初中級レベルの中国語文法能力テストの開発
―日本人中国語学習者のデータによる評価―[1]

張　婧禕[2]
玉岡　賀津雄[3]

DOI: 10.18999/stul.35.51

要約:本研究は，中国語を第2外国語として学ぶ日本人学習者の文法能力を測定するためのテストを開発した。そして，それを日本人中国語学習者に実施し，結果および評価を報告した。本文法テストは，中国語検定試験の準4級および4級レベルに準じて，四者択一の形式で 32 問を設けた。これらの 32 問は，配当級で分けると，準4級に相当する 16 問，4級に相当する 16 問，表現形式で分けると，1文の表現が 16 問，1往復の会話が 16 問になる（表1を参照）。さらに，測定内容からは，品詞の理解と文型の理解の2つの下位分類で，それぞれ 16 問で構成されている。このテストを，入門から初中級レベルに相当する日本人中国語学習者 154 名に実施した。その結果，テストの総得点は，7点から 32 点までに分布し，平均が 16.47 点で，標準偏差が 3.28 点であった。さらに，テストの各項目について，項目応答理論に基づいて，項目困難度，項目実質選択肢数，項目弁別力の3つの指標を計算した。これらの指標を総合的に考察した結果，この文法能力テストは弁別力が高く，実用的であると判断された。最後に，この中国語文法能力テストをオンラインで自由に使えるように，無料でウエブサイトに公開し，今後の中国語教育に活用できるようにした（https://forms.office.com/r/iLkcyeBFqF）。

キーワード：中国語文法能力テスト　中国語教育　初中級　日本人中国語学習者
項目応答理論　項目困難度　項目実質選択肢数　項目弁別力

---

[1] English title: Development of a beginner-to-intermediate level Chinese grammatical knowledge test: Evaluation by data from native Japanese speakers learning Chinese

[2] ZHANG, Jingyi, University of Miyazaki, Japan, E-mail: jingyizhang@cc.miyazaki-u.ac.jp

[3] TAMAOKA, Katsuo, Hunan University, China; Nagoya University, Japan, E-mail: ktamaoka@gc4.so-net.ne.jp

1. 研究の目的

外国語教育においては，語彙と文法が独立した教授内容として教えられている。それは，語彙と文法がそれぞれ独立した能力であると考えられているからである。そして，語彙と文法の教授・学習効果を確認する手段として，テストが用いられる。中国語教育では，張・玉岡・勝川（2017）が語彙能力を独立した1つの変数として測定するために，中国語の語彙能力テストを開発した。そこで，本研究は，中国語の文法能力を独立した1つの変数として，最少の問題数で最大の学習成果を効率良く測定できるテストを目指し，入門から初中級レベルまでの（ここでは，中国語検定試験4級を初中級とし，以下では，「初中級レベル」とのみ呼ぶ）文法能力を判定するためのテストを開発した。項目応答理論（IRT: Item Response Theory）に基づいて開発されたプログラム（TDAP: Test Data Analysis Program）を用いて，日本語を母語とする中国語学習者に対して実施した後のデータを分析し，本文法テストの実用性，有効性，妥当性を総合的に検討した。

2. 中国語文法能力テストの作成手順および内容

テストを作成する際の目的からみると，テストはクラス編成・プレースメントテスト，進度測定用テスト，到達度確認用テスト，熟達度測定用テストと総合的な診断テストの5つの種類に分類される（オルダーソン他，2010；ブラウン，1999）。また，テストの測定方法では，多肢選択，空欄補充，組み合わせ法，書き換え，短文解答，小論文などの多様な問題形式がある。本文法テストは，初中級レベルまでに相当する学習者を対象に，中国語の文法能力がどのレベルに達しているかを測定する。そのために，テストにおける内容の信頼性，妥当性，実用性の3つの観点を考慮したうえで，上記の複数のテストの方法のうち，四者択一の多肢選択の出題形式を選定して，到達度確認用テストを作成する。本テストを作成する際には，初中級レベルの中国語学習者を実施対象とした。したがって，中国語検定の準4級および4級レベル（以下，準4級と4級と記す）に相当する文法項目の中から，項目として選択した。この文法能力テストで測定する項目は，配当級，設問形式および測定内容という3つの観点から統制した。配当級の観点からは，中国語検定試験に準じて，準4級と4級という2つのレベルに分ける。また，設問形式の観点からは，1文の長さで提示する形式および2人による1往復の会話で提示する形式という2つに分ける。本テストでは，前者の設問形式を「表

現」と，後者の設問形式を「会話」と呼ぶ。さらに，中国語は日本語と違って，孤立語であるため，形態素の変化はない。したがって，本テストでは，測定内容からは，文における語と語の繋がりと関連する「品詞の理解」，文と文の繋がりまたは中国語の文型を作ることに必要とする「文型の理解」という2つの測定内容に分類した。つまり，本テストにおける2つの配当級が準4級と4級で，2つの設問形式が表現と会話で，2つの測定内容が品詞の理解と文型の理解である。表1に示したように，本テストにおけるそれぞれのカテゴリー（下位分類）について各4問を設ける。そのため，総問題数は，2（配当級）×2（設問形式）×2（測定内容）×4問＝32 問となる。

表1 テストにおける中国語検定試験の配当級および下位分類

配当級	設問形式	測定内容	問題数
準４級	表現	品詞の理解	4
		文型の理解	4
	会話	品詞の理解	4
		文型の理解	4
４級	表現	品詞の理解	4
		文型の理解	4
	会話	品詞の理解	4
		文型の理解	4
合計			32

注:本テストでは，中国語検定試験の配当級（準4級・4級）に準じて作成した.

なお，文法能力のみを要因として測定するために，日本語学習者にとってすでに学習してじゅうぶんに理解していると思われる語彙を使って設問文を作成した。また，各設問文に対して，形態的・意味的・文法的役割に類似するもの3つを誤りの錯乱肢として作った。こうして作成した 32 問の設問文の各下位分類，正答と錯乱肢の詳細は表2に示した。たとえば，設問文の“A:这件 T 恤衫怎么样？B:我想试（　　）。”は4級レベルに相当する文法項目で，2人の1往復の会話形式で提示されたものである。その設問文を完成するため，学習者に最も相応しいものを4つの選択肢から1つだけ選んでもらった。ここで，“B 试一下（ちょっと試してみる）” に対して，“A 一点儿（ちょっと）” （文法的役割類似），“C 一个（一つ）”（形態的類似）および“D 了试（ちょっと～した）” （意味的類似）という3つの錯乱肢を用意して，ランダムに提示した。ここでは，“B 试一下 （ちょっと試してみる）”が正答である。

表２　文法テスト32問の配当級，設問文および錯乱肢

NO.	配当級	出題形式	設問文	正答	錯乱肢１	錯乱肢２	錯乱肢３
1	準４級	表現	他们不（　）是日本人。	也	都	多	太
2	準４級	表現	今天作业（　）多了。	很	不很	非常	太
3	準４級	表現	你（　）什么名字?	叫	是	姓	贵姓
4	準４級	表現	今天（　）冷。	一点儿	/	没	很
5	準４級	表現	她（　）我高。	比	跟	和	给
6	準４級	表現	学校旁边（　）公园?	在不在	有没有	没有没	有不有
7	準４級	表現	我妹妹（　）两把伞。	有	不有	在	不在
8	準４級	表現	冰箱（　）有一杯牛奶。	面	/	的里	里
9	準４級	会話	A:最近怎么样? B:最近我（　）忙。	有点儿	一点儿	一会儿	点儿
10	準４級	会話	A:那是什么？B:这是一（　）毛衣。	张	台	件	条
11	準４級	会話	A:今天星期（　）？B:今天星期三。	几	多长	多少	多久
12	準４級	会話	A:你家有几（　）人？B:我家有四（　）人	口	个	位	只
13	準４級	会話	A:你喝咖啡（　）喝红茶? B:我喝红茶。	还是	是不是	还不	喝不
14	準４級	会話	A:那本词典（　）哪里？B:那本词典（　）桌子上。	有	在	没在	没有
15	準４級	会話	A:你们的作业（　）？B:我们的作业不多。	多不多	不多不	多没多	没多没
16	準４級	会話	A:我去邮局，你（　）? B:我回家。	呢	吗	吧	了
17	４級	表現	我现在（　）游25米。	能	会	没会	应该
18	４級	表現	她昨天（　）迟到了。	再	又	就	才
19	４級	表現	他（　）书包里拿出一本书。	在	从	到	向
20	４級	表現	弟弟（　）毕业了。	快想	快就	快要	就想
21	４級	表現	妈妈（　）我打扫房间。	被	叫	把	使
22	４級	表現	我吃（　）一次中国菜。	过	得	着	在
23	４級	表現	我姐姐（　）会说日语，（　）会说英语。	越～越～	不但～而且～	不但～都～	即使～也～
24	４級	表現	我明天想（　）看电影。	来不	用	去不	去
25	４級	会話	A:你家远吗？B:我家（　）车站很远。	从	到	离	在
26	４級	会話	A:这件T恤衫怎么样？B:我想试（　）。	一点儿	一下	一个	了试
27	４級	会話	A:明天天气怎么样？B:明天（　）下雨。	能	会	可以	别
28	４級	会話	A:请问，苹果（　）卖？B:两块五一斤。	几	多少	什么	怎么
29	４級	会話	A:他会说汉语吗？B:他汉语（　）。	说得流利	说得很流利	说很流利	说流利
30	４級	会話	A:你什么时候来日本的？B:我（　）去年来日本的。	会	是	从	到
31	４級	会話	A:四川菜好吃吗？B:四川菜（　）很辣，（　）很好吃。	既然～可是～	一～就～	或者～或者～	虽然～但是～
32	４級	会話	A:最近你怎么没来学校？B:（　）我回国了，（　）以没来学校。	先～然后～	为了～所以～	又～又～	因为～所以～

こうして 32 問からなる文法能力テストを，1問1点として，満点が 32 点とした（設問文の詳細は補記を参照）。

3. 中国語語彙能力テストの実施対象

本中国語文法能力テストは，日本の大学において中国語を第2外国語（第1外国語は英語）として学んでおり，日本語を母語とする 154 名の大学生を対象に実施した。そのうち，中国語を半年間履修した1年生は 150 名，1年と6ヶ月履修した2年生は4名であった。これらの学習者は，最年長が 35 歳 11 ヶ月，最年少が 18 歳3ヶ月で，平均年齢は 19 歳2ヶ月，標準偏差は 1 歳 6 ヶ月であった。なお，中国語検定のホームページ（https://www.chuken.gr.jp/tcp/grade.html）に掲載された各級の時間数の目安によって，1年生の学習者は準4級，2年生の学習者は4級に相当する。つまり，本研究のテスト実施当日では，調査対象とする学習者の中国語能力の分布は入門から初中級レベルまでであると想定される。

4. 記述統計の結果と信頼度係数の計算

4.1 テストの得点の分布

154 名の学習者による中国語の文法能力テストの総得点の分布は，図1に示した。分布の歪度（skewness）は 0.71 であった。歪度は，分布の非対称性を示す指標で，左右対称でない程度を示す。正規分布であれば，左右対称であるため，歪みはなく0である。歪度の符号によって，正の歪みをもつ分布は得点の高いほうに集まるように歪んでおり，負の歪みをもつ分布は得点の低いほうに歪んでいる。0.71 は，若干得点の高いほうに歪んでいることがわかる。中国語の文法能力テストの総得点の分布の尖度（kurtosis）は 5.66 であった。尖度は，分布の鋭さを表す指標である。正規分布と比べて，尖度が大きければ鋭いピークを持つ分布であり，尖度が小さければより平坦なピークを持つ分布である。尖度が0より大きい場合は，尖りが急で裾が長い分布となり，0より小さい場合は，尖りが丸く緩やかな分布となる。5.66 は0よりもかなり大きいので，尖った分布であることがわかる。図1に示したように，歪度と尖度を合わせて，本研究の中国語の文法能力テストの総得点分布は，ある得点区間に多くの学習者が集まる鋭角の分布であったことがわかる。

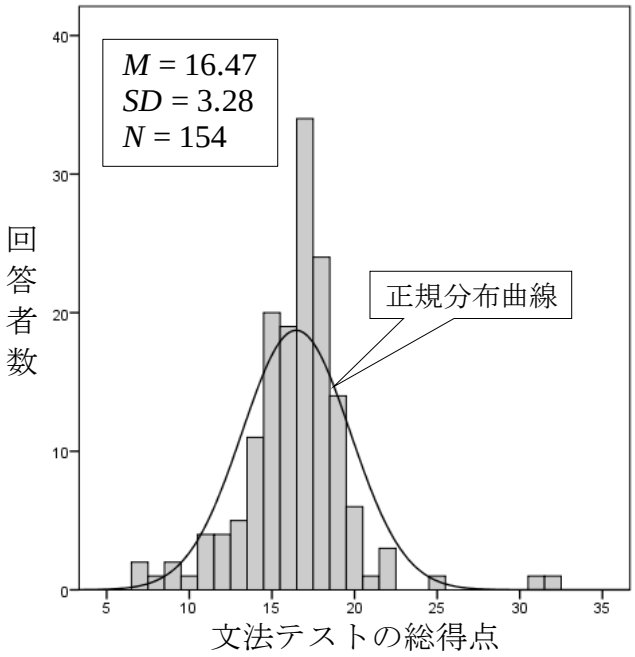

図1　154 名の学習者による文法テストの総得点分布

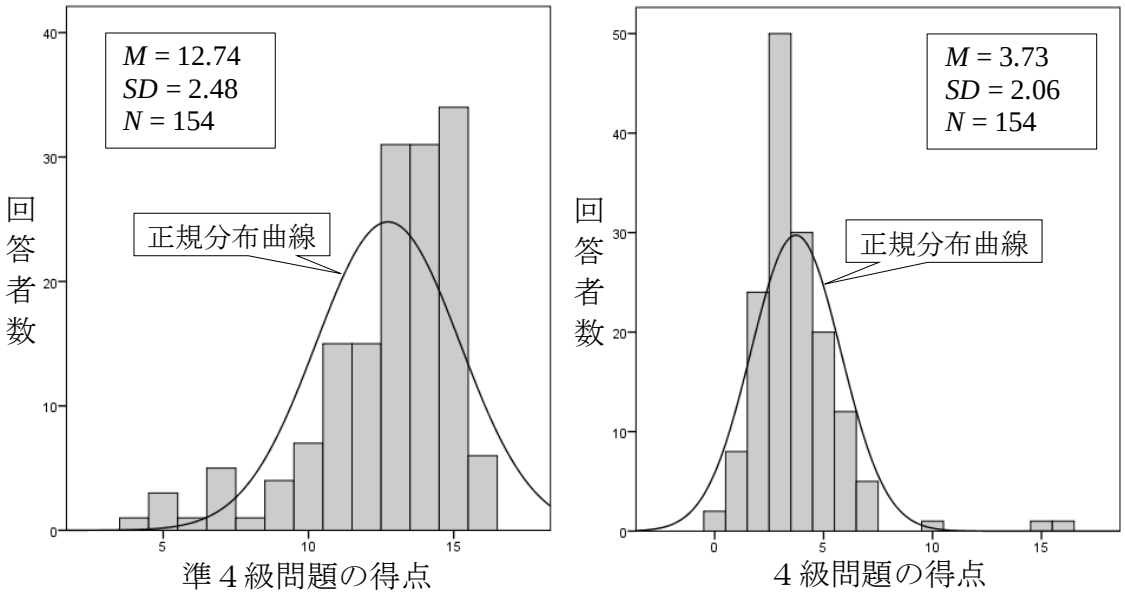

図2　配当級別の得点分布

さらに，配当級別の得点の分布は図2に示した。準4級の得点の分布では，歪度が-1.39で，尖度が 1.91 であった。16 点満点のうち，平均（*M*）が 12.74 点で，標準偏差（*SD*）が 2.48 点であった。一方，4級の得点の分布では，歪度が 2.61 で，尖度が 12.78 であった。4級レ

ベルのテストでは，低い得点領域に多くの中国語学習者が集まっていたことがわかる。16点満点のうち，平均（*M*）が 3.73 点で，標準偏差（*SD*）が 2.06 点であった。準4級は，平均が高く，得点が高いほうに偏っていたことに対して，4級は，平均が低くなり，得点が低いほうに偏っていた。

4.2 記述統計の結果と信頼度係数

32 問からなる文法能力テストにおける下位分類の得点について，配当級別（準4級・4級）で統計的特性を記述的に調べた。その結果は表3にまとめた。日本人中国語学習者154 名を対象に実施した文法能力テストの全体の平均（*M*）は 16.47 点で，標準偏差（*SD*）は 3.28 点であった。最高点は 32 点，最低点は 7 点であった。そのうち，準4級の問題については，16 点満点のうち，平均（*M*）が 12.74 点，標準偏差（*SD*）が 2.48 点，最高点は 16 点であり，最低点は4点であった。また，表現の形式で提示された問題の8点の満点のうち，平均（*M*）が 6.62 点，標準偏差（*SD*）が 1.47 点，最高点は8点で，最低点は2点であった。会話の形式で提示された問題の8点の満点のうち，平均（*M*）が 6.12 点，標準偏差（*SD*）が 1.28 点，最高点は8点であり，最低点は2点であった。

表3 テスト対象の 32 問の配当級および下位分類ごとの記述統計

下位分類	満点	*M*	*SD*	Max	Min	信頼度係数
準４級	16	12.74	2.48	16	4	−
表現	8	6.62	1.47	8	2	−
会話	8	6.12	1.28	8	2	−
４級	16	3.73	2.07	16	0	−
表現	8	1.83	1.28	8	0	−
会話	8	1.90	1.28	8	0	−
文法テスト	32	16.47	3.28	32	7	α=0.66

注:*N*=154, *M* は平均得点, *SD* は標準偏差, Max は最大値, Min は最小値を指す.

一方，準4級の問題の得点と比べて，4級の問題の得点は全体的に低かった。満点 16 点のうち，平均（*M*）が 3.73 点，標準偏差（*SD*）が 2.07 点，最高点は 16 点であり，最低点は0点であった。また，表現の形式で提示された問題の8点の満点のうち，平均（*M*）が 1.83 点，標準偏差（*SD*）が 1.28 点，最高点は8点であり，最低点は0点であった。会話の形式で提示

された問題の8点の満点のうち，平均（*M*）が 1.90 点，標準偏差（*SD*）が 1.28 点，最高点は8点であり，最低点は0点であった。

さらに，文法能力テストのクロンバックの信頼度係数（α）を計算した。クロンバックの信頼度係数とは，複数の質問項目で構成されるテストについて，それらの項目間の内的整合性を調べ，テストの測定精度を表す指標である。信頼度係数は，項目の測定誤差が小さいほど大きくなり，0から1までの連続数値をとる。1に近いほど信頼性が高くなる。154 名による本テスト全体のクロンバックの信頼度係数（α）は 0.66 であった。テスト全体の得点分布が示してくれたように，準4級と4級の得点分布に大きな違いがあるため，配当級別で信頼度係数を計算したほうが合理的であると考えられる。そこで，154 名の学習者による準4級のテスト問題のクロンバックの信頼度係数（α）を計算した結果，0.72 であった。これはかなりよい信頼度係数である。

一方，4級のテスト項目に対しては，得点の低い日本人学習者が多かった。6ヶ月間の中国語学習期間では，4級レベルに達するのは難しかったようである。そこで，4級に達していると想定される学生を選択して信頼度係数を計算することにした。全体の得点の 32 点満点のうち，19 点（満点 32 点×合格率 60%）以上を判断の基準とした結果，わずかに 27 名の学習者になった。27 名による 16 問の4級のテスト項目のクロンバックの信頼度係数（α）を計算し結果，0.68 の信頼度になった。4級の問題である 16 問について，27 名の人数で，0.70 に近い信頼度係数となった。これは，被験者数から考えると，測定内容の一貫性がある程度保たれていることを示しているといえよう。

5. 項目応答理論に基づくテスト評価

項目応答理論（Item Response Theory）は，古典的テスト理論（Classical Test Theory）と異なり，個々のテスト項目に対する反応（回答）を統計的モデルに基づいて分析する。そのため，項目応答理論は学習者がテスト項目の信頼性を改善することに有効的な情報を提供し，学習者依存性・テスト依存性に拘束されず，学習者の能力とテスト項目の難易度の不変性が求められる（大友，1996；大友・中村，2002）。

本研究では，学習者を準4級（*N*＝154）と4級（*N*＝27）に分けて，項目応答理論に基づいた開発した分析ソフト TDAP ver.2（Test Data Analysis Program; 大友・中村 2002）を用いて，各テスト項目の項目難易度，実質選択肢効果（以下，実質選択肢数），項目弁別力およびこ

れらの3指標の標準適切度の合計をテスト評価の有効な指標として算出し，配当級，正答者数，テストの測定内容などと共に表4に報告した。

5.1 各設問の項目困難度

項目困難度（item difficulty: DIFF）は，テスト項目がどれくらい難しいのかを示す指標である。項目困難度は正答者数をテストの参加者である全学習者数で割った数値である。そのため，実質的には正答率と同じで，0から1までの連続した値をとる。項目困難度の数値が0に近いほど，正答率が低くなり，最も困難な項目を意味する。それに対して，項目困難度の数値が1に近ければ近いほど，正答率が高くなり，最も容易な項目と解釈される。大友・中村（2002）は4つの選択肢をもつテストの場合，問題項目として最適な項目困難度は0.625とした。この基準を用いて0.625よりも高ければ高いほど，簡単な問題，逆に0.625よりも低ければ低いほど難しい問題とされる。以下，本テストの項目困難度は配当級別に報告する。

準4級の16問における各項目困難度（N=154）は，表4にまとめたとおりである。0.625よりも値が小さいのは問題17（0.494），問題27（0.234）の2問である。これらの問題は，正解者が少なかったことがわかる。一方，0.625よりも値が極めて大さく1に近いのは，問題3（0.987），問題14（0.935），問題20（0.929），問題30（0.948），問題32（0.935）との5問であった。これらの問題は，ほぼ全員が正解していたことを示しているため，問題があってもあまり意味がないとも考えられる。最適項目困難度で0.625に近いのは，問題22（0.643），問題24（0.812），問題28（0.831）などの9問であった。このように，このテストは，簡単な問題から難しい問題まで難易度の分散がほどよくあり，準4級の中国語の文法を適切に測定できるといえよう。

一方，4級レベルの各項目困難度（N=27）も，表4にまとめたとおりである。0.625よりも値が極めて小さかったのは，問題1（0.481），問題2（0.222），問題6（0.222），問題7（0.333），問題8（0.222），問題10（0.074），問題11（0.148），問題13（0.333），問題16（0.407），問題18（0.407），問題21（0.444），問題25（0.259）の12問があり，最適項目困難度の0.625に近いのは，問題12（0.704），問題15（0.593），問題23（0.741），問題31（0.593）の4問であった。つまり，このテストの4級レベルの16問は学習者にとって，問題の難易度の配置がかなり高く，中国語検定の4級レベルに達したかどうかを判断するために適切に測定できると考えられる。

表4 配当級別の正答者数, 項目困難度, 実質選択肢数, 項目弁別力および適切度

ID	問題番号	配当級	設問形式	測定内容	*N*	正答者数	DIFF	AENO	DISC	SATOT
1	4	準4級	表現	品詞の理解	154	134	0.870	1.664	0.624	1.759
2	9	準4級	表現	品詞の理解	154	134	0.870	1.630	0.491	1.531
3	19	準4級	表現	品詞の理解	154	104	0.675	2.571	0.606	1.900
4	24	準4級	表現	品詞の理解	154	125	0.812	1.808	0.588	1.600
5	3	準4級	表現	文型の理解	154	152	0.987	1.081	0.313	1.335
6	20	準4級	表現	文型の理解	154	143	0.929	1.389	0.502	1.561
7	22	準4級	表現	文型の理解	154	99	0.643	2.348	0.311	1.449
8	28	準4級	表現	文型の理解	154	128	0.831	1.772	0.360	1.430
9	5	準4級	会話	品詞の理解	154	138	0.896	1.546	0.447	1.538
10	14	準4級	会話	品詞の理解	154	144	0.935	1.299	0.335	1.264
11	17	準4級	会話	品詞の理解	154	76	0.494	3.091	0.514	1.594
12	32	準4級	会話	品詞の理解	154	144	0.935	1.353	0.420	1.477
13	26	準4級	会話	文型の理解	154	130	0.844	1.541	0.411	1.212
14	27	準4級	会話	文型の理解	154	36	0.234	3.505	0.157	1.149
15	29	準4級	会話	文型の理解	154	129	0.838	1.747	0.595	1.645
16	30	準4級	会話	文型の理解	154	146	0.948	1.285	0.532	1.553
17	11	4級	表現	品詞の理解	27	4	0.148	2.271	0.606	1.305
18	13	4級	表現	品詞の理解	27	9	0.333	3.746	0.300	1.528
19	18	4級	表現	品詞の理解	27	11	0.407	3.446	0.380	1.571
20	23	4級	表現	品詞の理解	27	20	0.741	2.179	0.208	1.583
21	1	4級	表現	文型の理解	27	13	0.481	2.726	0.398	1.508
22	6	4級	表現	文型の理解	27	6	0.222	2.480	0.657	1.449
23	25	4級	表現	文型の理解	27	7	0.259	3.210	0.532	1.476
24	31	4級	表現	文型の理解	27	16	0.593	2.791	0.356	1.670
25	2	4級	会話	品詞の理解	27	6	0.222	3.833	0.717	1.795
26	7	4級	会話	品詞の理解	27	9	0.333	2.988	0.459	1.443
27	8	4級	会話	品詞の理解	27	6	0.222	2.948	0.597	1.462
28	10	4級	会話	品詞の理解	27	2	0.074	2.739	0.459	1.271
29	12	4級	会話	文型の理解	27	19	0.704	2.169	0.286	1.560
30	15	4級	会話	文型の理解	27	16	0.593	2.998	0.306	1.711
31	16	4級	会話	文型の理解	27	11	0.407	2.258	0.151	1.274
32	21	4級	会話	文型の理解	27	12	0.444	2.277	0.446	1.395

注: DIFF=項目困難度, AENO=実質選択肢数, DISC =項目弁別力, SATOT=適切度.

5.2 各設問の実質選択肢数

実質選択肢数（actual equivalent number of options: AENO）は多肢選択形式（multiple choice format）の問題項目で, 正答以外の魅力的である錯乱肢がどれほど実質的に機能したかを示す指標であり, 1から選択肢の数までの連続的な数値をとる（大友・中村, 2002）。本研究は張ほか（2017）と同様に, 四者択一の問題形式であるため, 実質選択肢数の数値

が1から4までの連続的な値をとる。実質選択肢数が1であれば，全員が正答したことになり，錯乱肢の機能が働いていないことを意味する。4であれば，1つの問題に対して設けた正答を含む4つの選択肢がランダムに選択されたことを示す。したがって，実質選択肢数はどれくらいの錯乱効果があるのかを数値化してくれ，選択肢の分布状況または学習者による選択状況をわかりやすく反映する。正答率が低くなると，項目困難度も0に近づき，正答以外の選択肢を選ぶ確率も高くなるので，実質選択肢数も高くなる傾向がある。そこで，本研究では，準4級および4級の項目の項目困難度と実質選択肢数のピアソンの積率相関係数を計算した。その結果，準4級は問題項目数が 16 問で，r=-0.98 であった。それに対して，4級は問題項目数が 16 問で，r=-0.33 であった。いずれも逆相関を示した。これは，項目困難度は難しいほど数値が小さくなり，実質選択肢数は四者択一で最低の1から最高の4まで変化するので，逆相関を示すと考えられる。本テストで設けた 32 問の実質選択肢数を配当級別で報告する。

表4に示したとおり，準4級の 16 問のうち（N=154），実質選択肢数が最も高かったのは問題 27（3.505）で，最も低かったのは問題3（1.081）であった。16 問の実質選択肢数の平均は 1.852 であった。学習者全員に対して，この 16 問の錯乱肢の魅力をほどよく発揮したことがわかる。それに対して，4級の 16 問のうち（N=27），実質選択肢数が最も高かったのは問題2（3.833）で，最も低かったのは問題 12（2.169）であった。16 問の実質選択肢数の平均は 2.816 であった。32 問のテストに合格した 27 名の学習者に対して，この 16 問の選択状況には錯乱肢効果がみられ，正答以外の錯乱肢がよく機能していたことがわかる。

5.3 各設問の項目弁別力

項目弁別力（item discrimination power: DISC）は，ある項目が能力の高い学習者と能力の低い学習者をテストでその程度弁別できるかを示す指標であり，-1から1までの連続数値をとる（大友・中村，2002）。Ebel（1979）と大友・中村（2002）は，項目弁別力が 0.30 以下なら，能力を弁別する項目として不適切であり，改定または除外されるべきであることを提案している。表4に示したように，準4級の 16 問のうち（N=154），0.30 を満たしていないのは問題 27（0.157）のみであった。また，4級の 16 問のうち（N=27），0.30 を満たしていないのは問題 12（0.286），問題 16（0.151）と問題 23（0.208）の3問であった。これらの最適とはいえない問題が4問含まれているが，全体としては，今回の中国語の文法能力テストは学

習者の文法能力をじゅうぶんに弁別できると評価されよう。

5.4 項目分析結果の総合評価

どのテスト項目が最も有効であるかを示す指標として，項目困難度，実質選択肢数，項目弁別という3つの指標による標準適切度の合計（standard appropriateness total：SATOT）を計算し，表4に報告した。SATOT 値が大きければ，大きいほど優れていると解釈される。表3に示したように，準4級の問題 19（1.900）の SATOT 値が最も大きかったため，最も良好な項目となった。それに対し，準4級の問題 27（1.149）の SATOT 値が最も小さかったため，最も不良な項目とされる。学習者（*N*＝154）による回答からみると，問題 27 が測定する選択疑問文の文法項目に対して，全員がほぼ答えられず，床効果（floor effect; フロア効果ともいわれる）があった。

6. おわりに

本研究は，初中級レベルの中国語文法能力テストを開発し，その評価を報告した。中国語検定の準4級および4級に相当する日本人中国語学習者に対しテストを実施，回答に基づいて，テストの信頼性，有効性および実用性を報告した。結果と今後の検討は，以下のようにまとめられよう。

第1に，今回のテストの文法項目は，中国語検定の準4級と4級の文法項目の出題範囲に準じて，設問形式および測定内容を厳密に統制した。入門および初中級レベルに相当する日本人中国語学習者 154 名に実施した結果，準4級および4級のテスト項目の信頼度係数は 0.65 以上となり，高い値であった。本テストは，テスト問題全体として測定の精度を保っており，信頼性があると評価できよう。

第2に，項目応答理論に基づいて開発された分析ソフトである TDAP ver.2 を使って，項目困難度，実質選択肢数，項目弁別力とそれらの3つの指標による標準適切度の合計という4つの指標から本テストの問題の適切性および特徴を総合的に考察した。まず，今回実施した 32 問のうち，両配当級併せて最適項目困難度の 0.625 に近かったのは 13 問（40.63%）であった。0.625 よりも大きかったのは5問（15.63%）であり，0.625 より小さかったのは 14 問（43.75%）であった。つまり，本テストの問題は，難易度に適度な分散があることがわかる。本テストは，学習者の中国語の文法能力として，ほぼよい難易度であると考えられ，

文法能力をかなり効果的に測定しているといえよう。しかし，総合評価の指標としての標準適切度からでは，選択疑問文を測定する問題 27 に床効果がみられ，項目困難度が 0.234 であり，実施選択肢数が 3.505 となり，正答を含む4つの選択肢をほぼランダムに選んだことがわかる。この設問は会話形式で提示しているため，学習者の読解能力がある程度干渉したと思われる。これについては，今後，より適切な項目と差し替え，テストを改定する必要があろう。

最後に，Ebel（1979）と大友・中村（2002）が提案した弁別力の良好な問題の基準によって，本テストにおいては，項目弁別力が 0.30 を超えており，弁別力の良い項目が 28 問（全体項目の 87.5%）もあった。したがって，本文法テストは，弁別力が大きく，学習者による文法能力がどのレベルに達したかを効率的に判断できる。本文法能力テストは今後，初中級レベルまでの中国語教育において活用できるように，オンラインで自由にアクセスして受験できるように，無料でウエブサイトに公開した（https://forms.office.com/r/iLkcyeBFqF）。

付記

本研究は，科学研究費補助金・若手研究「日本人学習者の中国語の声調および韻律の理解を促進する背景諸要因」（課題番号：21K13055，研究代表者，張婧禕）の助成を受けた研究成果の一部である。また本研究は，科学研究費補助金・基盤研究（S）「OS 言語からみた「言語の語順」と「思考の順序」に関するフィールド言語心理学的研究」（研究課題番号：19H05589，研究代表者，東北大学・教授 小泉政利）の助成を受けた。

［参考文献］

大友賢二（1996）『項目応答理論入門: 言語テスト・データの新しい分析法』，大修館書店.

大友賢二（監修）・中村洋一（著）（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―』，桐原書店.

オルダーソン，チャールズ・クラッファム，キャロライン・ウォール，ダイアン（著）・渡辺良典（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.

張婧禕・玉岡賀津雄・勝川裕子（2017）「中国語語彙能力テストの開発―HSK3級の日本人中国語学習者のデータに基づいた評価―」『ことばの科学』31，21-37.

ブラウン，J. D.（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.

Ebel, R. L. (1979) *Essentials of Educational Measurement* (3rd ed.). Englewood Cliffs, Prentice Hall.

[Web 参考サイト]

中国語検定各級の目安（時間数）一覧表：
https://www.chuken.gr.jp/tcp/grade.html
中国語文法能力テストサイト：
https://forms.office.com/r/iLkcyeBFqF

……

張 婧禕
（宮崎大学多言語多文化教育研究センター・講師）
Email: jingyizhang@cc.miyazaki-u.ac.jp
玉岡 賀津雄
（湖南大学外国語学院・教授，名古屋大学大学院人文学研究科・名誉教授）
Email: ktamaoka@gc4.so-net.ne.jp

［補記］

中国語文法能力テスト

文の意味が成り立つため，最も適切なものを４つの選択肢から１つ選んでください。

1 （　） 我姐姐（　　）会说日语，（　　）会说英语。

A 越～越～　B 不但～而且～　C 不但～都～　D 即使～也～

2 （　） A：这件 T 恤衫怎么样？

B：我想试（　　）。

A 一点儿　B 一下　C 一个　D 了试

3 （　） 她（　　）我高。

A 比　B 跟　C 和　D 给

4 （　） 他们不（　　）是日本人。

A 也　B 都　C 多　D 太

5 （　） A：最近怎么样?

B：最近我（　　）忙。

A 有点儿　B 一点儿　C 一会儿　D 点儿

6 （　） 我吃（　　）一次中国菜。

A 过　B 得　C 着　D 在

7 （　） A：明天天气怎么样？

B：明天（　　）下雨。

A 能　B 会　C 可以　D 别

8 （　） A：你家远吗？

B：我家（　　）车站很远。

A 从　B 到　C 离　D 在

9 （　） 你（　　）什么名字?

A 叫　　B 是　　C 姓　　D 贵姓

10 （　） A：请问，苹果（　　）卖?

B：两块五一斤。

A 几　　B 多少　　C 什么　　D 怎么

11 （　） 他（　　）书包里拿出一本书。

A 在　　B 从　　C 到　　D 向

12 （　） A：最近你怎么没来学校?

B:（　　）我回国了，（　　）没来学校。

A 先～然后～　　B 为了～所以～　　C 又～又～　　D 因为～所以～

13 （　） 弟弟（　　）毕业了。

A 快想　　B 快就　　C 快要　　D 就想

14 （　） A：你家有几（　　）人?

B：我家有四（　　）人

A 口　　B 个　　C 位　　D 只

15 （　） A：四川菜好吃吗?

B：四川菜（　　）很辣，（　　）很好吃。

A 既然～可是～　　B 一～就～　　C 或者～或者～　　D 虽然～但是～

16 （　） A：他会说汉语吗?

B：他汉语（　　）。

A 说得流利　　B 说得很流利　　C 说很流利　　D 说流利

17 （　） A：那是什么?

B：这是一（　　）毛衣。

A 张　　B 台　　C 件　　D 条

18 （　） 她昨天（　　）迟到了。

A 再　　B 又　　C 就　　D 才

19 （　） 今天作业（　　）多了。

A 很　　B 不很　　C 非常　　D 太

20 （　） 我妹妹（　　）两把伞。

A 有　　B 不有　　C 在　　D 不在

21 （　） A：你什么时候来日本的?

B：我（　）去年来日本的。

A 会　　B 是　　C 从　　D 到

22 （　） 学校旁边（　　）公园?

A 在不在　　B 有没有　　C 没有没　　D 有不有

23 （　） 我现在（　　）游 25 米。

A 能　　B 会　　C 没会　　D 应该

24 （　） 今天（　　）冷。

A 一点儿　　B ／　　C 没　　D 很

25 （　） 妈妈（　　）我打扫房间。

A 被　　B 叫　　C 把　　D 使

26 （　） A：那本词典（　）哪里?

B：那本词典（　）桌子上。

A 有　　B 在　　C 没在　　D 没有

27 （　） A：你喝咖啡（　　）喝红茶?

B：我喝红茶。

A 还是　　B 是不是　　C 还不　　D 喝不

28 （　） 冰箱（　　）有一杯牛奶。

A 面　　B ／　　C 的里　　D 里

29 （　） A：我去邮局，你（　　）？

B：我回家。

A 呢　　B 吗　　C 吧　　D 了

30 （　） A：你们的作业（　　）？

B：我们的作业不多。

A 多不多　　B 不多不　　C 多没多　　D 没多没

31 （　） 我明天想（　　）看电影。

A 来不　　B 用　　C 去不　　D 去

32 （　） A：今天星期（　）？

B：今天星期三。

A 几　　B 多长　　C 多少　　D 多久