

TMU聴解テストの開発について

神村 初美* 劉 永亮* 柳 悦* 彭 韻* 林 香淑*

神谷 英里* 陸 黎莉** 十市 佐和子*** 西郡仁朗****

1. はじめに

本稿では、「TMU 聴解テスト」の開発について報告する。同テストは本学留学生のためのプレースメントテストとしての使用や、本学の日本語教育プログラムの一環として国内外での聴解能力の測定と分析などに利用されることとなる。そのため、できるだけ正確な聴解能力測定ができるように問題案の作成の際に留意した点、これを試行テストとして用いた場合の統計的結果、ならびにその結果から最終的に選出された聴解問題の信頼性を報告する。

2. 聴解テスト開発の背景

2-1. 本学での標準的な日本語テストの必要性

本学には 2009 年度に国際センターが設立され、留学生の受け入れ推進と日本語教育の充実が喫緊の課題となっている。留学生に対する日本語教育としては、学部留学生については、正規の言語科目「日本語Ⅰ」「日本語Ⅱ」があり、上級以上のレベルの学習者に対し、大学での勉強や発表・レポート作成等に必要な日本語教育を行っている。学部留学生の場合は、入学時に上級を超える日本語能力を有することが前提になっており、学習者の日本語能力について一定以上のレベルを想定することが可能である。それ以外の学習者、具体的には交換留学生（特別科目等履修生）、大学院生、研究

*首都大学東京 大学院生

**首都大学東京 研究生

***首都大学東京 国際センター 日本語講習会講師

****首都大学東京

生に関しては日本語のレベルはまちまちである。彼らに対する日本語のクラスは学生サポートセンター主催の「日本語講習会」という非正規の形態で行われてきたが、2010年度からは国際センターが「基礎日本語」として多様なレベルと内容の日本語クラスを設定し、センターが各レベルの修了認定を独自で行うこととなっている。その前提として個別の背景や能力・到達目標・学習環境などについてのディネス調査やプレースメントが必要である。今回の聴解試験の作成には、こうした全学的な背景がある。

こうした背景のもとでの試験開発なので、聴解試験だけでは不十分であり、今後は読解・文法の試験も開発する予定である。今回、聴解を優先したのは、本学日本語教育学教室で、これまで聴解問題開発と分析の実績が積み上げられ（西郡・宮田, 2003 ; 西郡, 2004）、開発体制が比較的整っていたためである。

また、本聴解テストでは、後に詳述するように日本語能力試験2級、すなわち、所謂中級を修了したレベルの能力測定が目的となっている。試験のレベルについても多様なものが求められるが、これまでの本学の留学生受け入れにおけるレベル測定の必要性やプレースメントでの試験の有用性を勘案するとこのレベルを優先的に開発すべきであると判断された。

2-2. 『日本語能力試験』について

今回開発した聴解試験の問題項目を作成する際にモデルとしたのは、『日本語能力試験 (Japanese Language Proficiency Test, 略称 JLPT) ・聴解』である。この日本語能力試験は、財団法人日本国際教育支援協会と独立行政法人国際交流基金が主催している日本語を母語としない人を対象に日本語能力を認定する公的な検定試験である¹。2008年時点で、日本を含め世界52の国と地域で試験を受けることが出来る大規模試験であり、世界で35万8千人あまりが受験している²。同試験のテストシラバスである『日本語能力試験出題基準』（日本国際教育協会³・国際交流基金, 1994 ; 以下『出題基準』）によると、聴解テストの出題についての基本的な理念では、日本語

¹ 国内の実施は日本国際教育支援協会、海外での実施は国際交流基金が行っている。

² 日本語能力試験公式サイト <http://www.jlpt.jp> (2009年12月29日アクセス)。

³ 同協会は2004年に改組され日本国際教育支援協会となった。

学習者が会える日常生活の中で解決を求められる聴解の課題を想定してそれを遂行する能力を測定することが標榜されており、「1. 学習者が教室を含めた日常生活で会える課題と同様の課題を解決する。2. 情報の正確な聞き取りにとどまらず、聞いて何かをすること、つまり広い意味での問題解決を求めるもの」であることが特徴であるとされている。

本学における留学生のためのプレースメントテストならびに本学関係の国内外での聴解能力分析も、大学での勉強だけでなく、学生という生活者として日常生活の中で解決が求められる課題を解決・遂行する面が重要であると考えられたので、『出題基準』での指標を重要な参考とした。

なお、『日本語能力試験』は2008年度までは試験問題が実施後公開されていた。世界の日本語学習者にとって日本語能力試験の過去問題は容易に入手でき、また、各教育機関でも頻繁に利用されているものなので、目や耳に触れる機会があまりに多く、プレースメントテストなどで利用することはできないものであった。また、2010年度からは試験全体が改定されることとなり、その一環として尺度の標準化のため、実施後も問題は公開されず、また、『出題基準』も公刊されないこととなっている（『新しい「日本語能力試験」ガイドブック』日本国際教育支援協会・国際交流基金, 2009）。こうした点を見ても、今回の聴解テスト開発のように、各教育機関で独自の試験の作成が求められている状況であると思われる。

3. 問題項目開発の方法

3-1. 概要

今回の聴解テストの問題項目案作成は、日本語教育学を学ぶ大学院生のプロジェクトワークとして行われ⁴、参加者が『日本語能力試験出題基準』（1994）における「聴解試験課題一覧」2級で強調されている点、西郡・宮田（2003）による聴解問題作成の留意点を参照した上で作成した素案を持ち寄った。グループによる検討、統括する教員からのコメントを受け改良・修正を経て問題案Aセット26問、問題案Bセット26問の合計52問の

⁴首都大学東京大学院人文科学研究科日本語教育学教室「日本語教育学研究・特論 談話と音声教育2」の2009年度前期コースワークの一環として行われた。

スクリプトを作成した。その後、東京方言話者が問題項目の発話を担当し、日本語教育関係者が録音を担当する作業を行った。これらの問題項目について国内外で試行テスト実施、項目分析を行った。その結果、算出された正答率と識別力などから優良問題 15 問と良問題 15 問を選出した。

3-2. 問題項目案作成の留意点

プロジェクトワークとして聴解問題項目を作成する際、先述の通り、『日本語能力試験出題基準』（1994）における聴解試験シラバスである「1. 試験課題基準一覧」および西郡・宮田（2003）による聴解問題作成の留意点を参考に問題作成の素案を検討するとともに聴解問題の改良を重ねた。

表 1 『日本語能力試験・聴解』2級出題における重要課題の基準一覧

A. 情報完成
A-2. 情報完成による論理的判断
B. 規則・法則・傾向の適用/検証
B-1. 規則・法則・傾向の具体例への適用
B-2. 規則・法則・傾向の検証/修正
B-3. 具体例に基づく規則・法則傾向の判断
C. 情報統合
C-1. 情報の比較（相違点）
C-2. 情報の比較（共通点）
C-3. 情報の統合
D. 意図の汲み取り
D-2. 意向の判断
E. 照合
E-2. 一つの事物・人物の照合
F. 順序の再構成
F-1. 時間的順序の再構成
F-2. 手順・手続きの再構成

『日本語能力試験出題基準』（1994）「V 聴解試験課題の解説」より

『日本語能力試験出題基準』「1. 試験課題基準一覧」は、日本語能力試験の聴解試験として行われる試験問題の類型を示したものである。出題者が問題作成にあたってこのリストをひとつの目安として聴解問題を考案するリソースでもある。この試験課題基準一覧表において、特に2級において中心的に出題される課題として記されている点に留意した（表1参照）。また、西郡・宮田(2003)の留意点は『日本語能力試験出題基準』と実際の聴解問題制作の長年の経験を踏まえてまとめられたものである。

表2 西郡・宮田(2003)における聴解問題作成の留意点

- A. 語彙・文型以上のコミュニケーション能力の要素を盛り込む。
- B. はじめにプリクエスションを設け、場面と課題を必ず提示する。
- C. 特定の知識や思想・観念・文化背景が必要な問題は避ける。
- D. 倫理的な問題に配慮する。
- E. 聴解能力以外の能力に大きく依存した問題は避ける。
- F. 日本語の話し言葉の特徴を積極的に取り入れる。
- G. 音声聴取上の問題を考慮する。
- H. その他
 - ・四肢択一課題とする。
 - ・聴取者が登場人物の一人に自分を投影できるもの
(原則上盗み聞きできないもの)
 - ・学習者の母語、男女、年齢による理解の差が出ないもの。
 - ・日本語音声を聞き分けるだけのものは避ける。
 - ・正解があいまいになりやすいものは避ける。
 - ・場面設定が長すぎるものは避ける。
 - ・問題を聞かなくても常識で答えられるものではないこと。

3-3. 聴解問題項目案の録音と編集

問題項目案の実際の録音に当たっては、音声録音小スタジオにおいて20代から40代の東京方言話者女性4名、男性4名が聴解問題の発話を担当し

た。発話者は日本語教育関係および大学の劇団に所属する者であり、ある程度のアナウンス訓練経験を有している。録音と編集は日本語教育関係者で視聴覚機器の技術訓練を受けた者が当たった。使用機器は、marants 社製ポータブル・ソリッドステート・レコーダーPMD660, RODE 社製コンデンサーマイク NT100 などである。音声の編集に関しては、Cycle of%th 社製フリーウェア Sound Engine Free を用いた。

4. 試行テストとその項目分析結果

4-1. 試行テストの実施

今回制作した聴解問題項目案は、2009年6月下旬から7月初旬にかけて海外4機関、国内1機関、計5つの日本語教育機関において延べ総数476名を対象に試行テストを行った⁵。

被験者は、合計242名であり、日本語学校または大学において中級レベルの日本語を集中的に学んでいる学習者である。同年度末の日本語能力試験の受験を予定していた者も多く、テストへの参加意欲は比較的高かったものと思われる。

これら被験者が学んでいる日本語教育機関は、日本、韓国、中国に所在し、国籍は、中国、韓国の大学では現地国籍であるが、日本国内の実施機関での被験者国籍は韓国、台湾、米国、アルゼンチン、イギリス、イタリア、シンガポール、ドイツ、タイ、フランス、マケドニアであった。被験者の母語は、中国語、韓国・朝鮮語、英語、イタリア語、スペイン語、タイ語、ドイツ語、フランス語と多岐にわたっていた。試行テストで使用したセットおよび各機関での被験者数などの内訳を表3に記す。

試験は、CDプレーヤーを用いて再生し、教室内に流すという集団実験方式をとった。LL教室など音響的な環境がよい場所で行うことを依頼したが、機関によっては一般教室を使用したところもあった。しかし、音声の聴取について問題があったという報告はなかったので、聴取上の問題は

⁵ 協力機関(協力者)は、中国・四川大学(車小平氏)東北林業大学(薄紅昕氏)、韓国・中央大学校(任栄哲氏・磯野英治氏)仁川専門大学(金慶恵氏)、学校法人長沼スクール(鈴木潤吉氏・岩田之男氏)である。記して感謝する。

なかったと思われる。

なお、試行テストに当たっては各機関において、被験者の半数がAセット（26問）、Bセット（26問）の順で、半数がその逆の順で受験し。順序性の影響が出ないように配慮した。

表3 試行テスト実施概要

	実施校	Aセット	Bセット	合計
日本	JN	54	54	108
中国	CT	60	60	120
	CS	25	25	50
韓国	KS	55	55	110
	KI	48	40	88
合計		242	234	476

(2009年6月下旬～7月初旬試験実施)

4-2. 試行試験全データの項目分析

一般的なテストの問題項目の適切さを吟味する方法に従い、項目の難易度と弁別力⁶を使用した項目分析を行った。52の問題項目案（Aセット26問、Bセット26問）、述べ476データの結果を表4と表5に記す。なお、難易度については各項目案の正答率、弁別力には各項目案の得点とテスト全体の得点との相関を表す点双列相関係数を使用して検証を行った。

4-3. 優良問題・良問の選定方法

石田（1992）によると望ましい項目における正答率は20%から80%の間であり、かつ弁別力が0.4以上とされている。

⁶できる者とできない者を選び分ける程度を表すもので、識別力とも言う。弁別指数または、点双列相関係数で表される。弁別指数は、得点の上位群の正答率と下位群の正答率の差であり、点双列相関係数は各項目の得点とテスト全体の得点との相関を表す。本稿では後者を採用している。

表4 Aセットの項目分析結果一覧表

問題 番号	受験 者数	正答 者数	標準 偏差	正答率	識別 力	選択率				
						1	2	3	4	無回答
*A-1	242	167	0.462	0.690	0.600	9%	<u>69%</u>	19%	2%	2%
*A-2	242	136	0.496	0.562	0.516	11%	<u>56%</u>	25%	7%	1%
**A-3	242	181	0.434	0.748	0.396	5%	<u>75%</u>	12%	6%	2%
**A-4	242	115	0.499	0.475	0.395	12%	13%	26%	<u>48%</u>	1%
*A-5	242	176	0.445	0.727	0.418	4%	6%	17%	73%	1%
A-6	242	106	0.496	0.438	0.356	<u>44%</u>	16%	23%	17%	1%
A-7	242	135	0.497	0.558	0.345	10%	19%	56%	14%	2%
*A-8	242	166	0.464	0.686	0.474	6%	10%	<u>69%</u>	13%	2%
*A-9	242	154	0.481	0.636	0.551	<u>64%</u>	8%	13%	13%	2%
*A-10	242	164	0.467	0.678	0.473	3%	24%	<u>68%</u>	4%	1%
*A-11	242	191	0.408	0.789	0.534	7%	<u>79%</u>	8%	5%	1%
**A-12	242	108	0.497	0.446	0.382	15%	31%	<u>45%</u>	8%	1%
*A-13	242	168	0.461	0.694	0.563	<u>69%</u>	13%	5%	12%	1%
*A-14	242	161	0.472	0.665	0.547	19%	8%	6%	<u>67%</u>	1%
*A-15	242	158	0.476	0.653	0.600	<u>65%</u>	6%	11%	16%	2%
*A-16	242	115	0.499	0.475	0.594	3%	<u>48%</u>	29%	20%	0%
*A-17	242	143	0.492	0.591	0.583	7%	14%	59%	19%	1%
*A-18	242	163	0.469	0.674	0.430	1%	14%	<u>67%</u>	18%	0%
*A-19	242	191	0.408	0.789	0.558	5%	79%	6%	9%	1%
**A-20	242	144	0.491	0.595	0.384	5%	<u>60%</u>	14%	21%	0%
*A-21	242	143	0.492	0.591	0.425	5%	3%	32%	<u>59%</u>	1%
*A-22	242	146	0.489	0.603	0.412	16%	8%	15%	<u>60%</u>	0%
*A-23	242	183	0.429	0.756	0.640	5%	5%	<u>76%</u>	14%	0%
*A-24	242	141	0.493	0.583	0.553	3%	12%	26%	<u>58%</u>	1%
A-25	242	204	0.364	0.843	0.442	2%	9%	4%	<u>84%</u>	1%
A-26	242	199	0.382	0.822	0.366	2%	11%	82%	5%	1%

(クロンバックの α 係数⁷=0.8666, *については表7参照, 下線部は正答選択肢)

⁷内的整合性の観点から見た、テストの測定値の信頼性係数。

表6 Bセットの項目分析結果一覧表

問題 番号	受験 者数	正答 者数	標準 偏差	正答率	識別 力	選択率				
						1	2	3	4	無回答
B-1	234	185	0.407	0.791	0.213	6%	<u>79%</u>	8%	6%	0%
B-2	234	191	0.387	0.816	0.481	3%	<u>82%</u>	6%	9%	0%
*B-3	234	178	0.427	0.761	0.657	3%	13%	<u>76%</u>	7%	1%
*B-4	234	155	0.473	0.662	0.622	5%	<u>66%</u>	12%	17%	0%
B-5	234	180	0.421	0.769	0.367	3%	10%	9%	<u>77%</u>	1%
*B-6	234	125	0.499	0.534	0.540	10%	22%	13%	53%	1%
B-7	234	60	0.437	0.256	0.292	46%	8%	18%	<u>26%</u>	2%
*B-8	234	129	0.497	0.551	0.420	4%	55%	14%	26%	1%
*B-9	234	123	0.499	0.526	0.521	13%	18%	14%	53%	2%
*B-10	234	185	0.407	0.791	0.588	3%	7%	10%	79%	0%
*B-11	234	161	0.463	0.688	0.526	8%	16%	69%	6%	2%
B-12	234	46	0.397	0.197	0.365	<u>20%</u>	20%	38%	21%	1%
*B-13	234	130	0.497	0.556	0.537	21%	<u>56%</u>	11%	12%	0%
*B-14	234	114	0.500	0.487	0.503	2%	33%	49%	16%	0%
*B-15	234	128	0.498	0.547	0.664	15%	<u>55%</u>	22%	7%	1%
*B-16	234	162	0.462	0.692	0.393	14%	69%	12%	4%	1%
**B-17	234	165	0.456	0.705	0.386	10%	9%	<u>71%</u>	10%	1%
*B-18	234	162	0.462	0.692	0.628	6%	69%	19%	4%	1%
B-19	234	75	0.467	0.321	0.212	8%	15%	32%	44%	0%
B-20	234	103	0.496	0.440	0.354	17%	22%	<u>44%</u>	16%	1%
*B-21	234	142	0.488	0.607	0.496	9%	<u>61%</u>	8%	22%	0%
*B-22	234	96	0.492	0.410	0.525	13%	<u>41%</u>	18%	27%	0%
B-23	234	19	0.273	0.081	0.173	<u>8%</u>	12%	32%	48%	0%
*B-24	234	134	0.495	0.573	0.502	11%	57%	18%	14%	0%
*B-25	234	114	0.500	0.487	0.581	<u>49%</u>	28%	9%	14%	1%
*B-26	234	129	0.497	0.551	0.446	5%	55%	16%	23%	1%

(クロンバックの α 係数=0.8556,*については表7参照, 下線部は正答選択肢)

石田（1992）は、どのくらいの信頼性係数が得られればそのテストは信頼できるかは、研究者によって意見が異なり、また、テストの種類によっても望ましい係数は異なってくるとしながらも、聴解力テストの望ましい数値として 0.80～0.89（Lado1961：332）をあげている。

そこで本テスト開発では、優良問題と良問題を選出するに当り、石田（1992）によるこれら望ましい項目の基準を援用し、表 7 のように「優良問題選出基準」を設けた。優良・良問題とする対象は、これらの基準を通過した問題のみとした。

表 7 優良問題選出基準

優良問選出基準	正答率	識別力
基準 1 *	0.2～0.8	0.4 以上
基準 2 **	0.2～0.8	0.39
基準 3 ***	0.2～0.8	0.38

これに従うと、基準 1 では 34 問（A：18 問 B：16 問）、基準 2 は 3 問（A：2 問 B：1 問）、基準 3 においても 3 問（A：2 問 B：1 問）の合計 40 問（A：22 問 B：18 問）が問題項目として望ましいという結果が計量的に算出された。各問の各種計数については、表 8 を参照されたい。

4-4. 項目分析結果に関する質的な信頼性と妥当性の検証

計量的検証から選出された優良問題・良問題の候補 40 問に対しさらに質的な検証を加えた。これは、項目分析において数値的に望ましい問題項目の基準を備えていると確認された選出問題にさらに質的な信頼性と妥当性の検証を加えることによって、より精度の高い聴解問題を抽出するためである。これまでの日本語能力試験の問題数や本学を含む各機関でのプレースメントテストの実施時間などの経験を踏まえると、優良問題 15 問、良問題 15 問を選出することが適当であると判断された。

この作業のため、まず開発グループによる検討をおこなった。8 名が 2 グループ（各 4 名）に別れ、各グループ毎に下記項目（参考：西郡・宮田，2003）

を質的検証の基準とした上で、討論を通し、上記の計量的検証第一段階にて選出された対象 40 問に対する検証を行った。その結果、それぞれのグループごとに優良問題 15 問および良問題 15 問に絞り込んだ。

表 8 優良問題 15 問・良問題 15 問の項目分析結果表
(A セットと B セットから選出統合)

水準	問題 番号	受験 者数	正答 者数	標準 偏差	正答 率	識別 力	選択率				
							1	2	3	4	無回答
優良	A-1	242	167	0.462	0.690	0.600	9%	69%	19%	2%	2%
優良	A-2	242	136	0.496	0.562	0.516	11%	56%	25%	7%	1%
優良	A-9	242	154	0.481	0.636	0.551	64%	8%	13%	13%	2%
優良	A-10	242	164	0.467	0.678	0.473	3%	24%	68%	4%	1%
優良	A-15	242	158	0.476	0.653	0.600	65%	6%	11%	16%	2%
優良	A-17	242	143	0.492	0.591	0.583	7%	14%	59%	19%	1%
優良	A-18	242	163	0.469	0.674	0.430	1%	14%	67%	18%	0%
優良	A-22	242	146	0.489	0.603	0.412	16%	8%	15%	60%	0%
優良	B-4	234	155	0.473	0.662	0.622	5%	66%	12%	17%	0%
優良	B-8	234	129	0.497	0.551	0.420	4%	55%	14%	26%	1%
優良	B-9	234	123	0.499	0.526	0.521	13%	18%	14%	53%	2%
優良	B-15	234	128	0.498	0.547	0.664	15%	55%	22%	7%	1%
優良	B-18	234	162	0.462	0.692	0.628	6%	69%	19%	4%	1%
優良	B-24	234	134	0.495	0.573	0.502	11%	57%	18%	14%	0%
優良	B-26	234	129	0.497	0.551	0.446	5%	55%	16%	23%	1%
良	A-4	242	115	0.499	0.475	0.395	12%	13%	26%	48%	1%
良	A-8	242	166	0.464	0.686	0.474	6%	10%	69%	13%	2%
良	A-13	242	168	0.461	0.694	0.563	69%	13%	5%	12%	1%
良	A-14	242	161	0.472	0.665	0.547	19%	8%	6%	67%	1%
良	A-20	242	144	0.491	0.595	0.384	5%	60%	14%	21%	0%
良	A-21	242	143	0.492	0.591	0.425	5%	3%	32%	59%	1%
良	A-24	242	141	0.493	0.583	0.553	3%	12%	26%	58%	1%
良	B-6	234	125	0.499	0.534	0.540	10%	22%	13%	53%	1%
良	B-11	234	161	0.463	0.688	0.526	8%	16%	69%	6%	2%
良	B-13	234	130	0.497	0.556	0.537	21%	56%	11%	12%	0%
良	B-14	234	114	0.500	0.487	0.503	2%	33%	49%	16%	0%
良	B-16	234	162	0.462	0.692	0.393	14%	69%	12%	4%	1%
良	B-21	234	142	0.488	0.607	0.496	9%	61%	8%	22%	0%
良	B-22	234	96	0.492	0.410	0.525	13%	41%	18%	27%	0%
良	B-25	234	114	0.500	0.487	0.581	49%	28%	9%	14%	1%

- ① 問いまたは回答選択肢に1級・級外の語彙が含まれている場合はその問題を排除。
- ② 会話文内に1級・級外の語彙が含まれている場合はその難易度、正答への影響力を検討。
- ③ 問題の内容および類型に関し、『日本語能力試験出題基準』「V. 聴解試験課題の解説 1. 試験課題一覧」(pp205-217)の出題課題における重要度(網掛け濃度)を目安として精査を行う。
- ④ 項目分析結果に見られた天井効果および床面効果を検証。

次に全体での検証を行った。各グループ内の討論により絞り込んだ優良問題・良問題を照らし合わせ、まず、両グループとも選出結果が一致した15問(優良問題または良問題での一致)を優良問題として選出した。さらに参加者全員により上述の選出基準を踏まえ、さらに内容に関する妥当性の討論精査を残りの問題項目に対し行い良問題15問を選出した。

優良問題と良問題だけを取り出して集計したクロンバックの α 係数は0.85となっている。

これらの問題については、テストの性質上内容は公開できないが、本学日本語教育学教室で管理しているので、学内外で必要な場合には対応を協議する。

5. おわりに

これまで述べてきたような手法で開発された『TMU聴解テスト』は、上述した一連の多角的な検証をもって日本語聴解能力を測る際の妥当性と信頼性を満たしていると思われる。本学の留学生に対するプレースメントテストだけでなく、さまざまな利用が可能であろう。例えば、『日本語能力試験の概要』(2003)によると、海外受験者と日本国内受験者との間で、もっとも大きな差の出る科目(類)は『聴解』の分野であるとされる。その原因を詳細に検討する際に、同様の良質の試験を用いて量的かつ質的な分析を行う方法があろう。しかし、その場合、日本語能力試験の過去の問題では、

多くの学習者に知られすぎており適当ではない。本TMU聴解テストは、こうした場合での利用が可能である。

今回の試行テストも、日本国内、韓国、中国で行われ、詳細なデータが残されている。この分析については別稿に譲る。また、今後、このテストの実施と結果の分析を繰り返し、教育研究に利用できる様々な知見を積み上げていきたい。

参考文献

- 石田敏子(1992)「第17章 テストの改良」『入門日本語テスト法』pp. 190-203
大修館書店
- 小野塚若菜・島田めぐみ(2008)『日本語教師のためのExcelでできる分析入門』スリーエーネットワーク
- 海保博之・柏崎秀子(2002)「13章 日本語能力を評価する」pp. 197-206, 「付録 テスト改善に役立つ技法 S・P表による分析と基本統計」pp. 207-219 『日本語教育のための心理学』新曜社
- 国際交流基金・日本国際教育協会(1994)『日本語能力試験出題基準』凡人社
- 国際交流基金(2004)『日本語能力試験の概要 2003年版』凡人社
- 国際交流基金・日本国際教育協会(2009)『新しい「日本語能力試験」ガイドブック』
- 坂上恵・中村則子・大木理恵(2009)「上級聴解試験評価—学習者の聴読解能力を適正に反映する題とはどのような問題か—」『東京外国語大学留学生日本語教育センター論集』35号 pp. 103-109
- 島田めぐみ(2004)「日本語聴解テストの項目正解率に影響を与える要因」『東京学芸大学紀要2部門 55』 pp. 145-153
- 西郡 仁朗・宮田 剛章(2003年)「上級レベル聴解素材のWEB公開と項目分析による素材の評価」『日本語研究』22号, 東京都立大学国語学研究室, pp. 155-171
- J. D. ブラウン(著)・和田稔(訳)(1999)『言語テストの基礎知識—正しい問題作成・評価のために—』大修館書店