

論 文

日本語学習者の研究論文の Analytic Scoring 評価基準の 妥当性と信頼性に関する一考察

高 帥

長春大学助理研究員・広島大学大学院文学研究科博士課程後期

Examination of the Validity and Reliability of the Analytic Scoring Evaluation
Criteria for Japanese Learners' Research Papers

GAO Shuai

Abstract: This study discusses the validity and reliability of the evaluation criteria conducted as part of preparing the analytic scoring evaluation criteria for research papers on learners of Japanese as a second language. With the publication of the “National Standard for Education of Undergraduate Specialty in Colleges and Universities” (a new national standard) and “Bachelor thesis (design) lottery inspection method (trial)” by the Chinese Ministry of Education, the importance of writing has been recognized more than before. However, there are still no writing evaluation standards that can be shared in Japanese language education. Using the Delphi method, the author conducted a research study, on a collection of evaluation items constructed based on the abilities required by the new national standard. Subsequently, the author created a draft of the analytic scoring evaluation criteria for Japanese learners' research papers based on the survey results. This study comprises conducting the research, using the draft, examining the validity and reliability of the draft for the evaluation criteria, and revising the draft for the evaluation criteria based on the results. The revised draft of the evaluation criteria was named the “Japanese research paper evaluation criteria”. From the results of the factor analysis and Cronbach's alpha, “Japanese research research treatise evaluation criteria” was proven to be a high reliability and validity scale.

Keywords: evaluation criteria, construct validity, reliability, analytic scoring, research papers of Japanese learners

1. はじめに

2018年1月、中国教育部は「四年制大学各専攻教育の国家スタンダード」(普通高等学校本科專業類教學質量國家標準)(以下、「新国標」)を公表した。そこには優秀な人材について新たな定義がなされ、人材を育成するために到達すべき目標が掲げられている。論文は、求められる能力の習得状況を検証する重要な手段であり、年一回各省レベルの教育行政部門が「新国標」の内容を参照し、専門別に抽選検査をすることが要求されている¹。

中国には、論文評価基準に関わる研究として、研究論文の作成に関する指導を目標としたものがあるが、それらは中国語を対象としており、外国語で書かれた論文を如何に評価するかについての研究は極めて少ない。

中国における第二言語(以下、L2)としての日本語研究論文²評価研究の問題点には、以下の3つが挙げられる。

(1)「新国標」の要求に適応する「卒業論文(設計)抽選検査の方法(試行)」(本科畢業論文(設計)抽檢辦法(試行))により、論文審査の要点である研究意義、論理構築、専門能力及び学術規範を評価指標にする論文作成指導向けの研究論文評価基準がない。

(2)多くの論文評価基準はHS評価であり、評価基準や観点を詳細に示すAS評価の研究は少ない。評価の研究では、評価のプロセスや留意点に関する研究が多く、具体的な評価基準が示されている研究はほとんどない³。

(3)試験や課題における作文評価の研究⁴はあるが、日本語の研究論文に特化した評価研究は極めて少ない。

以上の問題点を受け、筆者は2020年に中国における「新国標」に示されている人材育成の求められている能力に適応する、L2としての日本語研究論文評価基準を開発し始めた。L2としての日本語教育において、論文の評価基準が明確になると、日本語研究論文作成にも、日本語研究論文作成の指導にも、また、社会に出てからの文書作成にも役立つと考えられる。

本研究は、高帥(投稿中)の日本語学習者の研究論文のAnalytic scoring評価基準原案の妥当性と信頼性を検討し、最終的な評価基準原案を作成することを目的とする。本研究は、日本語学習者の研究論文に特化した妥当性と信頼性を備えた評価基準の完成を目指す。

2. 評価基準原案の信頼性の検討と探索的因子分析

2.1 調査概要

(1) 調査アンケートの概要

中国語母語の日本語教師（以下、中国人日本語教師）を対象にアンケート調査を行った。本研究および本調査の目的を簡単に説明し、アンケート（評価基準原案）を配布した（参考資料1）。中国人日本語教師には、日本語を学んでいる中国人大学生の研究論文の評価を依頼し、同時に評価に用いた「評価基準原案」の各評価小項目の必要度を5段階で採点するように依頼した。期間は14日間とした。アンケート調査に答えた137名の日本語教師のうち、回答に全部4点を選択した人が1名、全部5点を選択した人が8名で、真面目に回答されていないと筆者が判断した9名のデータを削除し、128名を分析対象とした。

(2) 被調査者の概要

中国人日本語教師 128名 男性：29名 女性：99名

年齢 (20代) 4名 (30代) 20名 (40代) 68名 (50代以上) 36名

日本語教育歴 (5年以下) 12名 (5-10年) 8名 (10年以上) 108名

論文指導歴 (経験なし) 17名 (5年以下) 11名 (5-10年) 5名

(10年以上) 95名

日本人日本語教師に論文を指導された経歴 (有り) 21名 (無し) 107名

2.2 調査結果

(1) 信頼性

測定の妥当性の条件としてしばしば挙げられるものに「信頼性」があり、「時間をおいてくりかえし測定してもデータが安定していること」や、「同じ人物についての評定値は、評価者が異なってもよく一致すること」といった結果の一貫性に関わるものがある。このような一貫性は、一般に測定の信頼性または尺度の信頼性とよばれている⁵。本研究は、態度・意見式テストの信頼性分析に用いられ、論文の評価に適している⁶クロンバックの α 係数による統計法を用いて、信頼性を測定した。

クロンバックの α 係数とは、項目群の可能な組み合わせ全てについて信頼性係数を算出し、平均した結果であり、 α 係数の値で信頼性を評価する。0から1までの値をとり、1に近いほど信頼性が高い。信頼性が高いと見るに

は少なくとも0.7以上は必要とされている。

アンケート調査のデータを入力し、SPSS で α 係数を計算した結果を表1にまとめた。この結果により、 α 係数は0.878であり（網掛け太字参照）、統計上、評価基準原案の信頼性は高いことが、また、項目が削除された場合のクロンバックの α 係数の結果では、25のいずれの項目も α 係数が0.85以上であり、評価項目として削除する必要がないことが示された⁷。以上、本調査で用いた「評価基準原案」全体の α 係数の値（0.878）により、信頼性が高いことが明らかになった。

(2) 探索的因子分析

探索的因子分析の目的は複数の質問項目において、共通に関連している因子を見つけることである⁸。つまり、より一般化され多くの観測変数に共通して影響を及ぼす因子を見つける為に変数をまとめ、因子構造の確認を行うのである。

本節では、本研究にとって、より妥当性の高い評価基準を作成するために、「評価基準原案」アンケートの回答結果を分析し、25項目（観測変数）を数少ない変数（多くの観測変数に共通して影響を及ぼす因子、潜在的な変数が共通因子とも呼ばれる）にまとめる探索的因子分析について述べる。

25項目の因子分析を行う前に、まず、アンケート調査のデータは因子分析を用いるのに適切かどうかを検定すべきである。本研究は「KMO および Bartlett の球面性検定」を用いて検定する。データが因子分析を用いるのに適切であるか（データに意味のある因子が発見できそうであるか）を判断するための基準として Kaiser-Meyer-Olkin という標本妥当性の測度（KMO 測度）がある。サンプルサイズに関しては絶対的な基準はない。多くの研究論文では100程度で因子分析を行っている場合が多いが、この程度で妥当なのか理論的根拠が必要である。KMO の標本妥当性の測度で検証することが推奨された⁹。また、Bartlett の球面性検定は、「観測変数間に無相関である」という帰無仮説¹⁰を検定する。簡単に言えば、「KMO および Bartlett の球面性検定」は変数間に共通因子があるかどうかを検定できる¹¹。KMO は0から1までの値をとり、1に近いほど共通因子の数が多い、かつ相関係数が適切に算出されたことを意味し、信頼できる因子の抽出及びサンプルの妥当性を示す¹²。0.5以下の時、因子分析を適用しない。少なくとも0.6以上であれば、因子分析を用いるのに適切である¹³。また、Bartlett の球面性検定では有意確率が0.05

より小さいのも必要である。

SPSS を利用し、「分析」→「次元分解」→「因子分析」→「記述統計」で「KMO および Bartlett の球面性検定」を行い、KMO の標本妥当性の測度は 0.823 (>0.6)、Bartlett の球面性検定の有意確率は 0.000 (<0.001) であった¹⁴ (参考資料 2)。この結果から、25項目は、因子分析を適用するのは妥当であるといえよう。

表 1 信頼性統計量の結果
信頼性統計量

Cronbach のアルファ	.878
項目の数	25
	項目が削除された 場合の Cronbach の アルファ
文体	.870
明快さ	.868
文型	.884
用言	.878
テンス	.880
語彙	.879
表記	.873
先行研究	.874
研究目的	.873
研究テーマ	.872
要旨	.872
研究対象	.869
信頼性	.872
研究結果	.873
独創性	.868
展望性	.870
一貫性	.875
研究方法	.875
説得力	.877
読み手	.874
バランス	.869
段落	.869
文	.870
メタ言語	.872
書式	.869

次に、因子構造を確認する前に、因子抽出を行う。因子抽出法は「主成分分析」、表示は「スクリープロット¹⁵」と因子抽出基準は「固有値の下限：1」¹⁶を選択する。

出力の結果は以下のように示されている。①25項目の中に共通性¹⁷の最小値は0.483 (文体)、全ての項目の共通性は0.2以上なので、全項目を保留にする。②固有値は1以上の因子数を6つ抽出し、因子のスクリープロットの結果¹⁸も参考にし、6つの因子解が適当だと判断した (参考資料 2)。

最後に、因子構造の確認を行う。ここでは、因子の解釈を容易にするためにバリマックス回転 (varimax rotation) を用いて因子構造を分析した。この方法により、因子負荷プロットを「サイズによる並び替え」で表示し、因子構造が確認した (表 2)。

第1因子は、「明快さ」「書式」「バランス」「段落」「文」「メタ言語」「文体」「研究対象」という文章を作成する際に必要な能力であり、「文章力」と名付けた。第2因子は「研究目的」「先行研究」「研究テーマ」「信頼性」「要旨」という他論文に対する批判や補充、また作者自身の考えが必要な能力であり、「思考力」と名付けた。第3因子は「一貫性」「読み手」「説得力」「研究方法」という自分の主張

をアピールする時必要な能力であり、「論証力」と名付けた。第4因子は「研究結果」「独創性」「展望性」という創造力を要求する項目であり、「革新力」と名付けた。第5因子と第6因子はそれぞれ「文型」「用言」「テンス」と「表記」「語彙」という外国語応用能力を考察する項目であり、「言語力1」と「言語力2」と名付けた。

表2を見ると、第6因子の項目数が2つしかなく、1因子あたりに適切な項目数は一般的に3以上である¹⁹ため、第5因子と第6因子を1つにし、「言語力」と名付けた。本研究の初期解では固有値1以上の因子は6つ抽出されたが、因子の解釈の観点から5因子解を最適解として採用した（累積寄与率²⁰は全分散の61.965%）。各因子に含まれる項目の因子負荷量が0.4以上のものである。いずれの項目も評価項目として削除する必要がないことがわかった。

3. 評価基準原案の検証的因子分析

3.1 検証的因子分析

因子分析は、探索的因子分析と検証的因子分析に大別される。前者は因子に関する明確な仮説がなく、因子を探るために使われるのに対して、後者は仮説が事前にあるときに、その過程を検証するために利用される。

探索的因子分析の結果を踏まえ、評価基準原案の構成概念妥当性を検証するために、ソフトAmosを利用した。2つのモデル（パス図）を設定した（参考資料3）。モデル1は全ての項目が日本語研究論文評価に負荷することを仮定した1因子モデルである。モデル2は探索的因子分析で抽出された5因子間に相関関係があることを仮定した5因子斜交モデルである。なお、実際のデータに対して、これらのモデルがどの程度適合しているかを確認するために、検証的因子分析²¹を行った。アンケート調査のデータを2つのモデルに代入し、パス係数²²と適合度を計算した。

適合度指標はモデル全体の適合度の度合いを以下の4指標を用いて検定した。① GFI (goodness-of-fit index): 推定された共分散行列と標本共分散行列の関連から規定される適合度指標、② AGFI (adjusted goodness-of-fit index): 調整済み適合度指標、③ CFI (comparative fit index): 妥当性を検証しているモデルと独立モデルの適合度を比較する適合度指標、④ RMSEA (root mean square error approximation): 近似誤差2乗平均平方根であり、GFIではパラメータが多くなるのに伴う点を補正した適合度指標（呉明隆、涂金堂、2012²³;

王培剛、2020²⁴)。

一般的には、適合度指標により、モデルの適合についての判断基準は① GFI と② AGFI がそのモデルが適合していると判断される一応の目安は0.9、③

表2 日本語研究論文評価基準原案の探索的因子分析

	因子負荷量					
	1	2	3	4	5	6
第1因子「文章力」						
明快さ	.767	.200	.181	.038	.059	-.171
書式	.759	.127	.088	.206	-.049	-.002
バランス	.749	.055	.176	.082	.128	.009
段落	.718	.142	.093	.133	.071	.039
文	.634	.011	.069	.310	.175	.305
メタ言語	.611	.376	-.018	.080	-.253	.035
文体	.550	.237	.158	.132	.126	.258
研究対象	.502	.358	.151	.296	-.090	-.184
第2因子「思考力」						
研究目的	.187	.725	.088	.101	.062	-.069
研究テーマ	.377	.653	-.111	.098	.018	.124
先行研究	-.010	.633	.121	.447	.022	-.079
信頼性	.446	.618	.092	-.123	-.004	.187
要旨	.141	.536	.127	.470	-.019	.182
第3因子「論証力」						
一貫性	.114	.232	.744	-.129	.068	.055
読み手	.236	.050	.707	.097	.008	-.157
説得力	.062	-.014	.706	.121	.153	.229
研究方法	.115	-.056	.627	.318	-.050	.021
第4因子「革新力」						
研究結果	.155	.161	.167	.755	-.044	-.003
独創性	.476	.059	.144	.632	.091	.076
展望性	.442	.197	-.032	.599	.045	.153
第5因子「言語力1」						
文型	.044	-.162	-.145	.154	.781	-.021
用言	.058	.174	.065	-.050	.746	.156
テンス	.041	-.004	.238	-.066	.725	-.230
第6因子「言語力2」						
表記	.188	.276	.344	.133	.038	.652
語彙	.088	.158	.146	-.039	.524	-.640

CFIの目安は0.9もしくは0.95とされている。また、④ RMSEAについては0.08以下であれば適合しており、あるいは0.05以下であれば良好、0.1以上であればNGといった判断基準が示されている（豊田、2007²⁵；呉明隆、涂金堂、2012；王培剛、2020）。

適合度の検討を行った結果、モデル1はモデル2と比べ、適合度が低いことが明らかになり、モデル2の方を採用して修正する（表3）こととした。モデル2のパス係数を計算すると、言語力と他の因子間の係数が小さくて（<0.4）、関連が薄いとわかった。

そこで、モデル3（参考資料3）を作成したが、計算の結果はGFI = 0.794、AGFI = 0.752、CFI = 0.826、RMSEA = 0.074であり、モデル2とほぼ変わらなかった。ゆえに、外国語語学力を考察する項目とする言語力と研究論文作成能力を考察する能力（ここでは研究力と名付けた）を分けて、モデルを設定した。また、項目「表記」と言語力のパス係数は-0.1から、0.4より遠く離れている。その観測変数は矢印を引いた因子に属しないと判断し、矢印を引き直し共分散構造分析をやり直すべきであ

表3 モデル1とモデル2の適合度指標

	GFI	AGFI	CFI	RMSEA
モデル1	.704	.650	.654	.104
モデル2	.796	.750	.825	.075

表4 言語力・斜交・直交モデルの適合度指標

	GFI	AGFI	CFI	RMSEA
言語力モデル	.999	.994	1.000	.000
斜交モデル	.938	.908	1.000	.000
直交モデル	.934	.904	1.000	.000

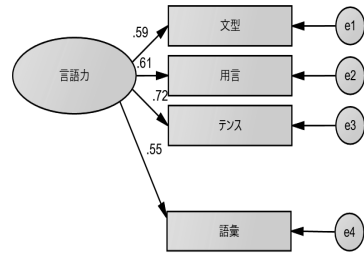


図1 言語力モデル

る。つまり、「表記」は他の因子に属すかもしれない。仮に他の因子に属すとして計算してみると、文章力、思考力、論証力と革新力に属す時のパス係数はそれぞれ0.41、0.47、0.57、0.45である。従って、「表記」を因子「論証力」に入れた。この時、設定したモデルは「言語力モデル」（図1）と「モデル4」（参考資料3）である。言語力モデルの因子と項目間のパス係数は全部0.4以上で、高い適合度が得られた（表4）。

モデル4の因子と項目間のパス係数も全て0.4以上である。しかし、モデル2より適合度が高いとは言え、高い適合度とは言えない（モデル4適合度の計算結果はGFI = 0.828、AGFI = 0.783、CFI = 0.876、RMSEA = 0.071である）。Amosが提示されたモデル適合度の向上につながりそうな修正候補を指数として示すModification indices (MI修正指数)で修正する必要がある。MI値が一番大きな項目から、修正し、高い適合度を得るまでモデル4を調整した。最後に得たモデルは図2に示し「斜交モデル」と名付けた。適合度の結果を表4に記入した。

また、探索的因子分析で抽出された5因子間に相関関係がないことを仮定し、5因子直交モデルを設定した。同じ方法で、「言語力モデル」と他の因子を分け、MI修正を行い、高い適合度得るまでモデルを調整した。得たモデル（直交モデル）は図3のように示されている。適合度の結果を表4に記入した。

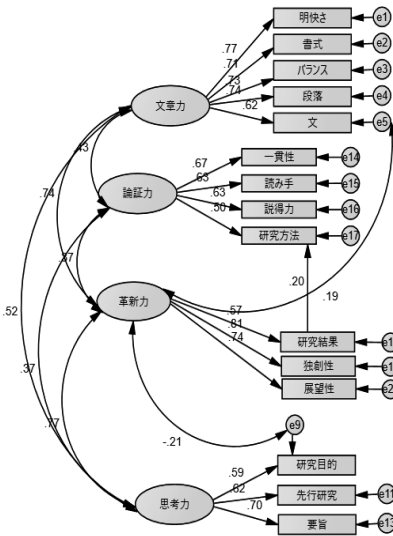


図2 斜交モデル

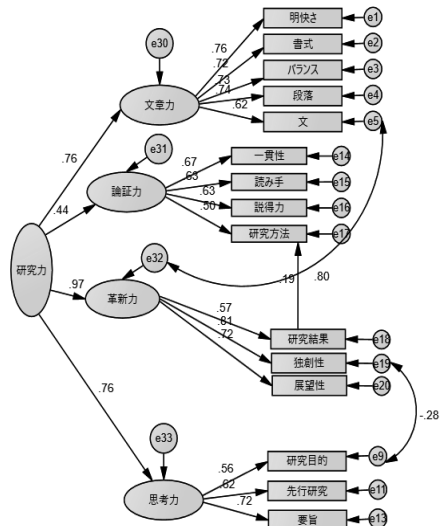


図3 直交モデル

3.2 構成概念妥当性と評価基準の構築

検証的因子分析の結果により、斜交モデルも直交モデルも高い適合度があることが分かった。斜交モデルの適合度は相対的に適合度の高いモデルであるから、日本語研究論文評価基準については、言語モデルと斜交モデルを合わせるモデルが最も適合するモデルと判断した。

さらに、図2に示す検証的因子分析の結果を参考に、モデルの適合度を高めるために、第1因子「文章力」の「メタ言語」「文体」「研究対象」項目と第2因子「思考力」の「信頼性」「研究テーマ」項目を削除した。また、因子「言語力」から第3因子「論証力」に移した「表記」項目も削除した。「メタ言語」は「読み手」、「文体」は「文型 / 表現」、「研究対象」と「信頼性」は「説得力」、「研究テーマ」は「研究結果」及び「表記」は「語彙」により評価できる。これまでの作業により、残った計19項目で構成した言語力モデルと斜交モデルの構成概念の妥当性は、検証できた。

19項目全体のクロンバックの α 係数は0.834であり、信頼性が高いと考えられる。

以上、計19項目の日本語研究論文評価基準の評価項目を決定した。評価基

準原案に基づき、「評価大項目」「段位」「得点」と「成績」の欄を追加した。「評価大項目」は因子分析の結果により、「言語力」「思考力」「革新力」「論証力」と「文章力」を分けた。各項目のレベルは上位（既定要求以上に満足できるレベル）、中位（既定要求に満足できるレベル）、下位（既定要求に満足できないレベル）と無段位（完全に既定要求に達していないレベル）を設置した。19項目の各項目について3点満点で、評価基準の満点を57点にした。社会の変化により、学校により、目的により、評価項目に対するポイントが異なる。従って、本研究の評価基準では各項目の重みづけは同じ値に設定した。即ち、1/19、約5.26%である。最後の成績は実際の得点から100点満点の最終点数に変換できる。優、良、中、合、不合の5つのレベルで成績を表示することができ、個人差を区別するために、点数で成績を表示することもできる。エクセルで作成しているため、各項目の段位を入力すれば、実際得点と最終得点は自動的に計算が可能である。どの項目の能力が低いのか、一目瞭然であり、論文修正にも論文指導にも役立てられる。これまでの作業を踏まえ、「新国標」に適応し、かつ妥当性と信頼性を備えた評価基準を表5に示す。

表5 日本語研究論文評価基準

能力別	評価大項目	番号	評価項目(簡易版)	評価項目(詳細版)	説明(具体例)			段位			得点			
					上位 3点	中位 2点	下位 1点	無段位 0点	満点	実際 得点	重み 付け	最終 得点		
言語力 4項目	言語力	1	文型/ 表現	文型、表現の適切さ	・文型、慣用表現、話し言葉と書き言葉が適切か ・方言ではなく、標準語で書かれているか ・用言(動詞、形容詞、形容動詞)の活用形が適切か ・付属語(助詞、助動詞)の活用が脱落があるか					12.00	0.00	21%	0.00	
		2	用言/ 付属語	用言と付属語の適切さ	・文末のテンス(時制) ・ヴォイス(態) ・アスペクト(相) の適切さ	・テンス(過去、現在、未来) ・ヴォイス(能動、受身、使役) ・アスペクト(開始、進行、完了...) が適切か ・自己動詞、単語意味の誤りがあるか ・意味がとれないものや他の違う意味を取れる誤りがあるか ・連体詞、副詞、接続詞、感動詞の誤りがあるか ・指示語(こそあと言葉)の指している対象が明確か								
		3	テンス/ ヴォイス/ アスペクト	・文末のテンス(時制) ・ヴォイス(態) ・アスペクト(相) の適切さ	・テンス(過去、現在、未来) ・ヴォイス(能動、受身、使役) ・アスペクト(開始、進行、完了...) が適切か ・自己動詞、単語意味の誤りがあるか ・意味がとれないものや他の違う意味を取れる誤りがあるか ・連体詞、副詞、接続詞、感動詞の誤りがあるか ・指示語(こそあと言葉)の指している対象が明確か									
		4	語彙	・単語意味の正確さ ・(助詞以外)活用のない品詞の正確さ	・自己動詞、単語意味の誤りがあるか ・意味がとれないものや他の違う意味を取れる誤りがあるか ・連体詞、副詞、接続詞、感動詞の誤りがあるか ・指示語(こそあと言葉)の指している対象が明確か									
思考力	思考力	5	先行研究	・先行研究の豊富さ ・先行研究や文献調査に対する検討	・先行研究や文献調査をまとめ、研究背景が十分に反映されているか ・先行研究や文献調査を批判的検討が行えられているか 以上のいずれかに該当するか					9.00	0.00	16%	0.00	
		6	研究目的	・研究課題の研究価値 ・研究目的、背景、意義の明確さ	・研究価値がある課題が見つけれられるか ・研究目的が明確か ・研究背景、意義が明記されているか ・要旨が目的に応じて、洗練的に詳しく書いているか ・要旨は本文の内容を反映しているか									
		7	要旨	要旨の明確さ	・研究価値がある課題が見つけれられるか ・研究目的が明確か ・研究背景、意義が明記されているか ・要旨が目的に応じて、洗練的に詳しく書いているか ・要旨は本文の内容を反映しているか									
		8	研究結果	結論の明確さと課題意識	結論が十分に明確に論じられているか、研究限界や不足が意識されているか									

4. おわりに

中国の人材能力に対する要求の変化に応じ、従来の研究にはなかった評価基準の作成を試みた。中国における「新国標」に言及された求められている能力に適応した、日本語学習者の研究論文に特化したAS評価基準の作成を目指し、その基準の原案の妥当性と信頼性について検討し、調査分析結果により、高い妥当性と信頼性を備えた「日本語研究論文評価基準」を構築した。

具体的な方法としては、まず、「評価基準原案」の各評価小項目を必要度により5段階評価するアンケート調査を行った。この調査結果を用いてクロンバック α 係数を計算し、「評価基準原案」の高い信頼性を確認した。次に、探索的因子分析を行い、5因子を持っている因子構造が確認できた。また、その因子構造の妥当性と信頼性を検証するために、検証的因子分析を行った。モデルの適合度の検討に基づき、計19項目で構成した言語力モデルと斜交モデルは良いモデルの条件を満たしていることがわかった。構成概念妥当性が検証できたと考えられる。さらに、「評価基準原案」に「評価大項目」「段位」「得点」と「成績」の欄を追加改良し、最終的に「日本語研究論文評価基準」を完成させた。最後に、「日本語研究論文評価基準」のクロンバック α 係数を計算した結果、高い信頼性があることも示された。

本研究で構築した「日本語研究論文評価基準」には以下の特徴がある。

(1) 本基準の評価内容は「新国標」の要求により、論文審査の要点である研究意義、論理構築、専門能力及び学術規範を評価指標にする論文作成指導に役立つ。

(2) 本基準には、評価基準や観点を詳細に示すAS評価とする具体的な評価基準が示されている。そのため、評価者が論文を評価することにも、教師が論文作成を指導することにも、作者自身が論文を修正することにも広く利用できる。

(3) 本基準は、高い妥当性と信頼性を備えている。

本研究では、新しい評価基準が妥当性と信頼性を備えることを検証したが、実用性についてはまだ調査する必要がある。中国人日本語教師が日本語研究論文を「日本語研究論文評価基準」と「現行論文評価基準」を用いて別々に評価し、その両者を、信頼性、評価者間の評価一致度と評価者内省の3つの面から比較するのが、今後の課題であると考えられる。

注：

- ¹ 中国教育部『本科畢業論文（設計）抽檢办法（試行）』、http://www.gov.cn/zhengce/zhengceku/2021-01/07/content_5577651.htm、2020年12月24日。
- ² 本研究で扱う評価対象の論文は、8000～15000字程度の専門性を有する研究論文として研究を進める。
- ³ 評価方法としては、さまざまな方法があるが、その中に Holistic scoring (HS) と Analytic scoring (AS) という評価方法が知られている。HS とは、作文の全体的印象を基に単一の得点で作文の評価をする評価方法のことであり、AS は作文を複数の観点から採点し、最後に個々の観点の得点を合計して評価する方法である。
- ⁴ 例として、李清华、孔文（2010）「TEM4写作新分項式评分标准的多层次 Rasch 模型分析」『外语电化教学』p19-25
- ⁵ 南風原朝和、市川伸一、下山晴彦（2001）p75-76
- ⁶ 朱徳全（2016）p114-115
- ⁷ 項目が削除された場合の Cronbach のアルファの数値が全体の α 係数と比べて小さくなるようならば、その項目を除くと残り全体の同質性は高まる。尺度を作り直す時に削除する目安となる。
- ⁸ 松尾太加志、中村知靖（2002）p12-14
- ⁹ 平井明代（2017）p194
- ¹⁰ 帰無仮説はある仮説が正しいかどうかの判断のために立てられる仮説である。大抵は否定されることを期待して立てられる。検定量の値が得られる確率が極端に低ければ、帰無仮説を棄てて、帰無仮説を否定する形式の対立仮説を採用する。
- ¹¹ 呉明隆、涂金堂（2012）p695
- ¹² 平井明代（2017）p194
- ¹³ 呉明隆、涂金堂（2012）p701
- ¹⁴ 有意確率は0.000 (<0.001)、帰無仮説の「観測変数間に無相関である」という可能性は千分の一未満と示し、帰無仮説を棄却されるべき、「観測変数間に相関がある」と判断された。観測変数間に相関があるということで因子分析を行うには妥当であるといえる。（平井明代（2016）p134）
- ¹⁵ 固有値を縦軸、因子の数を横軸にとって、固有値の変化をプロットしたもので、因子の数を定めるときに、参考にする。固有値のグラフがなだらかなる前までで因子の数とする。そのようなグラフのことをスクリープロットと言う。（松尾太加志、中村知靖（2002）p54）
- ¹⁶ 固有値とは各因子の全項目に対する支配度である。通常、カイザーガットマン基準により因子数の基準となる固有値の最小値は「1」に指定してある。（HF Kaiser, “The application of electronic computers to factor analysis”, Educational and Psychological Measurement, 20, 1960, p.141-151）
- ¹⁷ 共通性は抽出された全ての因子が各観測変数と共有する分散の割合を示す。0から1の値を取る、一般的には、0.2以下の項目があれば、該当項目の削除を検討してもよ

い。(朱徳全 (2016) p134)

¹⁸ 6番目までとその以降では変化のしかたがなだらかになった。

¹⁹ Velicer W. F., Fava J. L. (1998) p.231-251

²⁰ 因子寄与は因子が質問項目に対してどの程度寄与しているかという指標である。因子寄与率はそれを割合で示すものである。その寄与率を合計していったもので、累積寄与率と言われる。60%以上で良い。(松尾太加志、中村知靖 (2002) p84-85)

²¹ 確認的因子分析とも呼ばれる。実態は共分散構造分析である。

²² 因果関係の仮説は項目間を矢印で結んだパス図と呼ばれる図で表し、項目間の関係の強さを表すパス係数と呼ばれる値が求められ、パス図の矢印線上に記載される。パス係数の大小によって因果関係が解明できる。パス係数は0.4以上が望ましい。

²³ 呉明隆、涂金堂 (2012) p746-748

²⁴ 王培剛 (2020) p162

²⁵ 豊田秀樹 (2007) p18

参考文献

日本語

南風原朝和、市川伸一、下山晴彦 (2001) 『心理学研究法入門』東京大学出版会

松尾太加志、中村知靖 (2002) 『誰も教えてくれなかった因子分析』北大路書房

平井明代 (2017) 『教育・心理系研究のためのデータ分析入門 (第2版)』東京図書

豊田秀樹 (2007) 『Amos (編) 共分散構造分析—共分散構造方程式モデリング—』東京図書

中国語

朱徳全 (2016) 『教育測量学』中国人民大学出版社

呉明隆、涂金堂 (2012) 『SPSS 与統計応用分析』東北财经大学出版社

王培剛 (2020) 『多元統計分析与 SAS 実現』武漢大学出版社

英語

Velicer W. F., Fava J. L. (1998) “ Affects of Variable and Subject Sampling on Factor Pattern Recovery”, Psychological Methods (3)

【参考資料1】

日本語研究論文の評価項目に関する調査

ご協力いただきみなさまへお願い。

本調査は、第二言語としての日本語による研究論文評価基準の評価項目を構成することのために利用いたします。協力者の先生方には評価基準に用いた25項目を評価していただき、簡単なアンケートに答えていただきます。所要時間は約20分です。

調査への参加は協力者の自由意思によるもので、参加したくない場合および途中で辞めたい場合などは辞退できます。また、辞退によって協力者が不利な扱いを受けることはありません。

入力いただいたデータと個人情報の管理には十分な注意を払い、学術的な研究以外の目的には使いません。また、データを匿名にて使用させていただくことがあります。

調査実施者

高帥

広島大学大学院 文学研究科

E-mail : 276055048@ qq.com

以上に同意いただける方は、アンケートに進んでください。

今から評価していただく評価項目は、第二言語としての日本語による研究論文を評価するために作成したものです。この研究論文は8000～15000字程度の専門性を有する大学卒論みたいなものです。説明文や意見文のような作文ではない。

研究論文を評価する時、自分にとって、それぞれの評価項目がどの程度重要であるかを「非常に重要である」「重要である」「普通である」「重要でない」「全く重要でない」の5件法で回答を求めます。それぞれ5点から1点を配点していただきたいです。

番号	評価項目 (簡易版)	評価項目 (詳細版)	説明 (具体例)	必要程度 (5~1)
1	文体	文末スタイルの統一	「普通体」と「丁寧体」が適切に使われ、統一されているか	
2	明快さ	文が分かりやすい	<ul style="list-style-type: none"> ・日本語らしくない文 ・意味不明な曖昧文 ・意味が複数の解釈できる多義文 ・主語、述語、目的語間の関係やそれらの順序の不適切文 ・文の長さが不適切で、洗練されていない文のような文がないか 	
3	文型/ 表現	文型、表現の適切さ	<ul style="list-style-type: none"> ・文型、慣用表現、話し言葉と書き言葉が適切か ・方言ではなく、標準語で書かれているか 	
4	用言/ 付属語	用言と付属語の適切さ	<ul style="list-style-type: none"> ・用言 (動詞、形容詞、形容動詞) の活用形が適切か ・付属語 (助詞、助動詞) の誤用や脱落があるか 	
5	テンス/ ヴォイス/ アスペクト	<ul style="list-style-type: none"> ・文末のテンス (時制) ・ヴォイス (態) ・アスペクト (相) の適切さ	<ul style="list-style-type: none"> ・テンス (過去、現在、未来) ・ヴォイス (能動、受身、使役) ・アスペクト (開始、進行、完了...) が適切か 	
6	語彙	<ul style="list-style-type: none"> ・単語意味の正確さ ・(助詞以外) 活用のない品詞の正確さ 	<ul style="list-style-type: none"> ・意味がとれないものや他の違う意味を取れる誤りがあるか ・接続詞、感動詞の誤りがあるか ・指示語 (こそあど言葉) の指している対象が明確か ・自己動詞、単語意味の誤りがあるか 	
7	表記	文字、カナ、外国語の正確さ	<ul style="list-style-type: none"> ・カナの不適切 (例: 中国じん× ⇒ 中国人 ○) ・カナの誤り (例: 信じる× ⇒ 信ぢる ○) ・外来語表記の誤り (academic× ⇒ academic ○) ・脱字、長音、促音などの誤りがあるか 	
8	先行研究	<ul style="list-style-type: none"> ・先行研究の豊富さ ・先行研究や文献調査に対する検討 ・先行研究に対する理解の正確さ ・先行研究から引用された内容の適切さ 	<ul style="list-style-type: none"> ・先行研究や文献調査をまとめ、研究背景が十分に反映されているか ・先行研究や文献調査を批判的検討が行えているか ・論証しようとする話題、史料、文学作品、著者の主張などに対する理解が正確か、適切か、先行研究が適切に引用されているか 以上のいずれかに該当するか 	
9	研究目的	<ul style="list-style-type: none"> ・研究課題の研究価値 ・研究目的、背景、意義の明確さ 	<ul style="list-style-type: none"> ・研究価値がある課題が見つけれられるか ・研究目的が明確か ・研究背景、意義が明記されているか 	
10	研究テーマ	研究テーマの適切さ	<ul style="list-style-type: none"> ・研究内容との一致性などが適切か ・研究範囲の広さが適切か、研究内容が明確か ・テーマが充分掘り下げられているか 	
11	要旨	要旨の明確さ	<ul style="list-style-type: none"> ・要旨が目的に応じて、洗練的に詳しく書いているか ・要旨は本文の内容を反映しているか 	
12	研究対象	(研究/対比) 対象の適切さ	<ul style="list-style-type: none"> ・(研究/対比) 対象が研究内容に適切か 	
13	信頼性	<ul style="list-style-type: none"> ・理論/データに基づき、研究資料/調査や実験の内容に対する分析 ・効果的な情報収集手段で、必要で信頼できる情報の収集 	<ul style="list-style-type: none"> ・研究分野に関する理論を踏まえているか ・研究対象から集めたデータの完成度、信頼性、妥当性などが証明されているか 以上のいずれかに該当するか 	
14	研究結果	結論の明確さと課題意識	結論が十分に明確に論じられているか、研究限界や不足が意識されているか	

15	独創性	<ul style="list-style-type: none"> 既存の研究に自分の主張/独自の着眼点 新しい資料や事実に基づき、独創的な研究成果 	<ul style="list-style-type: none"> 分析や比較により、先を予想し/共通点と相違点を論じ、自分の主張が示されているか 未知の抽象的な現象/実物の発見/発明の論述 新しい研究方法/理論の構築/展開 新しい概念/解釈の提出 以上のいずれかに該当し、新提案の優位性が示されているか 	
16	展望性	考察の適切さと客観性	考察が結果に基づいて適切に導き出されている	
17	一貫性	研究プロセスの一貫性	文章全体を通して、研究目的に対応した結論が導かれたまでのプロセスが緻密で、一貫か	
18	研究方法	<ul style="list-style-type: none"> 研究方法の適切さ 研究方法に対する説明の明確さと論理性 調査内容や実験に対する設計の緻密さ 	<ul style="list-style-type: none"> どんな方法で論証されているか(例文で、対比で...) どんな理論に基づき論証されているか どんな実験で論証されているか 調査内容や実験の設計が妥当か、安全か、参考資料が適切に引用されているか 以上のいずれかに該当するか 	
19	説得力	<ul style="list-style-type: none"> 絞られている論点 充実な研究内容と説得力がある根拠 	<ul style="list-style-type: none"> 物事を多面的に捉え、論点が絞られているか 課題に沿って、書き手の主張が、説得力のある根拠とともに明確に述べられている。推量「だろう」、伝聞「(する) そうだ」、判断「べきだ」のような主観的表現が過剰使用されていないか 	
20	読み手	<ul style="list-style-type: none"> 読み手に対する配慮 文と文の間のロジックの適切さ 	<ul style="list-style-type: none"> 読み手に対する配慮があるか(例えば、読み手にないと思われる文化的背景を説明する) 文と文の間、矛盾がないか 	
21	バランス	本文全体構成(論証の目的・プロセス・結論)のバランス	文章全体の組み立てが整っている。論文の長さにより、序論・本論・結論のバランスが良いか、目安として、1:3(～5):1	
22	段落	<ul style="list-style-type: none"> 段落分け 段落間のつながり 無意味な段落 	<ul style="list-style-type: none"> 段落分けが適当か、 段落間のつながりが自然で流暢か 論証に無意味な段落があるか(2文以上で構成されている段落) 	
23	文	<ul style="list-style-type: none"> 文と文のつながり 無意味な文 	<ul style="list-style-type: none"> 段落内、文と文のつながりが自然で流暢か 論証に無意味な文があるか 	
24	メタ言語	<ul style="list-style-type: none"> メタ言語の正確さと適切さ メタ言語がなくても、論理的に自然につながっている 	<ul style="list-style-type: none"> 予告のメタ言語「まず」、「次に」、 まとめのメタ言語「以上述べたように」のようなメタ言語があるか 	
25	書式	<ul style="list-style-type: none"> デジタル化した情報の統合と効果的な表現 書式の位置づけ、書き方、読みやすさ 	<ul style="list-style-type: none"> 目次、ページ番号 段落の間隔、行間、字間、文字種類、文字サイズ 引用(脚注)、注釈や参考文献の表示が適切で、読みやすい 	

性別： 男 女
 年齢： 20代 30代 40代 50以上
 日本語教育歴： 5年以下 5～10年 10年以上
 論文指導歴： 経験なし 5年以下 5-10年 10年以上
 日本人日本語教師に論文を指導された経歴：有り 無し

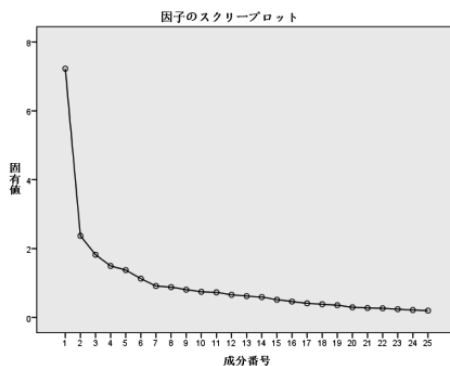
お疲れ様でした。
 どうもありがとうございました。

【参考資料2】

共通性		
	初期	因子抽出後
文体	1.000	.483
明快さ	1.000	.695
文型	1.000	.683
用言	1.000	.621
テンス	1.000	.642
語彙	1.000	.739
表記	1.000	.675
先行研究	1.000	.623
研究目的	1.000	.587
研究テーマ	1.000	.607
要旨	1.000	.578
研究対象	1.000	.532
信頼性	1.000	.639
研究結果	1.000	.650
独創性	1.000	.664
展望性	1.000	.619
一貫性	1.000	.645
研究方法	1.000	.513
説得力	1.000	.592
読み手	1.000	.592
バランス	1.000	.619
段落	1.000	.569
文	1.000	.626
メタ言語	1.000	.587
書式	1.000	.644

因子抽出法:主成分分析

KMO および Bartlett の検定		
Kaiser-Meyer-Olkin の標本妥当性の測度		.823
Bartlett の球面性検定	近似カイ 2 乗	1281.860
	自由度	300
	有意確率	.000

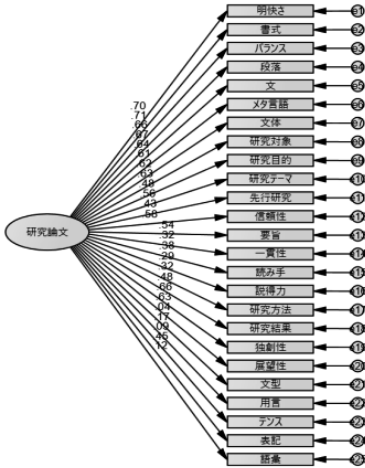


説明された分散の合計 (固有値 > 1 の部分)

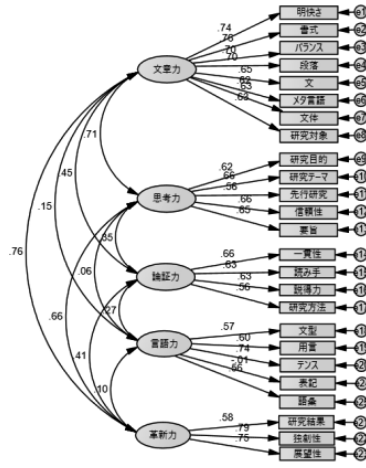
成分	初期の固有値			抽出後の負荷量平方和			回転後の負荷量平方和		
	合計	分散の%	累積%	合計	分散の%	累積%	合計	分散の%	累積%
1	7.227	28.908	28.908	7.227	28.908	28.908	4.548	18.193	18.193
2	2.372	9.489	38.397	2.372	9.489	38.397	2.712	10.848	29.041
3	1.824	7.294	45.691	1.824	7.294	45.691	2.408	9.634	38.675
4	1.496	5.985	51.676	1.496	5.985	51.676	2.252	9.006	47.681
5	1.377	5.508	57.184	1.377	5.508	57.184	2.162	8.646	56.327
6	1.128	4.511	61.695	1.128	4.511	61.695	1.342	5.367	61.695

因子抽出法:主成分分析

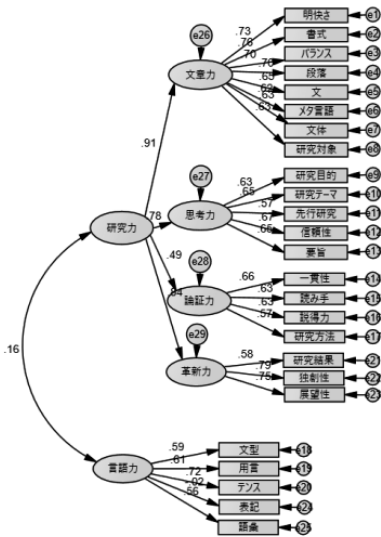
【参考資料 3】



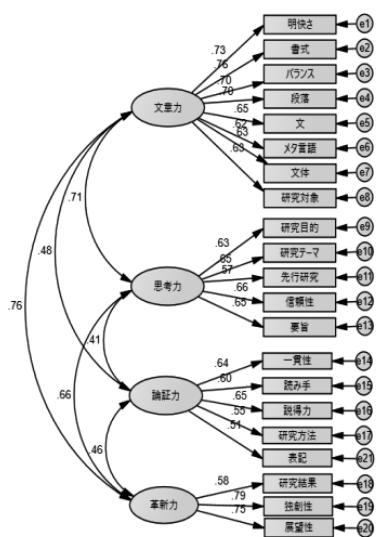
モデル 1



モデル 2



モデル 3



モデル 4