど、質的な調査も平行して行っていきたい。

最後に、今後の研究の展望について述べたい。日本語の能力が高いあるいはビジネスの専門的知識が豊富だからといって、その人が「仕事ができる」とは限らない。ビジネスの現場では、たとえば、メーカーの商品広告の制作などのように、複数の業種の社員がチームを組んでプロジェクトを行うことが少なくないが、その場合、それぞれの業種に特有の専門知識を持ち寄って、ビジネスやプロジェクトを進行させていくことになる。そこで求められるのは、どのような業種にも共通するビジネスの常識や言語的知識をツールとして、相手の知らない情報を伝えたり、自分の知らない情報を入手したり推測したりしながら課題を遂行していくことである。この際に必要となるのが、コミュニケーション能力である。

BJT はビジネス場面でのコミュニケーション能力を測定するテストとして社会的に広

く活用されているテストであるが、BJT の実施は 2016 年現在、年 2 回であり、2017 年度より CBT (Computer Based Testing) 方式に全面移行することが発表されているが、それでも受験回数には制限があり、受験地もアジア地域を中心として世界約 60 会場と限られている。BJT 以外に、いつでもどこでもビジネスコミュニケーション能力を客観的に測定するツールがあれば外国人材や企業の採用担当者にとって利便性が高い。今後は、コンピュータを使ってインターネット上で行うビジネス日本語のアダプティブ・テストの開発に取り組み、本論文と同様の調査・分析を行うなどして検討を重ねつつ、外国人材のビジネスコミュニケーション能力の測定ツールとして役立てることを目標とする。

本論文の成果が、外国人材の就職支援や企業の外国人材の採用活動の促進に寄与すべく、さらなる研究を重ねていきたいと考えている。

付記

本論文は、以下の初出論文をもとに執筆した。

第2章

小野塚若菜（2015）「第3章 BJT ビジネス日本語能力テスト」李在鎬 編『日本語教育のための言語テストガイドブック』, 49-66, くろしお出版

小野塚若菜, 前川眞一（2015）「設問回答率分析図によるビジネス日本語テストの項目分析 —「BJT ビジネス日本語能力テスト」における検証—」学芸国語国文学 嶋中道則教授・加藤清方教授 退職記念号, 47, 東京学芸大学学芸国語国文学会, 56-71

第3章

小野塚若菜, 篠崎佳子, 島恭子（2009）「ビジネスコミュニケーション能力の向上を目指す日本語教育に関する考察」日語日文学, 41, 大韓日語日文学会, 191-200

第4章

葦原恭子, 小野塚若菜（2014）「ビジネス日本語能力を評価するシステムとしてのビジネス日本語 Can-do statements の開発 —BJT ビジネス日本語能力テストの測定対象能力に基づいて—」日本語教育, 157, 日本語教育学会, 1-15

第5章

小野塚若菜, 加藤清方, 梅木由美子, 越前谷明子, 前川眞一（2015）「測定の質の向上を目的としたテスト項目の修正—ビジネス日本語テストを資料とした実証的研究—」日本テスト学会誌, 11, 日本テスト学会, 111-130

第6章

小野塚若菜（2016）「ビジネス日本語テストにおける DIF の分析—性別および居住地を下位集団として—」日本言語テスト学会誌, 19, 日本言語テスト学会, 86-108

なお、本研究は以下の助成を受けて行われた。

- 平成 28-29 年度文部科学省科学研究費（特別研究員奨励費 課題番号 16J00030）「ビジネス日本語能力テストにおける DIF の検出と改善に関する実証的研究」（研究代表者：小野塚若菜）
- 平成 26-27 年度文部科学省科学研究費（若手研究（B） 課題番号 26770186）「ビジネス日本語能力を測定するテストにおける DIF 項目の検出と改善に関する基礎研究」（研究代表者：小野塚若菜）
- 平成 22-26 年度文部科学省科学研究費（基盤研究（C） 課題番号 22520535）「外国人のビジネス日本語能力の評価に関する基礎研究—評価システムの確立を目指して—」（研究代表者：葦原恭子）
- 平成 20-22 年度文部科学省科学研究費（基盤研究（B） 課題番号 20401023）「ビジネス日本語能力テストの信頼性と妥当性の連関に関する実証的研究」（研究代表者：加藤清方→前川眞一）

謝辞

本研究を遂行し学位論文をまとめるにあたり、指導教員である今井新悟先生には終始多くのご指導を賜りました。衷心より感謝申し上げます。また、酒井たか子先生、加納千恵子先生、許明子先生、宇佐美慧先生には、副査として、時に応じ厳しくご指導いただき、そして励ましてくださいました。感謝に堪えません。

博士課程への進学を勧めてください、研究全般にわたる多大なご支援、ご指導を賜りました、東京工業大学大学院教授の前川眞一先生、東京学芸大学名誉教授の加藤清方先生に深く感謝しております。本研究におけるテストデータの使用にあたっては、資料や項目データを快く提供してくださった日本貿易振興機構の北条尚子所長に感謝申し上げます。

最後に、本研究への惜しめない協力と、貴重な時間をくださった日本語学習者の方々、日本語教育機関の教員の方々、ビジネス日本語能力テストの専門家の方々、そして多くの友人たちに対し、深い感謝の意を表して謝辞といたします。

参考・引用文献

1. 葦原恭子・小野塚若菜（2016）「琉球大学のインターンシップにおける業務経験が外国人留学生の自己評価に与える影響に関する事例研究—ビジネス日本語 Can-do statements の分析を通して—」琉球大学留学生センター紀要, 3, 1-16
2. アルク教育総合研究所監修（2015）『グローバル教育を考える』アルク
3. 池田伸子（1995）「教授システム開発におけるニーズ評価について」ICU 日本語教育研究センター紀要, 4, 国際基督教大学日本語教育研究センター, 43-62
4. 池田伸子（1996）「日本人ビジネスマンの話し言葉における語彙調査—ビジネスマン用日本語教育システム開発の基礎として—」日本語教育, 88, 117-127
5. 池田伸子（2001）『ビジネス日本語教育の研究』東京堂出版
6. 池田央（2001）『現代テスト理論』朝倉書店
7. 池田央（2008）「評価における信頼性と妥当性問題—古くて新しい課題に向けて—」指導と評価, 54-08, 4-7
8. 一般社団法人応用日本語教育協会「STBJ とは」
<http://www.ajlea.net/about_stbj/>（2016 年 9 月 1 日）
9. 井上俊哉・孫媛・野口裕之・酒井たか子（2007）「留学生対象の日本語能力テストにおける DIF について」日本テスト学会第 5 回大会発表論文集, 92-93
10. 植野真臣・吉村幸・荘島宏二郎・橋本貴充（2007）「大学入試センター試験「情報関係基礎」の統計分析」大学入試センター研究紀要, 36, 71-100
11. 大友賢二（2009）「項目応答理論—TOEFL・TOEIC 等の仕組み—」電子情報通信学会誌, 92(12), 1008-1012
12. 奥村太一（2014）「項目特性曲線と項目パラメータ」
<<http://www.juen.ac.jp/lab/okumura/test/sect0032.html>>（2016 年 9 月 1 日）
13. 小野寺志津・李徳奉・金久保紀子（2004）「ビジネス日本語教育のあり方—新入社員教育から見えるもの—」東京家政学院筑波女子大学紀要, 8, 127-137
14. 海外技術者研修協会（2007）「平成 18 年度 構造変化に対応した雇用システムに関する調査研究（日本企業における外国人留学生の就業促進に関する調査研究）」
<http://www.hidajapan.or.jp/jp/project/nihongo/asia/r_info/pdf/press070514_2

- pdf> (2016 年 9 月 1 日)
15. 海外技術者研修協会 (2009)「平成 19 年度 産業競争力強化高度人材育成事業(アジア人財資金構想共通カリキュラムマネージメントセンター事業報告書」
<http://www.hidajapan.or.jp/jp/project/nihongo/asia/r_info/pdf/press081012.pdf> (2016 年 6 月 25 日)
 16. 加藤清方(2010)『BJT ビジネス日本語能力テスト 体験テストと解説 改訂版』
日本漢字能力検定協会
 17. 株式会社 DISCO (2015)「『外国人留学生の採用に関する企業調査』アンケート
結果<2015 年 11 月調査>」
<<http://www.disc.co.jp/uploads/2015/12/2015kigyuu-gaikoku-report.pdf>> (2016
年 5 月 25 日)
 18. 神吉宇一 (2010)「ビジネス日本語教育に求められるもの」日本語教育学会教師
研修<ビジネス日本語 (ケースメソッド) >配付資料, 2010 年 1 月 9 日, 政策
研究大学院大学
 19. 企業活力研究所 (2010)「平成 22 年度アジア人財資金構想プロジェクトサポート
センター事業『日本企業における高度外国人材の採用・活用に関する調査』報告
書」<http://www.meti.go.jp/policy/asia_jinzai_shikin/surveydata_2010.pdf>
(2016 年 5 月 25 日)
 20. 岸学 (2010)「学力テスト問題の作成・分析・評価—教育心理学の視点から—」
富山県中学校教育研究会『学力診断サポート事業』
<http://www.u-gakugei.ac.jp/~kishilab/toyama_06.pdf> (2016 年 6 月 25 日)
 21. 吉陽・井上里鶴・ショリナ・ダリヤグル・孫思琦・小野塚若菜 (2016)「ビジネス
日本語の学習意識に関する調査—ビジネス経験の有無による違いに着目して—」
2016 年度日本語教育学会春季大会予稿集, 213-218
 22. 熊谷龍一 (2012)「統合的 DIF 検出方法の提案—“EasyDIF”の開発」心理学研
究 2012 年, 83-1, 35-43
 23. 熊谷龍一・脇田貴文 (2003)「特異項目機能検出方法の比較—BILOG-MG と
SIBTEST を用いた検討」Bulletin of the Graduate School of Education and
Human Development. Psychology and human developmental sciences, 50,
83-90

24. グローバル人材育成推進会議（2011）「グローバル人材育成推進会議 中間まとめ」経済産業省
<http://www.meti.go.jp/policy/economy/jinzai/san_gaku_kyodo/sanko1-1.pdf>
（2016 年 11 月 7 日）
25. 経済産業省「海外事業活動基本調査」
<<http://www.meti.go.jp/statistics/tyo/kaigaizi/result-2.html>>（2016 年 5 月 25 日）
26. 黄英哲（2016）「ビジネス日本語能力テストの分析と台湾人日本語学習者の受け止め方」『2016 年度日本語教育学会研究集会第 2 回予稿集』日本語教育学会, 37-40
27. 国際交流基金「JF 日本語教育スタンダード」
<<http://jfstandard.jp/top/ja/render.do;jsessionid=9F69A04284C15F6DC4F2027BB18285C4>>（2016 年 6 月 25 日）
28. 近藤彩（2007）『日本人と外国人のビジネス・コミュニケーションに関する実証研究』シリーズ言語学と言語教育, 9, ひつじ書房
29. 近藤ブラウン妃美（2012）『日本語教師のための評価入門』くろしお出版
30. 国際ビジネスコミュニケーション協会「ETS のブランドクオリティ」
<http://www.toeic.or.jp/toeic/about/what/philosophy/philosophy_03.html>
（2016 年 5 月 25 日）
31. 三枝令子（2003）「日本語 Can-do-statements の開発」日本行動計量学会 31 回大会発表論文抄録集, 31, 102-105
32. 三枝令子（2004）『日本語 Can-do-statements 尺度の開発』平成 13 年度～平成 15 年度科学研究費補助金（課題番号 13480068）研究成果報告書
33. 三枝令子（2008）『DIF を用いた日本語テスト改善に向けての基盤的研究』平成 18 年度～20 年度科学研究費補助金（課題番号 18320081）研究成果報告書
34. 酒井たか子・井上俊哉・浅見かおり・伊東祐郎・三枝令子・島田めぐみ・孫媛・野口裕之（2007）「日本語プレースメントテストにおける DIF 研究」2007 年度日本語教育学会春季大会予稿集, 201-206
35. 澤木泰代（2011）「大規模言語テストの妥当性・有用性検討に関する近年の動向」言語教育評価研究, 2, 54-63
36. 澤木泰代（2015）「言語教育プログラムのためのプレースメント・テスト開発」

ICU 日本語教育センター2015 年度連続講演会, 2015 年 10 月 21 日, 国際基督教大学

37. 財団法人海外技術者研修協会 AOTS 日本語教育センター (2011) 「教育機関のための外国人留学生ビジネス日本語教育ガイド」 アジア人財資金構想プロジェクトサポートセンター (編集), 経済産業省
<http://www.meti.go.jp/policy/asia_jinzai_shikin/studybusinessjapaneseguide.pdf> (2016 年 6 月 22 日)
38. 繁桝算男 (編著) (1998) 『心理測定法』放送大学教育振興会
39. 島田めぐみ (2010) 「自己評価 Can-do statements に関する一考察—客観テストとの比較を通して—」東京学芸大学紀要 総合研究科学系Ⅱ, 61, 267-277
40. 島田めぐみ・三枝令子・野口裕之 (2006) 「日本語 Can-do statements を利用した言語行動記述の試み—日本語能力試験受験者を対象として—」世界の日本語教育, 16, 国際交流基金日本語国際センター, 75-88
41. 清水裕子 (2005) 「測定における妥当性の理解のために—言語テストの基本概念として—」立命館言語文化研究, 16(4), 241-254
42. 新日本有限責任監査法人 (2015) 「平成 26 年度産業経済研究委託事業 (外国人留学生の就職及び定着状況に関する調査) 報告書」
<http://www.meti.go.jp/policy/economy/jinzai/global/pdf/H26_ryugakusei_report.pdf> (2016 年 9 月 17 日)
43. 鈴木敦 (2012) 「世界 ターゲット市場に強い人材発掘を」ジェトロセンサー2012 年 8 月号 日本貿易振興機構
<https://www.jetro.go.jp/ext_images/jfile/report/07001026/world_hr.pdf> (2016 年 11 月 7 日)
44. 鈴川由美・豊田秀樹・川端一光 (2008) 「わが国の数学教育は数学を日常の中で活用する能力を重視しているか : PISA2003 年調査の DIF による分析」教育心理学研究, 56(2), 206-217
45. 清ルミ (1995) 「上級日本語ビジネススピーブルのビジネスコミュニケーション上の支障点—インタビュー調査から教授内容を探る—」日本語教育, 87, 139-151
46. 孫媛・井上俊哉 (1995) 「アメリカにおける差異項目機能(DIF)研究」学術情報センター紀要, 7, 国立情報学研究所, 193-216

47. 田崎勝也 (2008) 『社会科学のための文化比較の方法』 ナカニシヤ出版
48. 田崎勝也・芳賀日登美・宮原哲・申知元 (2015) 「ビジネス・コミュニケーション能力を測定するスキルテストの開発」 *Aoyama Journal of International Studies*, 2, 123-140
49. 中小企業庁「地域中小企業の海外人材確保・定着支援事業」
<http://www.chusho.meti.go.jp/keiei/kokusai/2014/140226Teichaku.html> (2016年5月25日)
50. 中津原文代 (2013) 「能力基準としての Can-do statements とテストの妥当性を検証する『社会的・認知的枠組み』 (Socio-cognitive framework) について」 *言語教育評価研究*, 3, 44-53
51. 中村洋一 (2002) 『テストで言語能力は測れるか～言語テストデータ分析入門～』 大友賢二 (監), 桐原書店
52. 中村洋一 (2008) 「第一章 5. 古典的テスト理論と項目応答理論」 伊東祐郎『留学生の日本語能力測定のためのテスト項目プールの構築』平成16～平成19年度科学研究費補助金 基盤研究 (A) 研究成果報告書, 16-21
53. 日本漢字能力検定協会 (2013) 「外国人留学生の採用と日本語能力に関する調査」
<http://www.kanken.or.jp/bjt/survey_reports/data/survey_reports_report01.pdf> (2016年5月25日)
54. 日本学生支援機構 (2016) 「平成27年度私費外国人留学生生活実態調査」
<http://www.jasso.go.jp/about/statistics/ryuj_chosa/_icsFiles/afieldfile/2016/09/23/ryujchosa27p00.pdf> (2016年11月18日)
55. 日本語ビジネスコミュニケーション能力測定テスト
<<http://www.ask-digital.biz/training/japanese-biz-com-test.html>> (2016年6月25日)
56. 日本語能力試験 JLPT N1～N5 : 認定の目安
<<https://www.jlpt.jp/about/levelsummary.html>> (2016年6月25日)
57. 日本語能力試験実施委員会・日本語能力試験企画小委員会監修 (2011) 『平成21年度日本語能力試験 (第1回・第2回) 分析評価に関する報告書』 アスク出版
58. 日本テスト学会 (編) (2007) 『テスト・スタンダードー日本のテストの将来に向けて』 金子書房

59. 日本貿易振興機構人材開発支援課 (2005)「ジェトロテストに対する大連の日系企業の声」日本貿易振興機構
60. 塗師斌(1989)「二値データに基づく尺度の一次元性の評価の方法 Assessing the Unidimensionality of a Set of Dichotomous Variables」横浜国立大学教育紀要, 29, 137-148
61. 野口裕之・大隅敦子 (2014)『テストニングの基礎理論』研究社
62. 野元千寿子 (2005)「大学におけるビジネス日本語教育—受講者アンケートを通して見えるもの—」昭和女子大学大学院日本文学紀要, 16, 13-23
63. 南風原朝和 (2002)『心理統計学の基礎—統合的理解のために』有斐閣
64. 南風原朝和 (2011)『臨床心理学を学ぶ 7 量的研究法』東京大学出版会
65. 芳賀日登美・宮原哲・田崎勝也・申知元 (2015)「日本において企業が考えるコミュニケーション能力とは—半構造化面接法による探索的研究—」Aoyama Journal of International Studies, 2, 81-101
66. 藤森進 (1991)「第2章 いろいろな項目反応モデル」芝祐順 (編)『項目反応理論 —基礎と応用—』東京大学出版会, 33-50
67. 法務省入国管理局 (2015)「平成26年における留学生の日本企業等への就職状況について」<<http://www.moj.go.jp/content/001153834.pdf>> (2016年9月15日)
68. 堀井恵子・向山陽子 (2013)「学習目標達成のためのビジネス日本語 Can-do statements の開発」2013年度日本語教育学会春季大会予稿集, 313-314
69. 前川眞一 (1991)「第4章 パラメタの推定」芝祐順 (編)『項目反応理論 —基礎と応用—』, 東京大学出版会, 119-122
70. 松井治子 (1993)「ビジネス日本語とは」AJALT, 16, 8-9
71. 水本篤・竹内理 (2011)「効果量と検定力分析入門—統計的検定を正しく使うために—」2010年度部会報告論集『より良い外国語教育のための方法』外国語教育メディア学会関西支部メソドロジー研究部会, 47-73
72. 安永和央・石井秀宗 (2011)「国語読解テストにおける設問文中の単語の難しさが能力評価に及ぼす影響—具体例を回答させる設問の検討—」名古屋大学大学院教育発達科学研究科中等教育研究センター紀要, 58, 105-112
73. 安永和央・石井秀宗 (2012)「国語読解テストにおける設問設定が能力評価に及

- ばす影響—特色ある読解教育を実践している小中一貫教育校の事例」 *Journal of quality education*, 4, 111-130
74. 李志瑛 (2002) 「ビジネス日本語教育を考える」 言語文化と日本語教育 2002 年 5 月特集号, 245-260
75. 李在鎬 (2015) 「第 1 章 言語テストを捉える」 李在鎬 (編) 『言語テストガイドブック』くろしお出版, 8-30
76. 渡辺直登・野口裕之 (編著) (1999) 『組織心理測定論：項目反応理論のフロンティア』白桃書房
77. AJALT ビジネス日本語評価基準作成プロジェクト (2008) 「ビジネス日本語評価基準作成の試み」 *AJALT*, 31, 36-39
78. BJT ビジネス日本語能力テスト「受験者による自己評価『CAN DO レポート』」
<<http://www.kanken.or.jp/bjt/about/evaluation.html>> (2016 年 6 月 24 日)
79. BJT ビジネス日本語能力テスト「レベルガイド」
<<http://www.kanken.or.jp/bjt/about/levelguide.html>> (2016 年 6 月 24 日)
80. BJT ビジネス日本語能力テスト「BJT の特徴」
<<http://www.kanken.or.jp/bjt/about/feature.html>> (2016 年 6 月 24 日)
81. ETS 「ETS のテストに対するアプローチ」
<https://www.ets.org/jp/understanding_testing/> (2016 年 6 月 24 日)
82. ETS (2008) TOEIC Can-do Guide Executive Summary
<http://www.ets.org/Media/Tests/Test_of_English_for_International_Communication/TOEIC_Can_Do.pdf> (2016 年 5 月 27 日)
83. JETRO (2000) 『JETRO40 年の歩み』ジェトロ 40 年史編さん委員会
84. J.TEST 実用日本語検定「ビジネス J.TEST」
<<http://j-test.jp/sample-page/%E3%83%93%E3%82%B8%E3%83%8D%E3%82%B9%E3%82%8A%E3%82%8E%E3%82%94%E3%82%85%E3%82%94>> (2016 年 9 月 1 日)
85. Abedi, J. (2006) Language Issues in Item Development. In Downing, S.M. & Haladyna, T.M., (Eds.) *Handbook of Test Development*, 17, 377-398
86. Adler, N. J. & Bartholomew, S. (1992) Managing globally competent people.

Academy of Management Executive, 6 (3), 52-65

87. Alderson, J.C., Clapham, C., & Wall, D. (1995) *Language Test Construction and Evaluation*, Cambridge University Press. [チャールズ・オルダーソン, キャロライン・クラッファム, ダイアン・ウォール著, 渡部良典(編訳) (2010), 言語テストの作成と評価—あたらしい外国語教育のために, 春風社]
88. American Psychological Association. (1954) Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 51 (2, supplement).
89. American Psychological Association. (1966) *Standards for educational and psychological tests and manuals*.
90. American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1974) *Standards for educational and psychological tests*. American Psychological Association.
91. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999) *Standards for educational and psychological testing*. American Educational Research Association. [赤木愛和, 池田央(監訳) (1993), 教育・心理検査法のスタンダード, 図書文化社]
92. Angoff, W.H. (1982) Use of Difficulty and discrimination indices for detecting item bias, Berk, R.A., (Ed.) *Handbook of methods for detecting test bias*. Baltimore, Johns Hopkins University Press. 96-116
93. Angoff, W.H. (1993) Perspective on differential item functioning methodology. In P.W.Holland & H.Wainer., (Eds.), *Differential item functioning*. Hillsdale, NJ:Erlbaum Associates. 3-23
94. Arai, S. & Mayekawa, S. (2011) A comparison of equating methods and linking designs for developing an item pool under item response theory, *Behaviormetrika*, 38, 1, 1-16
95. Bachman, L. F. (1990) *Fundamental Consideration in Language Testing*. Oxford University Press.

96. Bachman, L. F. & Palmer, A. S. (1996) *Language testing in practice*. Oxford University Press [L.F.バックマン／A.S.パーマー著，大友賢二／ランドルフ・スラッシャー（監訳）（2000），＜実践＞言語テスト作成法，大修館書店]
97. Baranowski, R.A. (2006) Item Editing and Editorial Review. In Downing, S.M. and Haladyna, T.M., (Eds.) *Handbook of Test Development*, 15, 349-358
98. Burton, E. & Burton, N.W. (1993) The effect of item screening on test scores and test characteristics. In P.W.Holland & H.Wainer., (Eds.) *Differential item functioning*. Hillsdale, NJ:Erlbaum Associates. 321-335
99. Camilli, G. & Shepard, L. A. (1987) The inadequacy of ANOVA for detecting test bias. *Journal of Educational Statistics*, 12, 87-99
100. Chapelle, C.A. (2008) The TOEFL validity argument. In C. A. Chapelle, M. K. Enright & J. M. Jamieson., (Eds.) *Building a validity argument for the Test of English as a Foreign Language*. New York Routledge. 319-352
101. Cole, N. S. & Moss, P. A. (1993) Bias in Test Use, In R. L. Linn., (Eds.) *Educational measurement 3rd ed.*, New York: Macmillan, 211-220
102. Downing, S.M. (2006) Twelve Steps for Effective Test Development. In Downing, S.M. & Haladyna, T.M., (Eds.) *Handbook of Test Development*, 1, 3-26
103. Engelhard, G. (1990) Accuracy of Bias Review Judges in Identifying Differential Item Functioning on Teacher Certification Tests, *Applied measurement in education*, 3(4), Lawrence Erlbaum Associates. 347-360
104. Holland, P.W. & Thayer, D.T. (1988) Differential item functioning and the Mantel-Haenszel procedure. In H.Wainer & H.I.Braun., (Eds.) *Test Validity*, Hillsdale, NJ:Lawrence Erlbaum. 129-145
105. Hubley, A.M. & Zumbo, B.D. (1996) A dialectic on validity: Where we have been and where we are going. *The Journal of General Psychology*, 123, 3, 204-215
106. Kane, M. (2006) Validation. In R. Brennan., (Eds.) *Educational Measurement 4th ed.*, Westport. CT: American Council on Measurement in Education and Praeger Publishers, 17-64

- 107.Linn, R.L. (2006) The Standards for Educational and Psychological Testing: Guidance in Test Development. In Downing, S.M. & Haladyna, T.M., (Eds.) *Handbook of Test Development*, 2, 27-38
- 108.Livingston, S. A. (2006) Item Analysis. In Downing, S.M. & Haladyna, T.M., (Eds.) *Handbook of Test Development*, 19, 421-444
- 109.Lord, F. M.(1952) *A Theory of Test Scores.*, Psychometric Monograph, 7, Psychometric Society.
- 110.Lord, F. M. (1980) Applications of item response theory to practical testing problems. Hillsdale, NJ:Lawrence Erlbaum.
- 111.Messick, S. (1989) Validity. In R. L. Linn., (Ed.) *Educational measurement 3rd ed.*, New York: Macmillan, 13-103
- 112.Messick, S. (1996) Validity of performance assessment. In Philips, G., *Technical issues in large-scale performance assessment*. Washington, DC: National Center for Educational Statistics. 1-18
- 113.Muraki, E. (1999) Stepwise Analysis of Differential Item Functioning Based on Multiple-Group Partial Credit Model. *Journal of Educational Measurement*, 36(3), 217-232
- 114.Nering, M.L. & Ostini, R. (Eds.) (2010) *Handbook of Polytomous Item Response Theory Models*. Routledge.
- 115.Oller, J. W. Jr. (1979) *Language Test at School*. Longman Group Ltd. [堀口俊一 (訳者代表) (1994) , 言語テスト, 秀文インターナショナル]
- 116.O'Sullivan, B. & Weir, C.J. (2011) Test development and validation. In B. O'Sullivan., (Ed.) *Language testing: Theories and practices*, Basingstoke: Palgrave, 13-32
- 117.Ross. S. J. & Okabe, J. (2006) The Subjective and Objective Interface of Bias Detection on Language Tests. *International Journal of Testing*, 6(3), 229-253
- 118.Toulmin, S. E. (2003) *The uses of argument* (updated ed.) Cambridge, UK: Chambridge Univercity Press.
- 119.Van de vijver, F.J.R. & Leung, K., (1997) Methods and data analysis of comparative research 2nd ed. In: Berry, J.W., Poortinga, Y.H. & Pandey, J.,

- (Eds.) *Handbook of Cross-cultural Psychology*, 1. Allyn & Bacon, Boston, 257-300
120. Weir, C. J. (2005) Limitations of the Common European Framework for developing comparable examinations and tests. In B. O'Sullivan., (Ed.) *Language Testing*, 22(3), 1-20

資料

資料1 第4章で使⽤したビジネス日本語 Cds

【予備調査】

ビジネス日本語 Can-do Statements

2009(第1期)第3フェーズ		氏名: 記入日:	年	月	日
		国 籍:	大 学 名:		

日本語でどんなことができますか。最も当てはまるものを次の1～5の中から選んで、○をつけてください。

	5 かなり 問題なくできる (100%できる)	4 だいたいできる (70～80%できる)	3 ある程度できる (50%できる)	2 あまりできない (20～30%できる)	1 ほとんどできない (0%)
聴く					
01	5	4	3	2	1
02	5	4	3	2	1
03	5	4	3	2	1
04	5	4	3	2	1
05	5	4	3	2	1
06	5	4	3	2	1
07	5	4	3	2	1
08	5	4	3	2	1
09	5	4	3	2	1
10	5	4	3	2	1
読む 合 計					
読む					
01	5	4	3	2	1
02	5	4	3	2	1
03	5	4	3	2	1
04	5	4	3	2	1
05	5	4	3	2	1
06	5	4	3	2	1
07	5	4	3	2	1
08	5	4	3	2	1
09	5	4	3	2	1
10	5	4	3	2	1

【本調査】

ビジネス日本語 Can-do Statements (日本語版)

日本語でどんなことができますか。
最も当てはまるものを次の1～5の中から選んで、マークしてください。

1	2	3	4	5
ほとんどできない 0%	あまりできない (20～30%できる)	ある程度できる (50%できる)	だいたいできる (70～80%できる)	問題なくできる (100%できる)

聴 く	
1	上司から与えられた自分の担当業務に関する指示が理解できる。
2	社内で同僚同士が話している会話を近くで聞いて、どんな話題について話しているかわかる。
3	会議や打合せで報告を聞き、何が重要なポイントなのかを理解できる。
4	自分の担当分野に関する上司や先輩からのアドバイスやコメントの内容が理解できて、仕事に生かせる。
5	会議のときに、議論の内容を聞きながら、賛成や反対など主張の違いがわかる。
6	留守番電話のメッセージを聞いて、相手の要求や説明が理解できる。
7	社外の人からの苦情や要求の内容が理解できる。
聴きながら読む／読みながら聴く	
8	自分の担当分野についてのプレゼンテーションの内容が理解できる。
9	テレビの経済やビジネスに関するニュースの内容を理解することができる。
10	ビジネスに関する研修及び講演の内容が理解できる。
読 む	
11	広告、パンフレット・カタログなどから必要な情報を見つけることができる。
12	自分の仕事に必要な情報をインターネットで検索することができる。
13	自分の担当分野に関係する記事やニュースが理解できる。
14	社内ですら出されたお知らせや、閲覧されている文書、仕事の指示を書いたメモの内容が2、3分読んで理解できる。
15	社外から来たビジネス文書を2、3分で読んで、内容によって分類できる。
16	社内ですら提案された新しい企画についての案議書を読んで、企画の概要や提案理由がわかる。
17	社外の人と数回やり取りした文書やメールを、全部読んで、新しく決まったこと、変更があったことは何かわかる。
18	自分の担当業務に関する様々なビジネス文書を比較し、情報の優先順位をつけることができる。
19	丁寧な印象を与えるために婉曲的な表現で書かれた社外から来たビジネス文書の意図がわかる。
20	専門用語を使用した契約文書の内容が細部まで間違いなく理解できる。

話 す	
21	社内と社外で話している相手との関係を考慮して、場合に応じた適切な話し方を選ぶことができる。
22	敬語に注意しながら上司や先輩に依頼ができる。
23	専門的で複雑な内容でも、電話で要件をわかりやすくまとめて伝えることができる。
24	自分の担当している商品や企画について、長くなりすぎないように簡潔に、効果的にプレゼンテーションできる。
書 く	
25	同僚や上司に業務を依頼したり、業務の内容を伝えたりしたいときに、あまり時間をかけないで、さっとメールやメモを書くことができる。
26	上司や先輩に敬語を使用して業務の報告書を書くことができる。
27	自分が担当している商品や企画を他社や顧客に宣伝し、売り込む文書を書くことができる。
28	プレゼンテーションのための資料をわかりやすく作ることができる。
聴いて話す	
29	営業などの業務で社外の会社を訪ねたときや、社外の会議で集めた情報をまとめて、社内の人に報告することができる。
30	会議や打合せで議論の内容を理解して、自分の意見や疑問点を言うことができる。
31	上司に対して疑問点や問題点を感じたときに、婉曲的な言い方だが、正確に自分の思っていることを説明できる。
32	電話で話した内容の要点をまとめ、相手に確認することができる。
33	取引先の人と話しているときに、よく理解できないことがあれば、相手の気持ちを考えながら、丁寧に聞き返すことができる。
34	社外からの依頼や苦情の内容を理解して、相手の気持ちを考えながら、きちんと対応できる。
聴いて書く	
35	会議やプレゼンテーションを聞きながら要点をメモすることができる。
36	会議に出席して、会議の記録をとり、あとで正式な報告書を書くことができる。
読んで書く	
37	苦情や問い合わせのメールや文書が来たときに、相手の気持ちを考えながら、丁寧な返事を書くことができる。
38	新聞、インターネットなどから集めた情報をまとめて、報告書を書くことができる。
39	自社の競争相手の競合他社に関する情報を集め、自社と比較した報告書を書くことができる。
40	社外から依頼の文書が来たときに、相手の気持ちを考えながら、丁寧な断りの手紙やメールを書くことができる。

資料2 第5章で使用した実験テストフォーム

BTT ビジネス日本語能力テスト 実験テスト 02

もろたいようし

JLRT 問題用紙

JLRT Test Booklet

注意

(1) 試験監督者の指示があるまで、問題用紙を開けないでください。

(2) テストが始まったら、テストの途中で紙を返すことはできません。

(3) テストが終わった後、試験監督者の指示があるまで、紙を返さないでください。

(4) テスト中は、音源はでません。

(5) テープは一度しか使えないので、よく聞いてください。

(6) この問題用紙にメモをとることがあります。

(7) テスト開始後、試験監督者の指示に従ってください。

(8) テスト開始後、試験監督者の指示に従ってください。

(9) テスト開始後、試験監督者の指示に従ってください。

(10) テスト開始後、試験監督者の指示に従ってください。

(11) テスト開始後、試験監督者の指示に従ってください。

(12) テスト開始後、試験監督者の指示に従ってください。

(13) テスト開始後、試験監督者の指示に従ってください。

(14) テスト開始後、試験監督者の指示に従ってください。

(15) テスト開始後、試験監督者の指示に従ってください。

(16) テスト開始後、試験監督者の指示に従ってください。

(17) テスト開始後、試験監督者の指示に従ってください。

Attention

(1) Do not open this booklet until a test administrator instructs you to do so.

(2) You may not leave the room during the test.

(3) These will be a short intermission with music for about one minute during the test, but do not leave your seat. If you feel ill at any time, raise your hand.

(4) The tape will be played one time only so listen carefully.

(5) You may write notes in this booklet.

(6) Test questions may not be tape-recorded and this booklet may not be taken from the test room.

(7) Neatly write your registration number and name at the bottom of this page.

Write your name in alphabetical characters.

登録番号

Registration Number

名前

Name

229

だい ぶ
第1部
PART I

ちようかい
聴解テスト
Listening Test

しじ
指示があるまで、開けないでください。

もんだいようし
問題用紙に書かれたものを見ながら、
おとこ
音声を聴いて、質問に答えてください。

1

2

だい ぶ ちようかい
第1部 聴解テスト
PART 1 Listening Test

セクション1
Section 1

練習



3

第1部 聴解テスト:セクション1

1.



2.



3.



4.



5.



6.



4

Part1 Listening Test: Section 1

だい ぶ ちょうかい
第1部 聴解テスト
PART 1 Listening Test

セクション2
Section 2

練習



8

1.



2.



3.



4.



5.



6.

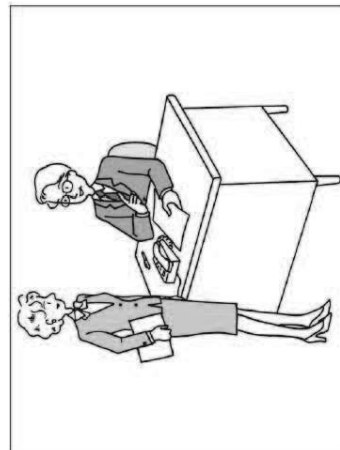


6

第1部 聴解テスト
PART 1 Listening Test

セクション3
Section 3

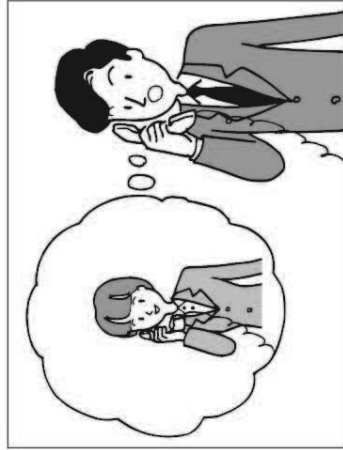
無題



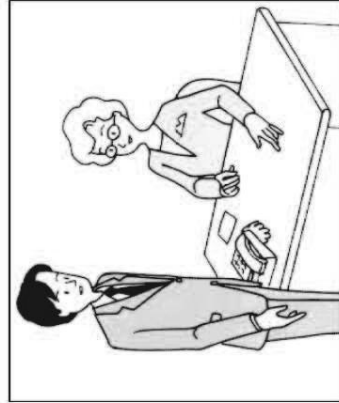
7

第1部 聴解テスト：セクション3
Part1 Listening Test: Section 3

1.



2.

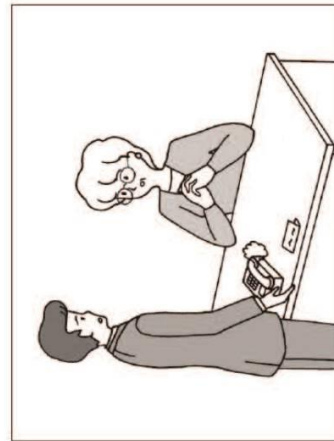


8

3.

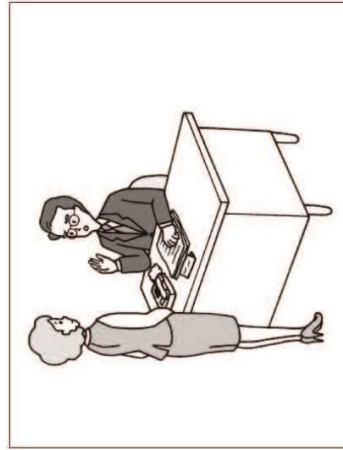


4.

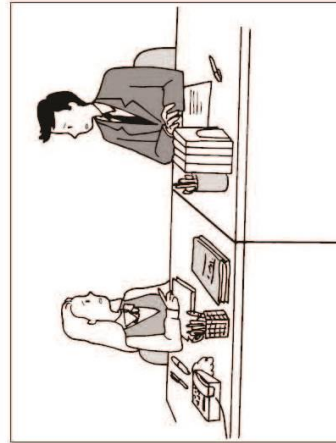


9

5.

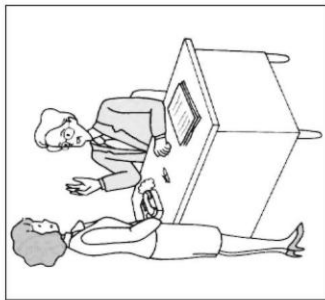


6.

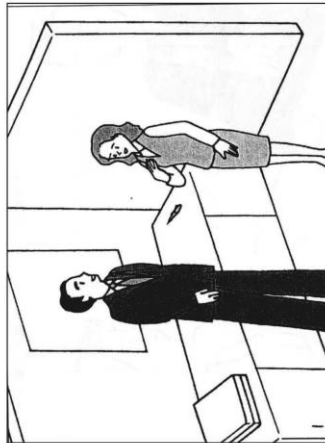


10

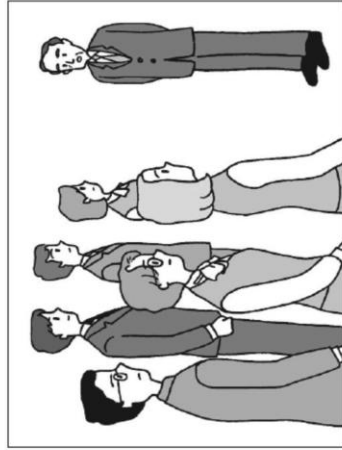
7.



8.



9.



だい ぶ
第2部
PART2

ちようどつかい
聴読解テスト
Listening & Reading Test

しじ
指示があるまで、開けないでください。

もふだいようし
問題用紙に書かれたものを見ながら、
おんせう
音声_きを聴いて、質問に答えてください。

13

14

だい ぶ
第2部 聴読解テスト
PART 2 Listening & Reading Test

セクション1
Section 1

練習



1. 3 時頃、部長の室内から電話がありました。
2. 3 時頃、部長の女房から電話がありました。
3. 3 時頃、部長の妻から電話がありました。
4. 3 時頃、部長の奥様から電話がありました。

15

第2部 聴読解テスト：セクション1 Part2 Listening & Reading Test : Section 1

1.



1. 輸送ルートの変更
2. 販売システムの変更
3. 新製品の仕様の変更
4. 生産ラインの見直し

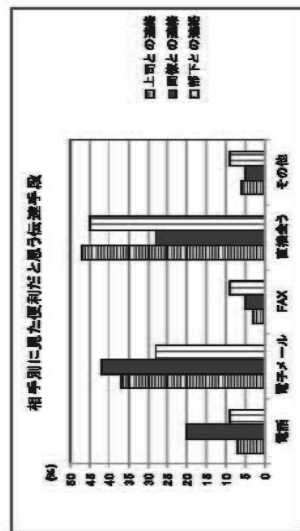
2.



1. 製品のスペック
2. 製品のデザイン
3. 製品のターゲット
4. 新製品のネーミング

16

၈၅



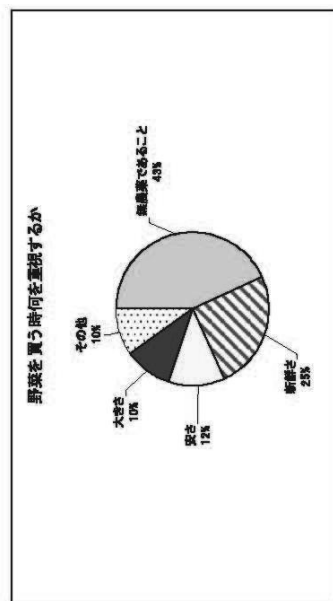
1. 電話
2. メール
3. ファックス
4. 直接会う

7

FROM : サトウ電機 山田 和男
TO : 田中一雄様
件名: 明日のアポイントの件
日付: ****年**月**日
田中様、
明日、1階に弊社では参りますがお断言でしたが、工事で事務室が非常
に生じ、急遽、出張することになってしまいました。
悪くしていいますのは明日の2時頃とさせて頂いております。
大塚様、野原さまですが、断断を調整していただければと
うか、できれば午前から受領がありますので、その前には参
ればと思います。
なお、出張してはお客様の予定ですが、その際、8時迄で明日の予定
が参ります。
以上、よろしくおねがいいたします。
山田和男

1. 1時過後
2. 2時過後
3. 3時過後
4. 4時過後

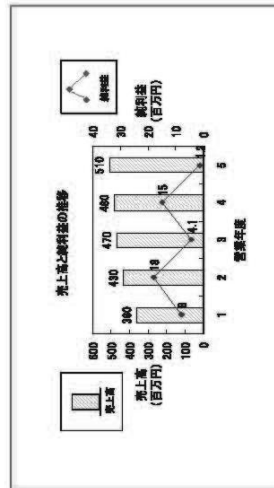
5.



1. 安全性
2. 鮮度
3. 値段
4. 形

19

6.



1. 売上高は減少している。
2. 純利益は増加している。
3. 売上高は増減をくり返している。
4. 純利益は増減を繰り返している。

20

7.

建物賃貸契約書

第1条 賃貸借料について

第2条 光熱費および水道料について

第3条 火災保険加入について

第4条 中途解約について

8.

中田部長へ

12時半ごろ、東都電機の鈴木様から電話あり。
明日の打ち合わせには、担当の方はお二人で、弁護士を一人、同伴されとのこと。

山田 受

1. 1人

2. 2人

3. 3人

4. 4人

9.

〇〇〇〇年度 社員募集要項

募集人員：正社員 6 名
年 齢：大卒・35 歳位まで
給 与：会社面談
休日休暇：完全週休二日制・有給休暇あり

応 募：
以下の書類をご郵送ください。書類選考の上、面接のご連絡を致します。
応募書類は返却いたしません。

・履歴書（写真貼付け・当社指定用紙）＊
・応募理由書（形式自由）

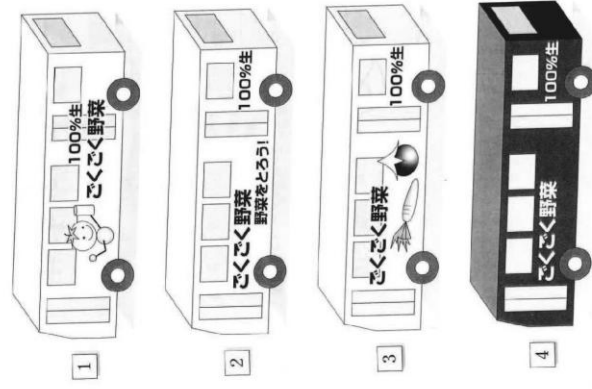
＊当社指定用紙は下記採用担当に電話またはファックスにてご請求ください。

応募締切：〇〇年〇月〇日必着

書類請求・送付先：
株式会社ミスト
〒〇〇〇-〇〇〇〇 東京都〇〇区〇〇北〇-〇〇
管理部人材開発課 採用担当
TEL.03-〇〇〇〇-〇〇〇〇
FAX.03-△△△△-△△△△

1. 応募書類を送る
2. 指定の用紙を取り寄せる
3. 応募理由を知らせる
4. 面接可能日を知らせる

1.



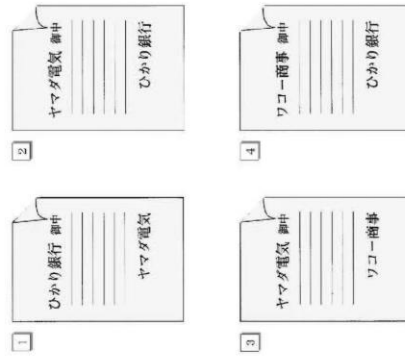
26

第2部 聴読解テスト
第2部 聴読解テスト

PART 2 Listening & Reading Test

セクション2
Section 2

練習



25

2.

1

シ
よりていあり。
来週月曜日の時間、
2時間遅らせてほしい、
とのこと。

2

シ
よりていあり。
来週月曜日の時間、
2時間遅らせてほしい、
とのこと。
返事は今日中に
ほしいそうです。

3

シ
よりていあり。
来週月曜日の時間、
1時間遅らせてほしい、
とのこと。

4

シ
よりていあり。
来週月曜日の時間、
1時間遅らせてほしい、
とのこと。
返事は今日中に
ほしいそうです。

3.

1

2

3

4

新感覚野菜ジュース
街頭キャンペーン用アンケート項目（案）

(1) 価格設定は？
☐ 高い ☐ 普通 ☐ 安い

(2) パッケージのデザインは？
☐ 良い ☐ 普通 ☐ 悪い

(3) 味は？
☐ おいしい ☐ 普通 ☐ まずい

(4) ネーミングは？
☐ 良い ☐ 普通 ☐ 悪い

4.

○○○○年○月○日
企画書

ポイント・キャンペーン
企画（案）

● 1 会員紹介キャンペーン

● 2 新規入会ポイント2倍キャンペーン

● 3 平日・時間限定ポイント2倍キャンペーン

● 4 平日・ポイント2倍キャンペーン

29

5.

		1		2		3		4	
		プラン1		プラン2		プラン3		プラン4	
	総工賃 (一箇月分)		総工賃 (一箇月分)		総工賃 (一箇月分)		総工賃 (一箇月分)		総工賃 (一箇月分)
基礎	¥200	前	¥250	後進	¥200	前	¥250	前	¥250
外壁	¥200	プランA	¥200	プランB	¥150	プランB	¥150	プランB	¥150
合計	¥400		¥450		¥350		¥400		¥400

仕 様 書

〇〇〇〇年〇月〇日

株式会社
ヤマダ印刷

30

6.

関係者各位

○○○○年○月○日

1
ABC株式会社
営業本部長 松野太郎

2
ABC株式会社
副社長 山田一郎

3
JFC銀行
頭取 佐藤浩氏

4
ミナト物産株式会社
営業課長 大山和男氏

販売プラットフォームのご案内

下記の通り、販売プラットフォームを開催いたしますのでご案内いたします。

記

日時：○○○○年×月×日 10:00～17:00

場所：ABC株式会社 大会議室

次第：

1. 開会の辞

2. 挨拶

3. 講演「金融戦略」

4. 講演「最先端技術について」

5.

6.

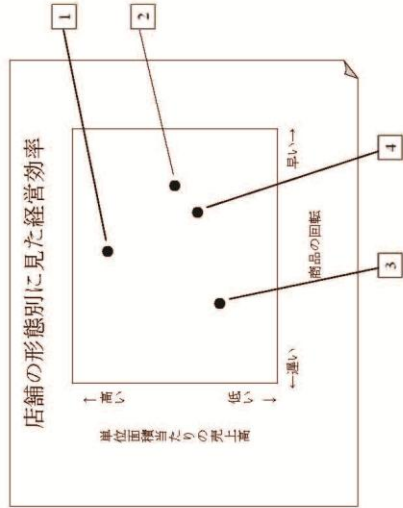
7. 閉会の辞

総観会 ○○ホテル ××の間にて

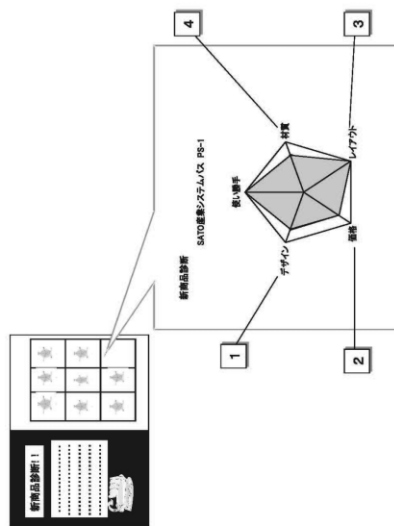
(昼食休憩)

以上

7.

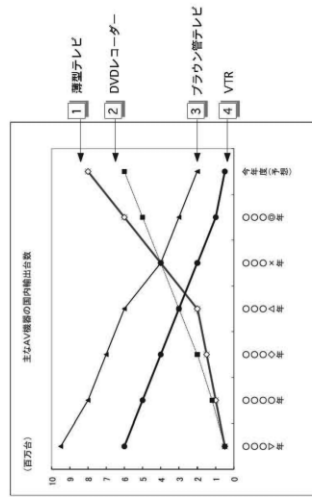


8.



33

9.



34

第3部
PART2

読解テスト
Reading Test
セクション1・2・3
Section 1 / 2 / 3

指示があるまで、開けないでください。

音声はありません。
問題を読んで答えてください。

第3部 聴解テスト：セクション1 Part3 Reading Test：Section 1

次の文の _____ に入る正しいことばを1、2、3、4の中から1つ選んでください。

1. _____ 所在地なんて、わざわざ、取引先に _____ ほどのことでもない。インターネットで調べればすぐわかる。

- 1. 聞いかける
- 2. 聞いたです
- 3. 聞いわせる
- 4. 聞いつめる

2. _____ 非常に多忙で、これ以上仕事を増やすと大変なのだが、条件によっては、_____

- 1. 受けるつもりもない
- 2. 受けるはずもない
- 3. 受けないことはない
- 4. 受けたくもない

3. _____ 残念なご報告ですが、当店は来月末日を _____、閉店させていただきます。

- 1. もちまして
- 2. まちまして
- 3. とりまして
- 4. きりまして

4. _____ 注文内容が違っているとお電話いただいた件ですが、当方には納得 _____ 点がございませ

- 1. いたしたい
- 2. いたすべき
- 3. いたしかねる
- 4. いたしやすい

<p>第3部 聴解テスト：セクション1</p> <p>Part3 Reading Test: Section 1</p> <p>5. 我が社の売上は、3期連続で、低迷が深まっている。 取引先を変えた____、売上不振が好転するとは限らない。</p> <p>1. あげく 2. すえに 3. とこで 4. ばかりに</p> <p>6. 御社の社長には____ご愛顧をいただき、厚く御礼申し上げます。</p> <p>1. ひとかどの 2. ひとすじの 3. ひとおもいに 4. ひとかたならぬ</p>	<p>第3部 聴解テスト：セクション2</p> <p>Part3 Reading Test: Section 2</p> <p>次の文の____に入る正しいことばを1, 2, 3, 4の中から1つ選んでください。</p> <p>1. A: 田中君、そのところ、ちょっと分りにくいんだけど。 B: すみません。では、もう一度、_____。</p> <p>1. 説明されますか 2. 説明なさいますか 3. 説明していただけますか 4. 説明させていただきますか</p> <p>2. 客: 先週、おたくで買ったパソコン、変な音をするんだけど。 店員: 申し訳ございません。____、ただいま田中の音を黙んでまいります。</p> <p>1. 私ではわかりかねますので 2. 私には関係ありませんので 3. 私の仕事ではありませんので 4. 私は運達させていただきますので</p> <p>3. _____。取締役にご伝言をお願いできないでしょうか。</p> <p>1. 恐縮ですが 2. 恐縮しますが 3. 恐縮しましたが 4. 恐縮するんですが</p> <p>4. 会議の結果がわかり次第、早急に対策を____必要がある。</p> <p>1. かかる 2. 使う 3. あやつる 4. 擇じる</p>
--	---

<div> <div>第3部 聴解テスト：セクション2</div> <div>Part3 Reading Test：Section 2</div> </div> <div> <p>5. 最近の若手社員の言葉の乱れについては、_____ものがある。</p> <ol style="list-style-type: none"> 1. 見過ごす 2. 見過ごすべき 3. 見過ごさない 4. 見過ごせない <p>6. 現在、当社で開発中のAX-300のスペックについては、当社にお送りいただければ、その詳細を_____。</p> <ol style="list-style-type: none"> 1. ご覧ください 2. ご覧いただきたい 3. ご覧になります 4. ご覧になれます </div>	<div> <div>第3部 聴解テスト：セクション3</div> <div>Part3 Reading Test：Section 3</div> </div> <div> <p>次の文章を読んで、質問に答えてください。 答えは、1、2、3、4の中から1つ選んでください。</p> <p>1. 取引先から、次のような文書が届きました。 工場の機械がストップしている直後の原因は、何だと言っていますか。</p> <div> <div>株式会社〇〇〇〇 営業部長××××様</div> <div>〇〇〇〇年〇月〇日</div> <div>株式会社△△△ 営業部□□□□</div> <div>弊社〇〇工場の機械一時中止のお知らせ</div> </div> <p>拝啓 時下ますますご清栄のこととお慶び申し上げます。</p> <p>この度、弊社〇〇工場の機械を〇月×日より一時中止する件につきまして、ご通知申し上げます。と申しますのも、皆様、ご存知のように、ただ今世界的に鉄鉱石の需要が高まっており、鋼板の入荷に支障をきたしております。韓国からの輸入に加えてオーストラリア、ブラジルからの調達が検討しておりますが、現在のところ見通しは立っておりません。</p> <p>誠に残念ながら、弊社従業員も自花件機を余儀なくされている状況でございます。日頃ご取引いただいております各社に多大なるご迷惑をおかけし、誠に申し訳ありません。</p> <p>機械の目処が立ち次第、直ちにご連絡いたしますので、今後ともよろしくお引き立てくださるようお願い申し上げます。</p> <div>敬具</div> </div> <div> <ol style="list-style-type: none"> 1. 人材不足のため 2. 鉄鉱石不足のため 3. 原材料不足のため 4. 資金不足のため </div>
--	--

第3部 聴読解テスト：セクション3

Part3 Reading Test：Section 3

2.
取引先から次のような文書が届きました。カタログ送付が遅れているのはなぜだと書いていますか。

ご担当者様 各位

〇〇〇〇年〇月〇日

XXカンパニー
営業推進部 〇〇〇〇

カタログ送付遅延について

拝啓 時下益々ご清栄のこととお慶び申し上げます。平素は弊社事業につぎまして、格別なるご高配を賜り、厚く御礼申し上げます。さて、先日は弊社のカタログをご送付いただき、誠にありがとうございます。通常であればご依頼から8日以内に発送するシステムとなっておりますが、現在新版カタログとの切り替え時期にあるため、一時送送を阻害させていただいております。新規格送時期につきましては、入荷次第早急にお送りさせていただきますが、これより一週間の猶予をいただければ幸いです。

なお、ご参考までに、現行のカタログを送付させていただきます。弊社の都合により、ご送達をおかけすることとなり、大変申し訳ございません。何卒ご寛容のほどご了承いただけますよう、お願いいたします。

敬具

第3部 聴読解テスト：セクション3

Part3 Reading Test：Section 3

3.
取引先から次のような文書が届きました。商品について、何をしてほしいと言っていますか。

〇〇商事株式会社
営業部業務開発室御中

〇〇〇〇年〇月〇日

△△株式会社
販賣部販賣促進課顧客サービス係

拝啓 貴社ますます御清栄のこととお喜び申し上げます。さて、この度は、弊社御品『炊飯鍋・たけるくん』をお買い上げいただき、心より御礼申し上げます。先日も直送お品を致しました通り、おかげ様で多くの皆様にご愛用いただき、大変、ご好評をいただいております。品質に關しまして、配付の自信を持っております。弊社では、日本のみならず海外にもお認めしております。今後も、同製品の品質改良には、たゆまぬ努力を続けて参る所存でございますが、同製品につきまして、使い勝手等、ユーザー各位に御用いたたくと共に、ご要望等、お伺いしたいと思います。

つきましては、別添のアンケート用紙にご記入の上、ご送付いただけましたら幸いです。

今後とも変わらぬご支援、ご愛顧を賜りますようお願い致します。取り急ぎ、お買い上げの御礼かたがたご挨拶申し上げます。

敬具

1. カタログが間違っていたため

2. カタログが足りないため

3. カタログができていないため

4. カタログを新しくしているため

250

4. 次のような社内文書が来ました。アンケートにはどのようにして回答してほしいと言っていますか。

社員各位

0000年〇月〇日

商品開発部

アンケートご回答のお願い

平素より、当商品開発部にご協力いただき、深謝いたします。さて、この度、先般の役員会決定を受け、販路に先立ち、新開発商品についてアンケートを実施する運びとなりました。つきましては、下記要項にてご回答いただをたく存じます。ご多忙中、恐縮ではございますが、ご協力の程、よろしくお願ひ申し上げます。

記

1. 回答方法：別紙調査票にご記入ください。

2. 回答期限：0000年〇月×日
回収に伺います。

3. 結果発表：0000年×月×日に公表

※お問い合わせ等は、内線（××××）、もしくは、メールにて、商品開発部担当・田中までお願いいたします。
tanaka@xxx.co.jp

以上

1. 別紙書式で回答してほしい
2. メールで回答してほしい
3. 電話で回答してほしい
4. 対面で回答してほしい

5. 次の記事は、新しい情報システムに関するものです。一人あたりの売上高がアップする根拠は何ですか。

営業部門 出社不要に

ここは、営業部門に新しい情報システムを導入し、週に一回程度出社するだけでよい勤務体制にする。また、事業部長の一人を意に一人づつ業務用パソコンと携帯電話を持たせ、会社に居なくても営業報告などの仕事ができるようにした。直営、営業活動に使える時間を増やして、一人あたり売上の五〇パーセントを目ざす。

新たな勤務体制とは、営業担当者は営業部全体の合算売上のために週に二、三回オフィスに集まるほかには出社しないこと。営業報告などの事務作業は外出先から電子メールで済ませる。通勤や顧客とオフィスの間の往還にかかっていた時間も営業活動に使われるようになり、営業質員もよりよく売上高を伸ばすことが可能になる。

まず電子コンポーネント販売営業部の担当者百六十人から導出した順次、順次を広げ二年後には国内の営業部門五千六百に適用する考え。

1. 直接営業活動に使える時間が増えるから
2. 国内営業部門の人員が5000人になるから
3. 電子コンポーネント販売営業部の担当者が100人になるから
4. 営業報告などの事務作業が必要なくなるから

6、
取引先から、次のような文書が届きました。取引先が取引を断る主な理由は、何ですか。

シークー株式会社
購買部 鈴木一広様

〇〇〇〇年〇月〇日

株式会社 エイビー
営業部 田中 徹

取引ご辞退の件

拝復 貴社ますますご繁栄のこととお喜び申し上げます。

さて、このたびは弊社製品〇〇につきまして見解書のご依頼をいただき、誠に有難うございました。

その御売価格と著しい開きがありますため、まことに残念ながら、今回のご依頼には応じかねます。

その背景といたしまして、今年に入って原料価格が急騰しましたものの、弊社では価格を抑え置いており、これ以上値引きいたしましたら、原価を割ってしまうような状況です。また、各販売店様とも、現行の価格にてお取引頂いておりますゆえ、御社のご要望にのみお応えするというわけにもまいりません。

以上のような事情でございますので、なにとぞご了承いただけますようお願い申し上げます。

まずはお詫びかたがたご返事申し上げます。

敬具

- 1、原料価格が上がったため
- 2、売価価格が下がったため
- 3、他社との取引を優先するため
- 4、製造を一時中止しているため

7、
次の文書は、販売部長からのお知らせです。どんな販売レポートを提出するように書いていますか。

特約店各位

〇〇〇〇年〇月〇日

△△本社
販売部長□□□□

販売報告会について

下記の通り、販売報告会を開催することになりました。

各位とも年度当初の目標数と販売実績を照らし合わせ、実証的なデータに基づいた、上半期の販売レポートを提出してください。以下の諸点を念頭に置き、報告するようにしてください。

- ・前年度の反省に基づいた販売促進が行われたか
- ・前年度とは異なる戦略がとられたか
- ・上半期の販売実績データの推移
- ・下半期の販売戦略について

レポート提出締切：6月23日

（メールに添付して送信のこと）

配

日時：〇〇〇〇年〇月〇日 午後1時～6時

場所：本社 大会議室

以上

- 1、前年度の販売実績の検証
- 2、前年度の販売戦略の目標
- 3、上半期の販売実績の検証
- 4、上半期の販売戦略の目標

8. 日米技術交渉に関する対談記事です。交渉のありかたについて専門家（山田氏）は何と書いていますか。

鈴木：それは、日米企業間の交渉は、以前に比べて、かなり複雑化しています。山田：それは、文化の違いが大きいからです。交渉では、通常かなりの情報開示が行われます。山田：それは、文化の違いが大きいからです。交渉では、通常かなりの情報開示が行われます。山田：それは、文化の違いが大きいからです。交渉では、通常かなりの情報開示が行われます。

鈴木：それは、日米企業間の交渉は、以前に比べて、かなり複雑化しています。山田：それは、文化の違いが大きいからです。交渉では、通常かなりの情報開示が行われます。山田：それは、文化の違いが大きいからです。交渉では、通常かなりの情報開示が行われます。

- 1. 日本企業の意思決定をスピードアップさせるべきだ
- 2. 日本企業の意思決定の遅さも、年報義務契約も全てケース・バイ・ケースで対応すればよい
- 3. 年報義務契約は当然である
- 4. コンピューター技術のようなものは時間をかけて交渉すべきだ

9. 次のような募集が来しました。どんなモニターを募集していますか。

社員各位

〇〇〇〇年〇月〇日

広報部

ご承知の通り、当社は、先週、健康飲料「ガーリーパワー」を発売いたしました。発売当日から、テレビ各局で同商品のコマーシャルを大々的に展開中です。当社の販売促進活動につきましては、すでに業界各方面から様々な感想、ご意見をいただいておりますが、この度、社員の皆様の声をお聞かせいただきたいというこ

とで、下記の通り、モニターを募集いたします。ふるってご応募いただきますようお願いいたします。

記

モニター対象：「ガーリーパワー」のテレビコマーシャル

モニター時期：〇月〇日（土）より〇月×日（日）

モニター事項：①同業他社商品と比較したアピール度

②同業他社商品と比較した好感度

③改善点

以上の3点につきまして、御署名および氏名を明記の上、メールまたはFAXで、広報部までご意見をお寄せください。

締切は〇月△日（水）とします。

以上

- 1. 自社製品
- 2. 他社製品
- 3. テレビ広告
- 4. テレビ番組

資料3 第6章で使ったDIF分析用プログラムコード (PERSCALE)

【性別によるDIF】

```
E&Blue-BJT_gender.PSL - E-BJT&Blue-BJT DIF ANALYSIS USING PARTIAL
CREDIT MODEL

                GENDER DIF FOR 160 ITEMS except item133,145

>COMMENT;

ALL ITEMS HAVE TWO CATEGORIES

>FILE

DFNAME='E&Blue-BJT_gender.txt',NFNAME='E&Blue-BJT_gender_dammy.txt',
SAVE;

>SAVE    PARM='E&Blue-BJT_gender.PRM',
INFORMATION='E&Blue-BJT_gender.INF';

>INPUT    NID=4,MGROUP=2,NTOTAL=158;
(4A1, 1X,1A1, 1X,132A1,1X,11A1,1X,15A1)

>TEST    TNAME='E&Blue-BJT',NBLOCK=1,ITEM=(1(1)158);

>BLOCK    NITEMS=158,NCAT=2,ORIGINAL=(0,1);

>MGROUP    DIF=(1,1,0,0), GNAME=('FEMALE','MALE'), GCODE=('1','2');

>CAL    LOGISTIC, PARTIAL, NQPTS=30, CYCLE=(20,1,1,1,1,1), POSTERIOR,
spripr, tpripr;
```

【居住地によるDIF】

```
E&Blue-BJT_naigai.PSL - E-BJT&Blue-BJT DIF ANALYSIS USING PARTIAL
CREDIT MODEL

                NAIGAI DIF FOR 160 ITEMS except item133,145

>COMMENT;

ALL ITEMS HAVE TWO CATEGORIES

>FILE

DFNAME='E&Blue-BJT_naigai.txt',NFNAME='E&Blue-BJT_naigai_dammy.txt',
```



```

SAVE;
>SAVE    PARM='E&Blue-BJT_naigai.PRM',
INFORMATION='E&Blue-BJT_naigai.INF';
>INPUT   NID=5,MGROUP=2,NTOTAL=158;
(5A1, 1X,1A1, 1X,132A1,1X,11A1,1X,16A1)
>TEST    TNAME='E&Blue-BJT',NBLOCK=1,ITEM=(1(1)158);
>BLOCK   NITEMS=158,NCAT=2,ORIGINAL=(0,1);
>MGROUP  DIF=(1,1,0,0), GNAME=('KOKUNAI','KOKUGAI'), GCODE=('1','2');
>CAL     LOGISTIC, PARTIAL, NQPTS=30, CYCLE=(20,1,1,1,1,1), POSTERIOR,
spripr, tpripr;

```

資料4 第6章で使用した実験テストフォーム

調査用紙（聴解1）

聴解 §1 描写問題

このセクションでは、写真または絵を見て答えてください。4つの文を読みます。写真または絵の内容を、最も正しく表している文は、どれですか。1、2、3、4の中から1つだけ選んでください。

実験テスト

#01 1. 問題名称：機材チェック

【視覚情報】



【音声情報】

1. 女の人が議事録を作成しています。
2. 女の人が資料を配付しています。
3. 女の人が出席者に質問しています。
4. 女の人がパソコンをチェックしています。

実験テスト

#02 2. 問題名称：名刺交換

【視覚情報】



【音声情報】

1. 商品を回収しています。
2. 名刺を交換しています。
3. 品物を入荷しています。
4. 契約書にサインしています。

実験テスト

#03 3. 問題名称：社員紹介

【視覚情報】



【音声情報】

1. 上司が、仕事の説明をしています。
2. 上司が、仕事の引継ぎをしています。
3. 上司が、新商品を紹介しています。
4. 上司が、新人を紹介しています。

調査用紙（聴解1）

実験テスト

#04 4. 問題名称：ビール部内会議

【視覚情報】



【音声情報】

1. 社員たちが、社員食堂で食事をしています。
2. 社員たちが、レストランでパーティーを開いています。
3. 社員たちが、会議室で部内会議を開いています。
4. 社員たちが、ビールを販売しています。

実験テスト

#05 5. 問題名称：商談終了

【視覚情報】



【音声情報】

1. コピーをしています。
2. 質問をしています。
3. レセプションを始めるところです。
4. 商談を終えたところです。

実験テスト

#06 6. 問題名称：会議室で説明

【視覚情報】



【音声情報】

1. 喫茶店で商談をしています。
2. 会議室で説明を聞いています。
3. 社長室であいさつをしています。
4. 会議室で契約書にサインしています。

聴解 §2 表現力問題

このセクションでは、音声を聴きながら、写真または絵を見て答えてください。質問のあと、4つの文を読みます。1、2、3、4の4つの中から正しい答えを1つだけ選んでください。

実験テスト

#07 1. 問題名称：アポなし訪問

【視覚情報】



【音声情報】

受付に突然の来客がありました。担当者は、今、席をはずしています。お客に少し待ってもらいます。なんと言いますか。

1. 少々お待ちいただけますか。
2. お取次ぎできかねますが。
3. お帰りになりますか。
4. お引取り願えませんか。

実験テスト

#08 2. 問題名称：お大事に

【視覚情報】



【音声情報】

同僚の奥さんから、電話がかかってきました。同僚は、病気で休むと言っています。何と答えますか。

1. わかりました。お困りでしょうね。
2. わかりました。しょうがありませんね。
3. わかりました。お大事になさってください。
4. わかりました。お氣をつけてください。

実験テスト

#09 3. 問題名称：荷物受け取り

【視覚情報】



【音声情報】

この写真は、ホテルのお客が自分の荷物を受け取りに来た場面です。お客は自分で荷物を受け取ります。どのように案内しますか。

1. あちらで荷物をお受け取りください。
2. あちらで荷物をお受け取りしてください。
3. あちらで荷物をお受け取りしてあげます。
4. あちらで荷物をお受け取りしてさしあげます。

実験テスト

#10 4. 問題名称：新商品の説明

【視覚情報】



【音声情報】

男の人が訪問先で新商品の説明をしています。この男の人は、どのように話を始めますか。

1. では、ご説明してあげます。
2. では、ご説明してさしあげます。
3. では、ご説明させていただきます。
4. では、ご説明させていただきます。

実験テスト

#11 5. 問題名称：プレゼン開始 reform

【視覚情報】



【音声情報】

プレゼンテーションをしています。参加者に、質問があるか聞きます。参加者に、なんと言いますか。

1. 何かご質問がしたいんですか。
2. 何かご質問がありますか。
3. 何かご質問でもあるんですか。
4. 何かご質問でもなさりたいんですか。

実験テスト

#12 6. 問題名称：資料請求

【視覚情報】



【音声情報】

取引先から見積もりを送ってほしいと電話がありました。送り先の担当者名を尋ねます。なんと言いますか。

1. どこにお送りいたしましょうか。
2. だれにお送りいたしましょうか。
3. どなたにお送りいたしましょうか。
4. どっちにお送りいたしましょうか。

聴解 83 聴解問題

問題用紙の絵は、場面を表しています。このセクションでは、問題用紙の絵を見ながら、音声を聴いて答えてください。質問のあと、4つの文を読みます。1、2、3、4の中から正しい答えを1つだけ選んでください。

実験テスト

#13 1. 問題名称：手伝う

【視覚情報】



【音声情報】

男の人と女の人が、取引先の入金状況について話しています。女の方は、このあと、何をしますか。

男2：うわー、まいったなあ。

女2：どうしたんですか？

男2：ヤマシタ物産からの入金なんだけど、今日も入ってないんだよ。

女2：えー、まだなの？

男2：うん、メールでプッシュするのも、もう、何回目だろ。

女2：金利が、ばかにならないもんね。

男2：うん。週明けには入ってるって、返事だったのに……。電話するしかないか。あ、でも、今日、これからずっと外回りだ。

女2：わたし、代わりにやっとうか？

男2：お願いできるかなあ。

女2：ええ、担当者と何度か話したことあるから。まかせといて。

男2：助かるよ。

女の方は、このあと、何をしますか。

1. 入金の確認をする。
2. メールを出す。
3. 電話をする。
4. 外出する。

調査用紙（聴解3）

実験テスト

#14 2. 問題名称：言付け

【視覚情報】



【音声情報】（男2にフィルタ）

男の人から電話がかかってきました。男の人は、何について相談をしたいと言っていますか。

女2：はい、ABC 商事営業課でございます。

男2：サン商事の伊藤と申しますが、山田さん、いらっしゃいますか。

女2：あいにく、山田はただいま外出中でございますが、昼過ぎには戻ると思いますが……。

男2：そうですか。それでは、伝言をお願いしたいんですが。

女2：はい。

男2：えー、このたび当社が開発した新商品の件なんですが。

女2：はい。

男2：先日そちらにお伺いして、特徴をご紹介させていただいたんですが、その時ペンディングとなっておりました販売価格について、最終的に相談したいと思います。

女2：価格の件ですね。

男2：はい。で、お約束どおり、午後伺いますとお伝え下さい。

女2：はい。申し伝えておきます。

男の人は、何について相談をしたいと言っていますか。

1. 商品の紹介
2. 商品の価格
3. 商品の特徴
4. 商品の開発

調査用紙（聴解3）

実験テスト

#15 3. 問題名称：請求書へのクレーム

【視覚情報】



【音声情報】

上司と部下が、取引先に送った請求書の内容について、話しています。請求書の内容は、どうなっていましたか。

女1：小林くん、この間ライオン商事に送った請求書のことなんだけど。

男2：はい。あ、あちらの本社に送ったやつですね。

女1：数量が違うんじゃないかっていうクレームが来てるんだけど。

男2：え？そんなはずは…。

女1：数が多いから、先月分も含めてないかって。

男2：…あ、それ、支店に送った分も入ってるからだと思います。

女1：なんでそれ、本社に送ったの？

男2：あ、あちらから、「請求は本社宛てに」って指示があったので。納品書の控えがありますので、すぐ確認できます。

女1：そう。じゃ、こちらのミスじゃなかったのね。

男2：そんな。いくらぼくでも数を間違えたりしませんよ。

女1：あら、前に、他の会社の分を請求したことがあったじゃない。

請求書の内容は、どうなっていましたか。

1. 数が間違っていた。
2. 支店の分も含まれていた。
3. ほかの会社の分も含まれていた。
4. 前の月の分も含まれていた。

調査用紙（聴解3）

実験テスト

#16 4. 問題名称：契約決済待ち

【視覚情報】



【音声情報】（女1にフィルタ）

男の人が、取引先の女の人と電話で話しています。女の方は、今回の契約について、なんと言っていますか。

女1：はい。トキワ食品でございます。

男1：もしもし、相沢物産の秋山ですが、前田さんお願いできますでしょうか。

女1：はい、わたくしです。いつもお世話になっております。

男1：ああ、どうも、秋山です。こちらこそ、いつもお世話になっております。あのう、本日お電話いたしましたのは、先日のご契約の件でございます。

女1：ええ、はい。こちらはその件で、お電話しようと思っていたところなんです。

男1：そうですか。では…、結論は、出ましたでしょうか。

女1：それが…結論は出ているんですが、部長が海外出張中で、最終決裁がおりてないんですよ。

男1：そうですか…。

女1：すみません。来週あたまにはご連絡するようにいたしますので。

男1：ありがとうございます。では、よろしくお願いします。

女1：いえ、わざわざありがとうございました。

女の方は、今回の契約について、なんと言っていますか。

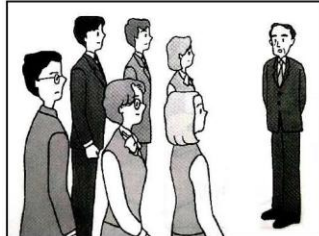
1. 部長がいないので待つてほしい。
2. 担当者でないからわからない。
3. 契約をやめる。
4. 契約書を直す。

調査用紙（聴解3）

実験テスト

#17 5. 問題名称：社長挨拶 reform

【視覚情報】



【音声情報】

社長が、社員の前で話しています。社長は、これから、どのような方針で経営を進めていく必要があると言っていますか。

男1：えー、わが社は、これまでの大衆化路線一辺倒から脱却する時期に来ています。マーケットには、必ずしも低価格を求める層だけが、存在しているわけではありません。品質の高いものを、それに相応しい価格で、夢や個性を求める層が数多く存在しています。今、なすべきことは、決してリストラではありません。リ・エンジニアリング、つまり、まったく新しい発想で会社を再生させる、この一言につきます。不採算部門を主力部門に統合し、採算部門を強化、商品の付加価値を高める必要があります。
わが社が行うべきことは、本業以外への進出や人員整理ではなく、全社員が一丸となった付加価値の創造にあると思います。

社長は、これから、どのような戦略を進めていく必要があると言っていますか。

1. 商品の高級化
2. 商品の低価格化
3. 経営の多角化
4. 会社のスリム化

調査用紙（聴解3）

実験テスト

#18 6. 問題名称：カスタマイズ

【視覚情報】



【音声情報】

コンピュータ会社で、上司と部下が話しています。部下は、このあと、会議でどんなことを提案しますか。

- 男1：このところ、売りが落ちてるね。
男2：ええ、そうなんです。去年の同じ時期と比べると、今月は8割減です。
男1：そんなに落ち込んでるのか。
男2：ええ。
男1：これじゃ値下げしても、追いつかんね。
男2：そうですね。
男1：いままで、デザインや機能を変えたり、いろいろやったけど……。
男2：はい。宣伝にも、人とお金をじゅうぶんにかけてます。
男1：なにか別のアプローチが必要かな。
男2：はい。実は……。
男1：うん？
男2：ユーザーのニーズに合わせて、カスタマイズして売るんです。
男1：カスタマイズ？
男2：はい。ユーザー一人ひとりの注文に応えるんです。
男1：なるほど。
男2：1度、会議にあげさせていただきませんか。
男1：うん。じゃあ、その前に、もう少し詳しく聞かせてもらえるかな。
男2：はい。

部下は、このあと、会議で、どんなことを提案しますか。

1. 価格の変更
2. デザインの変更
3. 宣伝の方法
4. 販売の方法

調査用紙（聴解3）

実験テスト

#19 7. 問題名称：パソコンの増産体制

【視覚情報】



【音声情報】

女の人が、パソコンの増産体制について、会議で報告しています。まだ解決していない問題は、何ですか。

女1：ノートパソコンを増産するにあたりまして、部品の供給、そして、ワーカーの確保が問題となっていました。今日は、調査の結果をご報告いたします。えー、まず、部品のサプライヤーですが、今回、定評のある2社を、新たに加えることに成功いたしました。また、現在ある5社も、供給能力に問題はないと思われます。次に、ワーカーの手配についてですが、常時採用可能な期間工を100名確保いたしました。この状況は、今後、半年間は変わらない見通しです。ただ、半年後に、他社の新工場が稼働することがわかりまして、その際には、そちらに、かなり移る可能性があると思われます。その時点で、なおワーカーを確保できるかが、今回の目標達成の成否を、大きく左右すると思われます。

まだ解決していない問題は、何ですか。

1. 部品の生産能力
2. 部品の供給元の確保
3. 新工場の稼働率
4. 今後の労働力の確保

調査用紙（聴解3）

実験テスト

#20 8. 問題名称：カーテンの在庫数 reform

【視覚情報】



【音声情報】

デパートの店員が、話しています。部下は、このあと取引先に何を問い合わせますか。

女2：あ、主任。納品されたカーテン、チェック、終わりました。

男1：新作のカーテンのことね。

女2：はい。

男1：今度のカーテンも人気が出そうだね。

女2：はい。前회가約3週間で完売してますから、今回の新作も同じくらいでの完売を見越しています。

男1：新作はパステルカラーの3色模様だったよね。

女2：はい、そうです。前は1色のグラデーション模様で、それはそれで好評でした。今回の新作に関しては、すでにお客様からの問い合わせも来てます。

男1：うん、ゴージャスな感じだからな。じゃ、追加注文も考えたほうがいいかな。ボーナスの時期だし、店頭に並べる数を増やすか。

女2：じゃ、追加注文しておきますか。まだ余裕ありますかね…。

男1：じゃ、とりあえず先方に、ファックスであとどのくらいこちらにまわせるか、確認しておいてくれる？

女2：はい、わかりました。

部下は、このあと取引先に何を問い合わせますか。

1. 商品の模様
2. 商品の在庫
3. 商品の納期
4. 商品の色

調査用紙（聴解3）

実験テスト

#21 9. 問題名称：派遣社員の採用

【視覚情報】



【音声情報】

上司と部下が、派遣社員の採用について話しています。上司は、派遣社員をどのように採用したいと言っていますか。

男1：高橋君、例の、キャンペーン期間中の派遣社員、手配のほうどうなってる？

男2：あ、はい、派遣会社から連絡がありまして、1か月を2、3名のローテーションでどうかと言ってきてるんですが。

男1：1か月を2、3名？

男2：はい、1日とか2日交代の形になると思いますが。

男1：うーん、何か月もっていうわけじゃないんだから、1人でフルに入ってもらえる人を頼むよ。

男2：そうですか。交代ではだめですか。

男1：うん、1か月だけのことだからねえ。人数が多いと、手間がかかって効率悪いからなあ。

男2：ええ。そうですね。

男1：うん、そういうことだから、もう1度、調整してみてくれないか。

男2：わかりました。連絡してみます。

上司は、派遣社員をどのように採用したいと言っていますか。

1. 同じ人を1か月以内
2. 3名の人を1か月以内
3. 同じ人を1か月以上
4. 3名の人を1か月以上

聴読解 §1 描写問題

このセクションでは、音声を聴きながら、問題用紙を見て答えてください。質問をよく聴いて、問題用紙の1、2、3、4の中から正しい答えを1つだけ選んでください。

実験テスト

#22 1. 問題名称：大阪

【視覚情報】



1. クレームの処理
2. 新任のあいさつ
3. 工場の見学
4. 新商品の説明

【音声情報】

2人の会社員が、出張先での仕事について話しています。出張の目的は、何ですか。

男1：大阪は久しぶりだなあ。

女2：そうですね。本当は、もっといい話をしに行きたいところですけどね。

男1：まあ、つきあいの長いところだから、こちらの事情もご存じとは思うけど。

女2：そうですか。

男1：でも、こういうときは、ちゃんと先方の顔を見て、ご説明するのが一番だね。

出張の目的は、何ですか。

実験テスト

#23 2. 問題名称：調子はどう？

【視覚情報】



1. 体の調子
2. 機械の調子
3. 売り上げの実績
4. 生産の効率

【音声情報】

上司と部下が話しています。何について話していますか。

男1：よう！

男2：あ、部長。

男1：どう？ 今月の生産ライン、調子はどう？

男2：第1ラインの生産性が上がりました。

男1：それじゃ、今月の目標台数の達成は、確実だね。

男2：はい。

何について話していますか。

調査用紙（聴読解1）

実験テスト

#24 3. 問題名称：1 ミリ縮めて

【視覚情報】



1. 部品の数进行少なくする。
2. 部品の金額を安くする。
3. 部品の納期を早くする。
4. 部品の大きさを小さくする。

【音声情報】

男の社員が、取引先と機械の部品のことについて、電話で話しています。取引先の人に、何をお願いしていますか。

男2：……ええ、ええ、そうです。なんとか、あと1 ミリ、ええ、できれば縮めてもらって。…
…そうなんです。……ええ。いいですよ、延ばしても、納期は。……はい。じゃ。なんと
か、その線をお願いします。

取引先の人に、何をお願いしていますか。

実験テスト

#25 4. 問題名称：社員募集要項 reform

【視覚情報】

〇〇〇〇年度 社員募集要項

募集人員：正社員 5名
年 齢：大卒・35歳位まで
給 与：委細面談
休日休暇：完全週休二日制・有給休暇あり

応 募：

以下の書類をご郵送ください。書類選考の上、面接の
ご連絡を致します。応募書類は返却いたしません。

・履歴書（要写真添付・当社所定用紙）＊
・応募理由書（形式自由）

＊当社所定用紙は下記採用担当宛に電話または
ファックスにてご請求ください。

応募締切：〇〇年〇月〇日必着

書類請求・送付先：
株式会社 ミスト
〒〇〇〇-〇〇〇〇 東京都〇〇区〇〇北〇-〇〇
管理部人材開発課 採用担当
TEL.03-〇〇〇〇-〇〇〇〇
FAX.03-△△△△-△△△△

1. 応募書類を送る
2. 指定の用紙を取り寄せる
3. 応募理由を知らせる
4. 面接可能日を知らせる

【音声情報】

次の広告は、ある会社の社員募集要項で
す。応募する時は、まず最初に何をする
ように言っていますか。

調査用紙（聴読解1）

実験テスト

#26 5. 問題名称：前売

【視覚情報】

第19回 ゲームソフト見本市TOKYO

テーマ | 近未来|

会 期 ○○○○年○月○日（土）ー○月○日（水）

会 場 ○○○ ○○○○

開催時間 平日10:00-19:00 土・日9:30-19:00

出品者数 15カ国から175社が出品

入場料 一般（高校生以上）：1,200円（前売1,000円）
（小学生以下） 小中学生：600円（前売 600円）
 ※当日売16時以降入場料
 一般（高校生以上）：1,000円
 小中学生：500円

1. 1,000 円
2. 1,200 円
3. 2,000 円
4. 2,400 円

【音声情報】

同僚と2人で、来月開催の見本市に行くことになりました。事前に入場券を買っておきます。2人分でいくら必要ですか。

実験テスト

#27 6. 問題名称：将来なりたい職業

【視覚情報】

順位	男 子	女 子
1	ジャーナリスト	保 育 士
2	教 師	看 護 師
3	医 師	教 師
4	会 社 員	デ ザ イ ナ ー
5	エ ン ジ ニ ア	会 社 員

1. ジャーナリスト
2. 保育士
3. 教師
4. 会社員

【音声情報】

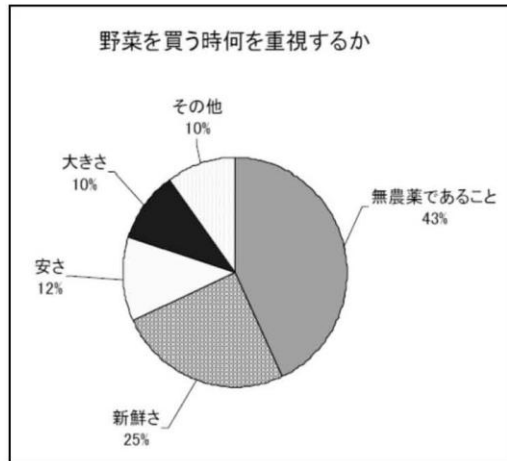
次の表は、ある生命保険会社が高校生を対象として、生徒が将来何になりたいかを調査した結果です。高校生の男子も女子も共通に多く希望しているのは、どんな仕事ですか。

調査用紙（聴読解 1）

実験テスト

#28 7. 問題名称：野菜を買うときに重視すること

【視覚情報】



1. 安全性
2. 鮮度
3. 値段
4. 形

【音声情報】

次のグラフは消費者が食品を買う時、何を重視しているかをあらわしたものです。このグラフから消費者は何を重視していることがわかりますか。

実験テスト

#29 8. 問題名称：健康診断

【視覚情報】

健康診断のお知らせ

新年度の健康診断を下記のとおり行いますので、各部の該当者は所定の日時・場所にて受診してください。

総務部長

	男 性	女 性
(1) 4月9日（月） 午前	営業部 セミナールーム	企画部 第3会議室
(2) 4月9日（月） 午後	企画部 セミナールーム	総務部 第3会議室
(3) 4月10日（火） 午前	総務部 大会議室	販売部 第1会議室
(4) 4月10日（火） 午後	販売部 大会議室	営業部 第1会議室

午前 10:00～12:00 午後 14:00～16:00

1. 月曜日 セミナールーム
2. 月曜日 第3会議室
3. 火曜日 大会議室
4. 火曜日 第1会議室

【音声情報】

次のお知らせは、健康診断についてのものです。企画部の男性は、何曜日にどこの部屋で健康診断を受けますか。

調査用紙（聴読解1）
実験テスト

#30 9. 問題名称：建物賃貸契約書

【視覚情報】

建物賃貸契約書

1 第1条 賃貸借料について

2 第2条 光熱費および水道料について

3 第3条 火災保険加入について

4 第4条 中途解約について

【音声情報】

次の文書は、建物の賃貸契約書です。契約期間内に引っ越し場合の条件は、どこに書いてありますか。

聴読解 §2 聴読解問題

このセクションでは、音声を聴きながら、問題用紙を見て答えてください。質問をよく聴いて、問題用紙の1、2、3、4の中から正しい答えを1つだけ選んでください。

実験テスト

#31 1. 問題名称：販売者会議 reform

【視覚情報】

関係各位

〇〇〇〇年〇月〇日

ABC株式会社
営業本部長 松野太郎

販売フォーラムのご案内

下記の通り、販売フォーラムを開催いたしますのご案内いたします。

記

日時：〇〇〇〇年×月×日 10:00~17:00
場所：ABC株式会社 大会議室
次第：

1. 開会の辞
2. 挨拶
3. 講演「金融戦略」
4. 講演「販売戦略について」
5.
6.
7. 閉会の辞

懇親会 〇〇ホテル ××の間にて

ABC株式会社 山田一郎
副社長

JFC銀行 頭取 佐藤浩氏

ミナト物産株式会社
営業課長 大山和男氏

1

2

3

4

【音声情報】（男1にフィルタ）

上司が外出先から、電話をかけて来ました。部下は、このあと、書類のどの部分を訂正しますか。

- 男2：はい、営業三課です。
- 男1：あ、松野だけど。
- 男2：あ、部長、お疲れ様です。
- 男1：うん、ちょっと頼みたいことがあるんだけど。
- 男2：はい。
- 男1：あの、例の販売者会議の件なんだけど。
- 男2：はい。
- 男1：うん、今、JFC銀行の頭取に会ってね。
- 男2：ああ、佐藤頭取ですね。
- 男1：佐藤さんは午前中しか出られないって。

男2：あ、はい。わかりました。懇親会、欠席ですね。

男1：うん。それで、うちの挨拶なんだけど、やっぱり、社長にお願いすることになったから。

男2：ああ、よかったですね。

男1：それで、案内状を、至急、変更しておいてほしいんだけど。

男2：はい、わかりました。

男1：あと、ミナト物産の大山さんは予定通り。

男2：はい、わかりました。

部下は、このあと、書類のどの部分を訂正しますか。

調査用紙（聴読解2）
実験テスト

#32 2. 問題名称：会議の議題

【視覚情報】

〇〇〇〇年〇月〇日

定例部課長会議

議題：新製品LⅢについて

- 1 デザイン
- 2 販売価格
- 3 販売チャネル
- 4 広告宣伝方法

資料 ① 〇〇〇〇〇〇〇〇〇
② 〇〇〇〇〇〇〇〇〇
③ 〇〇〇〇〇〇〇〇〇
④ 〇〇〇〇〇〇〇〇〇

実験テスト

#33 3. 問題名称：祝賀会挨拶

【視覚情報】

株式会社〇〇〇サービス
創立30周年記念祝賀会
プログラム

- 1 ■ 開会の辞 総務課長 〇〇〇〇
- 社長あいさつ 代表取締役社長 〇〇〇〇
- 2 ■ 来賓祝辞 〇〇市長 〇〇〇〇
- 来賓祝辞 〇〇電機社長 〇〇〇〇
- 3 ■ 乾杯 〇〇スチール社長 〇〇〇〇
- 歓談
- 4 ■ 閉会の辞 総務課長 〇〇〇〇

より一層の御社のご発展をお祈りし、ごあいさつに代えさせていただきます。本日はおめでとうございます。

プログラムのどこの話ですか。

【音声情報】

会議に遅れました。いま、どの議題について話されていますか。

男2：今回ののは、従来モデルとは違って、20代をターゲットにしたデザインですから、デパートや小売店だけではなく、より新しいルートを開拓していく必要があると思いますね。

女2：そうですね。郊外にあるディスカウストアなども、視野に入れておくべきでしょうね。

いま、どの議題について話されていますか。

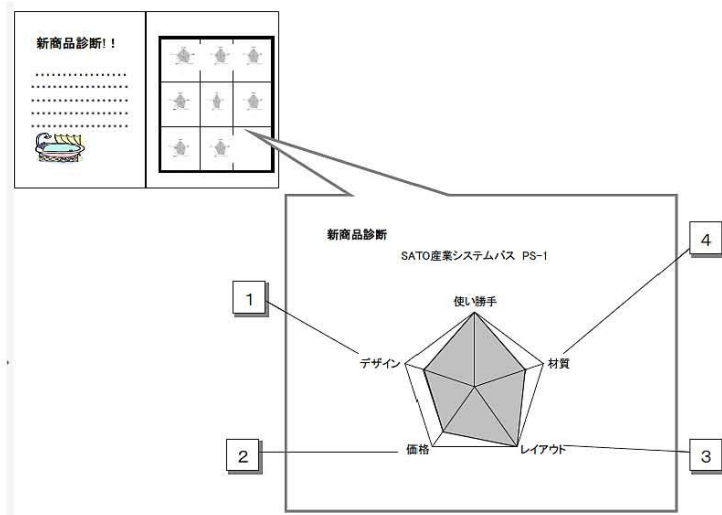
【音声情報】

パーティー会場に、遅れて着きました。誰かが話しています。プログラムのどこの話ですか。

男2：……思えば、弊社とお取引をいただき、十数年の月日が過ぎようとしています。御社のご支援、ご厚意により、わたくしどもも成長することができました。これもひとえに、社長さんをはじめとする従業員の皆様のご厚情と、心から感謝しております。今後とも相変わりがせず、よろしくお願いいたしますとともに、より

調査用紙（聴読解２）
実験テスト

#34 4. 問題名称：新製品の評価
【視覚情報】



【音声情報】

上司と部下が、自社の製品に関する雑誌の記事を見ながら、話しています。上司は、どの項目が問題だと言っていますか。

女2：課長、うちの新しいシステムバスの評価が出てますが…。

男1：え？どれ…。

女2：はい、これです。使い勝手は満点ですね。

男1：ああ、そうだね。デザインもまあまあだね。

女2：機能的にも今回、蛇口の形を工夫して、一段と掃除しやすくしましたから。

男1：うん、まあ、クリアしたって感じだね。

女2：ええ。

男1：それに、今回の目玉は、従来と同じレイアウトで、モジュールの配置を変えたってことだったよね。

女2：はい、そうです。それにしても、価格の評価は…。

男1：うん。前と変わらずだね。

女2：材質はいいので、変えるわけにいきませんしね。

男1：うん、でも、モジュールの変更だけじゃ、コストを下げるのは…。

女2：難しいですね。

男1：そうなんだよ。もう少し、ここのコストを抑えられればなあ。

女2：そうですね。

上司は、どの項目を改良すべきだと言っていますか。

調査用紙（聴読解 2）
実験テスト

#35 5. 問題名称：新入社員面接
【視覚情報】

1	<p>氏名 ○○○○</p> <p>学歴 ○○○○年○月 イギリスコベント中等学校卒業 ○○○○年○月 アメリカミズーリ高等学校卒業 ○○○○年○月 東都大学経済学部経営学科 卒業予定</p>	2	<p>氏名 ○○○○</p> <p>学歴 ○○○○年○月 吉田高等学校卒業 ○○○○年○月 中部大学経済学部経営学科卒業</p> <p>職歴 ○○○○年○月 丸中食品入社 ○○○○年○月 同社門前退社 ○○○○年○月 三井自動車入社 現在：営業2課 課長補佐</p>
3	<p>氏名 ○○○○</p> <p>学歴 ○○○○年○月 池野高等学校卒業 ○○○○年○月 京阪大学ビジネス学部卒業 ○○○○年○月 同大学大学院マネジメント学科 修士過程修了予定</p>	4	<p>氏名 ○○○○</p> <p>学歴 ○○○○年○月 田中高等学校卒業 ○○○○年○月 四国外国語大学英語学科卒業 ○○○○年○月 関東大学大学院経営実務学科 博士前期課程修了 ○○○○年○月 同大学大学院同学科 博士口頭課程修了予定</p>

【音声情報】

ある会社で、社長と人事課長が、新入社員の面接について話しています。社長は、最初にどの人に会うと言っていますか。

男2：社長、こちらが、今回、最終選考に残った4名の履歴書です。

男1：うん。

男2：彼女はいわゆるバイリンガルで、英語力はかなりものだそうです。

男1：そうだな。最近、海外との取引も増えてきてるし、こういう人材を増やしておく必要があるだろうね。

男2：ええ。それから、彼の場合は転職になります。

男1：うん。なるほど、いろんな業界を経験してるんだね。案外、こういう人間が、うちの体質に風穴をあけてくれるかもしれないね。

男2：はい。あとは、院卒の2人です。

男1：ほう。修士に博士か。うん？しかし、2人ともタイプミスがあるな。

男2：本当ですね。

男1：こんなことでミスを犯すようじゃ、いくら学歴が高くてもなあ……。やはり、実務経験だな、即戦力重視。よし、まずこの人に会うことにするよ。

社長は、最初にどの人に会うと言っていますか。

実験テスト

#36 6. 問題名称：見積書の返事

【視覚情報】

1	東西工業の件、 金額、見積もり通り。 納期、1週間早く。	2	東西工業の件、 金額、安く。 納期、1週間早く。
3	東西工業の件、 金額、見積もり通り。 納期、2～3日早く。	4	東西工業の件、 金額、安く。 納期、2～3日早く。

【音声情報】

男の人と女の人が、見積書の返事について話しています。女の方は、どのメモを書きますか。

男1：大塚さん。東西工業の見積もりの返事、メールしてくれないかな。

女1：はい。このあいだの見積もりですね。

男1：そうそう。

女1：なんて書きますか。

男1：うん、そうだな。金額はあれでいいんだけど……、こっちの在庫が危ないからね。納品、急げないかって。若干でも早くなるって言うてくれば、すぐそっちに発注するって書いてくれる？

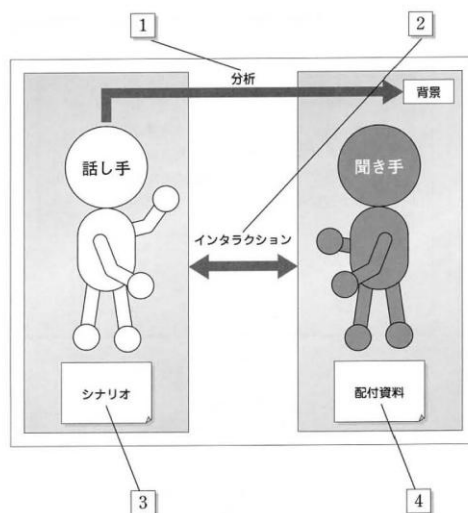
女1：若干って、2、3日ですか。

男1：いや、1週間かな。

女1：はい。

女の方は、どのメモを書きますか。

調査用紙（聴読解2）
実験テスト
 #37 7. 問題名称：プレゼンセミナー
【視覚情報】



【音声情報】

セミナーで、プレゼンテーションの方法について説明しています。この人は、図の中のどこが一番大切だと言っていますか。

女1：えー、プレゼンテーションに含まれる主要素としては、この図のようなことが考えられます。まず、プレゼンの流れを決めるシナリオ、それから、聞き手の背景の分析、聞き手とのインタラクション、そして、配付資料、の4つです。そこで最も大切なのは、プレゼンを一方的にダラダラ行わないということです。皆さん、通常、質疑応答をまとめて最後に回していませんか。聞き手から出る質問を恐れていませんか。聞き手が質問するのは、提案内容に興味がある証拠です。質問は、話の節目ごとに適宜取り上げ、その場その場で答えたほうが効果的です。プレゼンというのは、自分が話をする場ではなく、聴衆の話聞く場なんです。

図の中のどこが一番大切だと言っていますか。

調査用紙（聴読解2）

実験テスト

#38 8. 問題名称：飛行機のチケット

【視覚情報】

空 席 案 内				
出発地（時刻）	到着地（時刻）	便名	機種	空席状況
東京（0900）	札幌（1035）	BNB 057	744	×
東京（0930）	札幌（1105）	BNB 961	763	×
東京（1000）	札幌（1135）	BNB 059	74R	○
東京（1025）	札幌（1155）	BNB 4713	763	×
東京（1130）	札幌（1300）	BNB 061	74R	○
東京（1200）	札幌（1330）	BNB 963	744	△
東京（1250）	札幌（1420）	BNB 4715	763	△

【音声情報】

男の人が、女の人に飛行機のチケットの予約を頼んでいます。女の方は、どの便を予約しますか。

男1：森さん。すまないけど、航空券の手配、お願いできますか。

女2：はい。ご出張ですか。

男1：うん。来週の頭から、札幌なんだけだね。

女2：お時間は？

男1：えーと、会議が1時からだから……、向こうの空港に、遅くとも12時に着く便をお願いします。

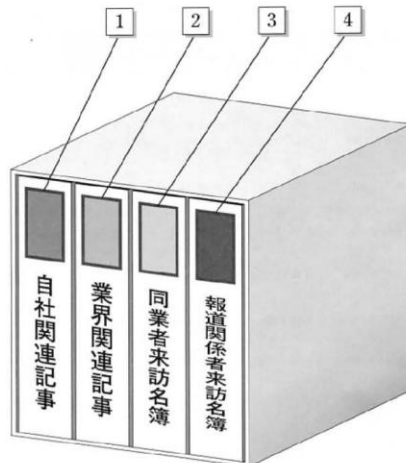
女2：はい。じゃ、予約を入れておきます。

女の方は、どの便を予約しますか。

実験テスト

#39 9. 問題名称：記事のファイル

【視覚情報】



【音声情報】

男の人と女の人が書類のファイルについて話しています。男の人は書類をどのファイルにしまいますか。

女1：佐藤さん、ちょっと。

男2：はい。

女1：先月のはじめに、アメリカから、ジャーナリストがうち取材に来たでしょ。

男2：ええ。たしか、経済誌の記者でしたよね。

女1：そう。そのときの記事を送ってきてくれたから、ファイルしといてくれる？

男2：はい。

男の人は書類をどのファイルにしまいますか。

読解 §1 文法・語彙問題

次の文の _____ に入る正しいことばを、1、2、3、4の中から1つ選んでください。

実験テスト

#40 1. 問題名称：売上高

A社の売上高はB社の売上高の _____ 3分の1にすぎない。

1. ごく
2. ただ
3. わずか
4. すこし

実験テスト

#41 2. 問題名称：所在地

所在地なんて、わざわざ、取引先に _____ ほどのことでもない。インターネットで調べればすぐわかる。

1. 問いかける
2. 問いただす
3. 問い合わせる
4. 問いつめる

実験テスト

#42 3. 問題名称：通勤

東京に転勤するまで、通勤がこんなに大変だとは思っても _____。

1. よりませんでした
2. ありませんでした
3. かかりませんでした
4. のこりませんでした

実験テスト

#43 4. 問題名称：閉店

残念ながら報告ですが、当店は来月末日を _____、閉店させていただきます。

1. もちまして
2. まちまして
3. とりまして
4. きりまして

調査用紙（読解1）

実験テスト

#44 5. 問題名称：納得できない

注文内容が違っているとお電話いただいた件ですが、当方には納得_____点がございます。

1. いたしたい
2. いたすべき
3. いたしかねる
4. いたしやすい

実験テスト

#45 6. 問題名称：水増し

受注工事のかなりの部分が架空であり、売り上げや利益を水増しする_____ものであった。

1. ための
2. ほどの
3. ことの
4. あまりの

読解 §2 表現力問題

次の文の _____ に入る正しいことばを、1、2、3、4の中から1つ選んでください。

実験テスト

#46 1. 問題名称：ちょっと一杯

A：ちょっと、一杯、寄っていかない？おごるよ。

B：いえ、今日は、ぜひ私たちに_____。

1. 払ってください
2. 払わせてください
3. 払わせてあげます
4. 払ってもらってください

実験テスト

#47 2. 問題名称：ご迷惑

課長にはいろいろご迷惑を_____。

1. くださいました
2. おかけしました
3. おかけになりました
4. さしあげました

実験テスト

#48 3. 問題名称：変な音

客：先週、おたくで買ったパソコン、変な音がするんだけど。

店員：申し訳ございません。_____、ただいま担当の者を呼んでまいります。

1. 私ではわかりかねますので
2. 私には関係がありませんので
3. 私の仕事ではありませんので
4. 私は遠慮させていただきますので

実験テスト

#49 4. 問題名称：ご存知ですか

A：営業1課の森さんをご存知ですか。

B：いいえ、_____。

1. ご存じないです
2. ご存じありません
3. ご存じではありません
4. 存じておりません

調査用紙（読解2）

実験テスト

#50 5. 問題名称：当社にお越しいただければ

現在、当社で開発中の AX-300 のスペックにつきましては、当社にお越しいただければ、その詳細を_____。

1. ご覧ください
2. ご覧いただきたい
3. ご覧になります
4. ご覧になれます

実験テスト

#51 6. 問題名称：認める

ASB 商事の田村ですが、先日の件、うちの課長が_____ので、近々お会いしたいのですが、ご都合はいかがでしょうか。

1. 認めました
2. 認めてあげました
3. お認めくださった
4. お認めいただいた

読解 83 読解問題

次の文章を読んで、質問に答えてください。答えは、1、2、3、4の中から1つ選んでください。

実験テスト

#52 1. 問題名称：仕事の引き継ぎ

先輩から、仕事の引き継ぎのメールが届きました。メールを受け取った社員は、一番はじめに何をしなければなりませんか。

Date: Mon, xx xxx xxxx xx:xx:xx
From: nakano@xxxxx.xxx.co.jp
To: yamamoto@xxxxx.xxx.co.jp

山本さんへ

明日は一日留守にします。
次の4点よろしく。

- ・会議のセット
明後日の朝の会議、準備よろしく。
出席者の出欠をおさえておくこと。
- ・資料のコピー
時間がある時に役員会資料を30セットコピー。バイトさんに指示。
- ・部長のスピーチ原稿
原稿を部長にチェックしてもらう。
部長、午後から出張、必ず朝イチで。
- ・翻訳
スピーチ原稿、部長からOKでたら、すぐ英訳に出す。
時間ないから、急ぎでお願いすること。

以上、悪いけどお願いします。

中野

1. 会議出席者の出欠をチェックすること
2. 役員会の資料をコピーすること
3. 部長に原稿をみてもらうこと
4. 原稿を英語に翻訳すること

調査用紙（読解3）

実験テスト

#53 2. 問題名称：定期購読

次のような手紙を受け取りました。用件は何ですか。

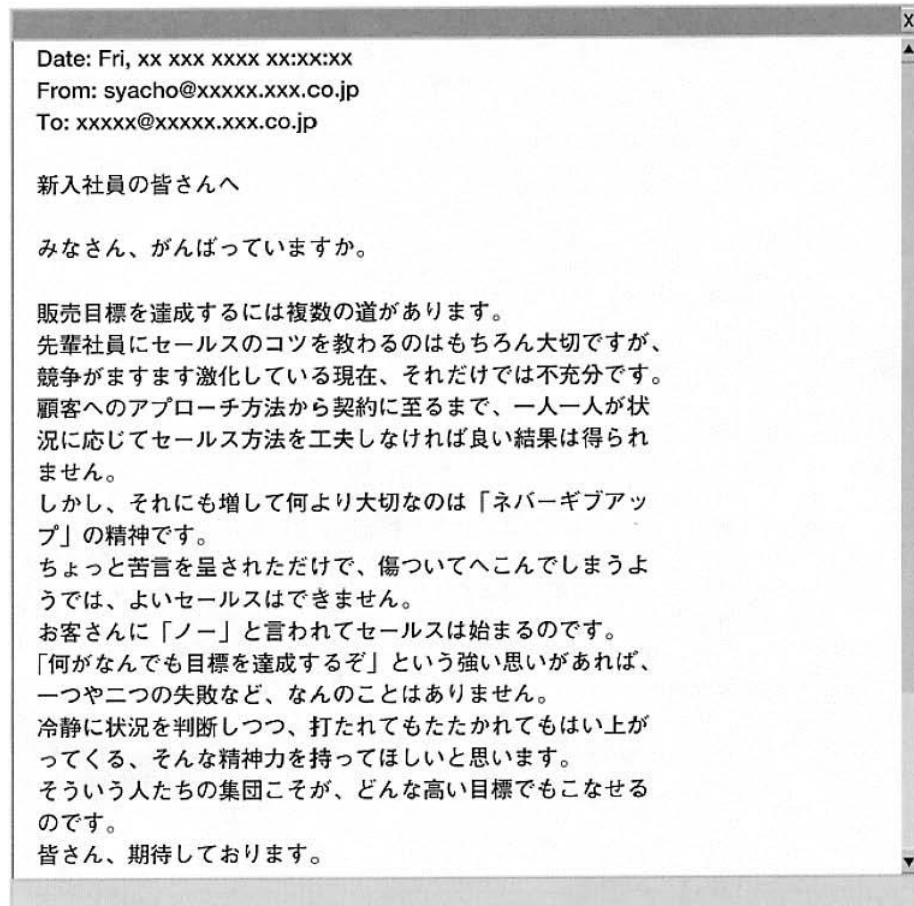
		〇〇〇〇年〇月〇日	
株式会社	〇〇〇		
総務部	〇〇〇〇様		
		株式会社	〇〇〇〇〇
		営業部	〇〇〇〇
<p>拝啓 時下ますますご盛栄のこととお慶び申し上げます。</p> <p>平素は弊社発行『週刊ビジネス世界』をご購読頂き、有り難うございます。</p> <p>現在、予約購読を頂いておりますが、来る4月に購読終了となります。3月31日までに更新の手続きをして頂き、一年分の購読料を前納して頂ける場合、当初の割引価格からさらに30%引きの特典がございます。この機会に是非、予約購読更新のお手続きをお薦めいたします。</p> <p>『週刊ビジネス世界』は、4月より誌面リニューアルを行い、ますます充実した内容を満載し、読者の皆様にご満足いただけるビジネスマガジンとして生まれ変わります。引き続き弊誌のご利用をよろしくお願いいたします。</p> <p style="text-align: right;">敬具</p>			

1. 購読を続けてほしい。
2. 購読を終了してほしい。
3. 未納の購読料を支払ってほしい。
4. 新規の購読を申し込んでほしい。

実験テスト

#54 3. 問題名称：社長からのメール

社長から、次のようなメールが届きました。社長は、セールス活動において何が一番大切だと言っていますか。



1. 先輩社員から、セールスの仕方を学ぶこと
2. 状況に応じて、セールスの仕方を変えられること
3. たたかれたら、たたき返せる精神力を持つこと
4. 失敗しても、やり直せる精神力を持つこと

調査用紙（読解3）

実験テスト

#55 4. 問題名称：炊飯器たけるくん

取引先から次のような文書が届きました。商品について、何をしてほしいと言っていますか。

〇〇〇〇年〇月〇日	
□□商事株式会社 営業部業務課調査係御中	△△株式会社 販売部販売促進課顧客サービス係
<p>拝啓 貴社ますます御清栄のこととお喜び申し上げます。</p> <p>さて、この度は、弊社製品『炊飯器・たけるくん』をお買い上げいただきまして、心より御礼申し上げます。</p> <p>先日直接お話を致しました通り、おかげ様で多くの皆様にご愛好いただき、大変、ご好評をいただいております。品質に関しましても、絶対の自信を持っておりまして、弊社では、日本のみならず海外にもお勧めしております。</p> <p>今後も、同製品の品質改良には、たゆまぬ努力を続けて参る所存でございますが、同製品につきまして、使い勝手等、ユーザー各位に評価いただくと共に、ご要望等、お伺いしたいと思います。</p> <p>つきましては、別添のアンケート用紙にご記入の上、ご返送いただけましたら幸いです。</p> <p>今後とも変わらぬご支援、ご愛顧を賜りますよう宜しくお願い致します。</p> <p>取り急ぎ、お買い上げの御礼かたがたご挨拶申し上げます。</p> <p style="text-align: right;">敬具</p>	

1. 調査を海外で実施してほしい。
2. 調査の報告をしてほしい。
3. 調査に協力してほしい。
4. 調査結果を集計してほしい。

調査用紙（読解3）

実験テスト

#56 5. 問題名称：販売報告会 reform

次の文書は、販売部長からのお知らせです。どんな販売レポートを提出するように言っていますか。

特約店各位	〇〇〇〇年〇月〇日
	△△本社 販売部長□□□□
販売報告会について	
下記の通り、販売報告会を開催することになりました。	
各位とも年度当初の目標数と販売実績を照らし合わせ、実証的なデータに基づいた、上半期の販売レポートを提出してください。以下の諸点を念頭に置き、報告するようにしてください。	
<ul style="list-style-type: none">・前年度の反省に基づいた販売促進が行われたか・前年度とは異なる戦略がとられたか・上半期の販売実績データの推移・下半期の販売戦略について	
レポート提出締切：6月23日 (メールに添付して送信のこと)	
記	
日時：〇〇〇〇年〇月〇日 午後1時～5時	
場所：本社 大会議室	
	以上

1. 前年度の販売実績の検証
2. 前年度の販売戦略の目標
3. 上半期の販売実績の検証
4. 上半期の販売戦略の目標

調査用紙（読解3）

実験テスト

#57 6. 問題名称：フレックスタイム制度

次の文書は、ある会社のフレックスタイム制度に関する社内文書です。フレックスタイム制度をどう見直すと言っていますか。

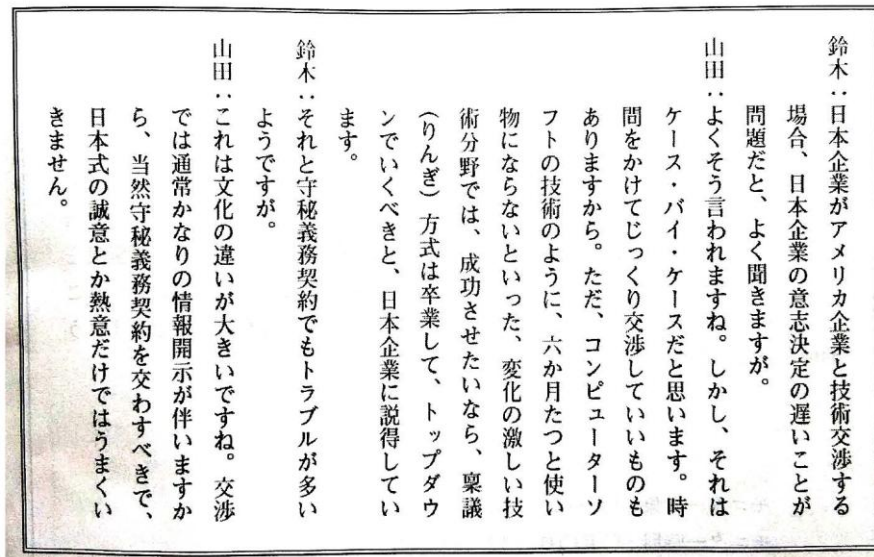
		〇〇〇〇年〇月〇日	
社員各位		総務部長	
フレックスタイム制度見直しについて			
<p>わが社で行っているフレックスタイム制度について、前回の役員会議で討議しました結果を皆様にお知らせいたします。</p> <p>かねてから定時に社員が揃わないことでお客様にご迷惑をかけ、また作業の能率も下げているとの声が各部署よりあがっておりました。</p> <p>そのような状況に鑑み、本年度をもって管理職に対する本制度の廃止を決定いたしました。また将来は、一般職についても取りやめる方向で検討中です。</p> <p>なお、育児や介護など、特別の理由のある社員にのみ、継続という方向で対応していく所存であります。</p> <p>皆様の生活に影響を与えることもあるかとは思いますが、ご理解のほどよろしくお願い申し上げます。</p>			

1. 全面的に廃止する。
2. 管理職のみ廃止する。
3. 一般職にのみ導入する。
4. 事情のある社員にのみ導入する。

実験テスト

#58 7. 問題名称：日米技術交渉対談記事

日米技術交渉に関する対談記事です。交渉のありかたについて専門家（山田氏）は何と言っていますか。



1. 日本企業の意志決定をスピードアップさせるべきだ。
2. 日本企業の意志決定の遅さも、守秘義務契約も全てケース・バイ・ケースで対応すればよい。
3. 守秘義務契約は当然である。
4. コンピューター技術のようなものは時間をかけて交渉すべきだ。

調査用紙（読解3）

実験テスト

#59 8. 問題名称：商品破損

納品した商品について、お客様から次のようなメールが届きました。用件は何ですか。

送信者：テルメニッヒ東京 PC営業部
日時：〇〇〇〇年〇月〇日 〇〇：〇〇
宛先：竹下流通センター
件名：パソコンラック(J-153)の件

竹下流通センター
〇〇〇〇様

前略
9月20日に注文いたしましたパソコンラック(J-153)30台が、本日
着荷いたしました。
入荷後の検品の結果、30台中、2台にキャスター部分の破損がみ
られました。破損原因は不明ですが、同一箇所のため、製造工程
での不具合かと思われます。
つきましては、至急、ご手配のうえ代替品をお送りいただきたく
お願い申し上げます。
なお、損傷品は、明細書を同封し、本日返送させていただきました。

草々

テルメニッヒ東京 PC営業部
〇〇〇〇

1. 壊れた商品を修理してほしい。
2. 壊れた原因を調べてほしい。
3. 代わりの商品を送ってほしい。
4. 明細書を返送してほしい。

実験テスト

#60 9. 問題名称：ガーリーパワー

次のような募集が来ました。どんなモニターを募集していますか。

〇〇〇〇年〇月〇日	
社員各位	広報部
<p>ご承知の通り、当社では、先週、健康飲料「ガーリーパワー」を発売いたしました。発売当日から、テレビ各局で同製品のコマーシャルを大々的に展開中です。当社の販売促進活動につきましては、すでに業界各方面から様々なご感想、ご意見をいただいておりますが、この度、社員の皆様の声をお聞かせいただきたいということで、下記の通り、モニターを募集いたします。ふるってご応募いただきますようお願いいたします。</p>	
記	
<p>モニター対象：「ガーリーパワー」のテレビコマーシャル モニター時期：〇月〇日（土）より〇月×日（日） モニター事項：①同業他社商品と比較したアピール度 ②同業他社商品と比較した好感度 ③改善点</p>	
<p>以上の3点につきまして、部署名および氏名を明記の上、メールまたはFAXで、広報部までご意見をお寄せください。 締切は〇月△日（水）とします。</p>	
以上	

1. 自社製品
2. 他社製品
3. テレビ広告
4. テレビ番組

資料5 第6章で使用した審査表

【審査方法についての説明書き（協力依頼文書）】

調査協力者各位

筑波大学 人文社会学研究科 国際日本研究専攻 小野塚 若菜

このたびは、調査にご協力いただきありがとうございます。以下の事項をお読みにになり、テスト問題の審査をお願いいたします。内容や方法についてご不明な点はいつでもお問い合わせください。

1. 調査の内容

本調査ではビジネス日本語テストの問題作成の専門家の皆様に、ビジネス日本語テストの問題ひとつにつき、以下のふたつの観点から審査をしていただきます。

- ① 日本語能力が同じであっても、性別（男／女）によって正答率に違いがあると予想されるか。
違いがあるとしたら、どちらのほうが正答率が高いと考えるか。
 - ② 日本語能力が同じであっても、居住地（日本国内／国外）によって正答率に違いがあると予想されるか。 日本国内の受験者のほうが正答率が高いと考えるか。
- ①および②について、正答率に違いがあると考える場合、その理由としてどのようなことが挙げられるか。

2. 方法

調査用紙には、テスト問題の視覚情報と音声スクリプト*（聴解および聴読解の場合）が書かれています。各問題について、上記①②の審査をしてください。

①の性別については、「男性のほうが正答率が高い（と予想する）」から「女性のほうが正答率が高い（と予想する）」の5段階、②の居住地については、「正答率に違いはない（と予想する）」から「国内受験者のほうが正答率が高い（と予想する）」の3段階で、その正答率の違いの程度を主観的に審査し、該当する番号の列に「○」を記入してください。なお、「正答率に違いがない（と予想する）」以外を選んだ場合は、右のコメント欄に、どのような点で正答率が高いと予想したのか、その理由を後で確認できるようにできるだけ具体的に書いてください。

*音声スクリプトには、話者の声の調子や効果音の情報（ex.紙をめくる音）、また、解答にかかる秒数の情報は書かれていません。審査をする上でこれらの情報が必要な場合は、音声ファイルを再生してご確認ください。

審査表.xlsx

調査用紙（テスト問題）

1.

□

2.

□

QID	部門	UID	問題名称	性別によって正答率に違いがあるかどうか				
				男性の正答率が高い	正答率に違いはない	女性の正答率が高い		
例	聴解 S1		書き方の例		○			
1	聴解 S1	1	機材チェック					
2	聴解 S1	2	名刺交換					
3	聴解 S1	3	社員紹介reform					
4	聴解 S1	4	ゼール部内会議					
5	聴解 S1	5	上司の電話を気にする					
6	聴解 S1	6	会議室で説明					

問題を見ながら、Excelの「審査表」に、ひとつひとつ入力してください。コメント欄のセルは入力すると自動的に高さが調整されます。

3. 送付物

- 1) PDF ファイル「調査用紙」（テスト問題）
- 2) Excel ファイル「審査表」（「性別」と「居住地」2枚のシートがあります）

【性別による DIF】

GID	部門	UID	問題名称	性別によって正答率に違いがあるかどうか			(「0:正答率に違いはない」以外の場合) どのような点でそのように予想したか
				男性のほうが正答率が高い	正答率に違いはない	女性のほうが正答率が高い	
				← -2	-1 0 1 2 →		
例	聴解	\$1	例 書き方の例		○		
1	聴解	\$1	1 機材チェック				
2	聴解	\$1	2 名刺交換				
3	聴解	\$1	3 社員紹介reform				
4	聴解	\$1	4 ビール部内会議				
5	聴解	\$1	5 商談終了				
6	聴解	\$1	6 会議室で説明				
7	聴解	\$2	1 アドなし訪問				
8	聴解	\$2	2 お大事に				
9	聴解	\$2	3 荷物受け取り				
10	聴解	\$2	4 新商品の説明				
11	聴解	\$2	5 プレゼン開始reform				
12	聴解	\$2	6 資料請求				
13	聴解	\$3	1 手伝う				
14	聴解	\$3	2 言付け				
15	聴解	\$3	3 請求書へのクレーム				
16	聴解	\$3	4 契約決着待ち				
17	聴解	\$3	5 社長挨拶reform				
18	聴解	\$3	6 カスタマイズ				
19	聴解	\$3	7 パソコンの増産体制				
20	聴解	\$3	8 カーテンの在庫数reform				
21	聴解	\$3	9 派遣社員の採用				
22	聴読解	\$1	1 大阪				
23	聴読解	\$1	2 調子はどう?				
24	聴読解	\$1	3 ミミ縮めて				
25	聴読解	\$1	4 社員募集要項reform				
26	聴読解	\$1	5 前売				
27	聴読解	\$1	6 将来なりたい職業				
28	聴読解	\$1	7 野菜を買うときに重視すること				
29	聴読解	\$1	8 健康診断				
30	聴読解	\$1	9 建物賃貸契約書				
31	聴読解	\$2	1 販売者会議reform				
32	聴読解	\$2	2 会議の議題				
33	聴読解	\$2	3 祝賀会挨拶				
34	聴読解	\$2	4 新製品の評価				
35	聴読解	\$2	5 新入社員面接				
36	聴読解	\$2	6 見積書の返事				
37	聴読解	\$2	7 プレゼンセミナー				
38	聴読解	\$2	8 飛行機のチケット				
39	聴読解	\$2	9 記事のファイル				
40	読解	\$1	1 売上高				
41	読解	\$1	2 所在地				
42	読解	\$1	3 通勤				
43	読解	\$1	4 閉店				
44	読解	\$1	5 納得できない				
45	読解	\$1	6 水増し				
46	読解	\$2	1 ちよつと一杯				
47	読解	\$2	2 ご迷惑				
48	読解	\$2	3 変な音				
49	読解	\$2	4 ご存知ですか				
50	読解	\$2	5 当社にお越しいただければ				
51	読解	\$2	6 認める				
52	読解	\$3	1 仕事の引き継ぎ				
53	読解	\$3	2 定期購読				
54	読解	\$3	3 社長からのメール				
55	読解	\$3	4 炊飯器たけるくん				
56	読解	\$3	5 販売報告会reform				
57	読解	\$3	6 フレックスタイム制度				
58	読解	\$3	7 日米技術交渉対談記事				
59	読解	\$3	8 商品破壊				
60	読解	\$3	9 ガーリーパワー				

【居住地による DIF】

GID	部門	UID	問題名称	居住地によって正答率に違いがあるかどうか 正答率に違いはない 国内受験者のほうが正答率が高い			(「0:正答率に違いはない」以外の場合) どのような点でどのように予想したか
				0	1	2	
例	聴解	§ 1	例 書き方の例			○	写真のビジネス場面は、日本国外では見聞きすることはほとんどないと思う。
1	聴解	§ 1	1 機材チェック				
2	聴解	§ 1	2 名刺交換				
3	聴解	§ 1	3 社員紹介reform				
4	聴解	§ 1	4 ビール部内会議				
5	聴解	§ 1	5 商談終了				
6	聴解	§ 1	6 会議室で説明				
7	聴解	§ 2	1 アポなし訪問				
8	聴解	§ 2	2 お大事に				
9	聴解	§ 2	3 荷物受け取り				
10	聴解	§ 2	4 新商品の説明				
11	聴解	§ 2	5 プレゼン開始reform				
12	聴解	§ 2	6 資料請求				
13	聴解	§ 3	1 手紙う				
14	聴解	§ 3	2 言付け				
15	聴解	§ 3	3 請求書へのクレーム				
16	聴解	§ 3	4 契約決済待ち				
17	聴解	§ 3	5 社長挨拶reform				
18	聴解	§ 3	6 カスタマイズ				
19	聴解	§ 3	7 パソコンの増産体制				
20	聴解	§ 3	8 カーテンの在庫数reform				
21	聴解	§ 3	9 派遣社員の採用				
22	聴読	§ 1	1 大阪				
23	聴読	§ 1	2 調子はどう?				
24	聴読	§ 1	3 リミ縮めて				
25	聴読	§ 1	4 社員募集要項reform				
26	聴読	§ 1	5 前売				
27	聴読	§ 1	6 将来なりたい職業				
28	聴読	§ 1	7 野菜を買うときに重視すること				
29	聴読	§ 1	8 健康診断				
30	聴読	§ 1	9 建物賃貸契約書				
31	聴読	§ 2	1 販売者会議reform				
32	聴読	§ 2	2 会議の議題				
33	聴読	§ 2	3 祝賀会挨拶				
34	聴読	§ 2	4 新製品の評価				
35	聴読	§ 2	5 新入社員面接				
36	聴読	§ 2	6 見積書の返事				
37	聴読	§ 2	7 プレゼンセミナー				
38	聴読	§ 2	8 飛行機のチケット				
39	聴読	§ 2	9 記事のファイル				
40	読解	§ 1	1 売上高				
41	読解	§ 1	2 所在地				
42	読解	§ 1	3 通勤				
43	読解	§ 1	4 閉店				
44	読解	§ 1	5 納得できない				
45	読解	§ 1	6 水増し				
46	読解	§ 2	1 ちよつと一杯				
47	読解	§ 2	2 ご迷惑				
48	読解	§ 2	3 変な音				
49	読解	§ 2	4 ご存知ですか				
50	読解	§ 2	5 当社にお誘いいただければ				
51	読解	§ 2	6 認める				
52	読解	§ 3	1 仕事の引き継ぎ				
53	読解	§ 3	2 定期購読				
54	読解	§ 3	3 社長からのメール				
55	読解	§ 3	4 炊飯器たけるくん				
56	読解	§ 3	5 販売報告会reform				
57	読解	§ 3	6 フレックスタイム制度				
58	読解	§ 3	7 日米技術交渉対談記事				
59	読解	§ 3	8 商品破壊				
60	読解	§ 3	9 ガーリーパワー				