

日本留学試験「記述問題」が測っているもの

村上京子（名古屋大学留学生センター）

要旨：

「記述問題」の得点に関わる要因として、書き手の諸要因のほか、課題・文章の種類・時間制限・評価基準などがあげられる。これらの要因によって、同一の書き手であっても、書かれた作文の評価の間には大きな差が出てくることを指摘した。

また、大学進学を目指す学習者に日本留学試験の聴解・聴読解・読解・記述を実施し、聴解・聴読解・読解の成績と記述問題の評価との関係を調べた。その結果から、記述問題で測られているものが、他の日本語試験で測られている聴解等の日本語能力とは多少異なる側面であることがわかった。以上から、信頼性を低める要因をできるだけ取り除くとともに、大学での勉学上必要な「書く能力」を反映するような課題の工夫が必要であることを論じた。

キーワード：記述問題、誤差要因、書き手要因、課題要因、聴解・聴読解・読解の成績

0. はじめに

日本留学試験において「記述問題」が加えられたことは、大学受験を目指す学習者にとっても、日本語を指導する教師にとっても大きな変化といえよう。大学入学後、レポートを作成したり、試験の答案を書いたりするといったまとまりのある文を書くことは頻度も高く、学部留学生自身も困難と感じる技能の1つである（村上,2003）。

しかし、作文の評価にはさまざまな困難が伴う。課題や出題形式など妥当性や信頼性に関わる要因、評定の安定性・一貫性など作文の評価一般について指摘される問題点のほかに、日本留学試験「記述問題」固有の問題も挙げられる。ここでは、まず「記述問題」の評価に関わる要因についてかんがえ、次に実際の学習者の得点間の相関から、「記述問題」が何を測っているのか考えていきたい。

1. 記述問題に関わる諸要因

Cooper(1984)は、小論文を使って Writing ability を評価する際にその得点に関して誤差要因として働くものに、書き手 (writer)、題目 (topic)、形式 (mode)、時間制限 (time-limit)、テスト状況 (examination situation)、そして評定者 (rater) を上げている。もちろん、評価の観点項目やその配点も重要な要因の1つである。ここでは、その中のいくつかについて取り上げ、考えてみることにする。

1. 1 書き手要因

書き手の要因としては、もちろんその人の文章を書く能力が測られるべきであることは当然だが、一般的に「書く能力」というものが存在し、1回の試験でそれが測れ

るものであるかどうかは疑問である。また実際の試験においては、疲れや慣れ、モチベーションなどの情動的な要因の他に「書く能力」以外の知識などさまざまな事柄が関係する。

特に「記述問題」では、受験者の受験準備が重要な要因となりうる。某国立大学に合格した私費留学生 27 名に受験準備に関するインタビューをした結果、全員が受験以前に日本語学校等で「記述問題」の講習を受け、試験と同じ形式の作文を書く練習をしていた。さらに、そこで受けた書き方の指導にもかなり共通したパターンがあることがわかった。

日本留学試験を運営する日本学生支援機構(Japan Student Services Organization : JASSO) が公開している採点基準(資料参照)では、「執筆者の意図が明快に理解可能である」ことと、「主張に根拠が示されており、かつ、主張と根拠との間に十分な論理的関係がある」ことが評価のポイントになると示されている。「文法・表記上の軽微な誤りや文体上やや不自然な点は許容」される。したがって、一定のパターンに沿って書いていけば高い得点を得ることは可能だと考えられる。

1. 2 トピック要因

記述問題に関して課題(topic)が異なれば、同一の学習者でもその評価に大きな変動がありうる。従来の研究でも、母語で書いた作文間では相関が 0.3 程度かそれ以下と低い(Coffman:1966,池田:1992,小嶋・村上:1991)。これに対し、学習者が母語以外で書いた作文では、評価されるものが母語話者の作文とは大きく異なるため、複数のトピック間の相関は母語の作文間に比べ高くなる。しかし、研究上の条件により多少相違があるものの 0.60~0.70 程度で十分に高い相関関係があるとはいえない(Jacobs et al.:1981,Carlson et al.:1985)。

廣瀬(2005)は同一の学習者 48 名に複数のトピックを与え、その作文間の評価の関係を調べている。その結果、2つの作文間に平均値の差は見られなかったが、相関係数では文法で 0.54、論理点では 0.35、合計で 0.57 とかなり低い値を報告している。この採点は JASSO の採点手続きに従って行ったものであるが、評価の信頼性の問題があったため、再度 4 名の評価者間一致度を高めて行い直しているが、やはり文法:0.58、論理:0.33、合計:0.62 となり、一方の作文から他方の作文の得点を推定するにはその標準誤差が大きく、2つのトピックから1つを選択させる方法に疑問を投げかけている。

1. 3 文の種類

文の種類・形式(mode)に関して、Brown et al.(1991)は英語のクラス分けテストで、文章を読んでそれに基づいて作文する課題と自分の経験から書く課題間の相関を求めている。その結果、相関は 0.36 と低く、同じ学習者を一貫して測定することは難しことを示した。このことから、利害関係のあるテストでは、複数の課題を与えて多方面から見ていくことの重要性を指摘している。

村上(2005)は、学部在籍する同じ日本語学習者の書いた文章を比較している。意見文と説明文の比較結果から、合計点間には 0.76 と比較的高い相関関係がみられたも

の、観点別にみていくと「文のわかりやすさ」や「内容」にはほとんど相関が見られない。反対に、「文法・語彙の正確さ」や「表現の多様性」「段落構成」などには、同じ個人にある程度一貫した傾向が見られた。また、「記述問題」で要求している 400 字程度という短い文章は、大学の勉強上書かなければならないレポートなどの文章とはいろいろな面で異なり、今後実際に大学で要求される文章と「記述問題」の関係も見ていく必要があると述べている。

1. 4 時間制限

時間制限 (time-limit) は、その時間内で考えをまとめ、指定された長さの文章を書けるかというスピードテストの側面と、時間切れで最後まで書ききれず本来の力を出せなかったという妥当性の問題が関係する。

日本留学試験では、20 分で 400 字の作文を課している。これは、TOEFL や MELAB など英語の大規模試験が 30 分で比較的短い文を書かせているのに準じたものだと考えられる。Jacobs et al.(1981)は、制限時間を変えた条件間で比較を行い、30 分という時間制限の中で十分妥当な評価をしようと結論付けている。しかし、従来の日本語教育において多くは1時間程度の比較的ゆったりした時間内で作文指導を行っている。日本留学試験という大規模な制約条件の厳しい中での実施のため、やむをえないのであろうが、前記のインタビューでも、多くの学生がともかく早く書く練習をしたと述べている。また中には読解問題を早く終え、記述に戻る時間を作ったという学生もあり、要領のよさも誤差要因に入るとは考えられる。大学に送られる記述問題の答案を独自の基準で採点したが、時間切れで文章の途中で切れている答案はほとんど見当たらなかった。その意味では 20 分という時間は適当なのであろうが、もしこれが「記述問題」対策の練習をしてきたからできた、また読解問題を早く終らせることができたからできた、のであれば問題があると考えられる。

2. 記述問題の評価

記述問題の得点が、日本語のほかの試験得点とどのような関係にあるのか、JASSO からは報告されていない。各大学には受験者の得点が知らされるため、独自にその関係を知ることはできる。しかし、その公表はできないことと、大学によっては受験者に偏りがあるため、ここでは某日本語学校で大学進学を目指して勉強している学習者のデータをもとに調べていきたい。

1. 2 でも述べたように記述問題の得点はトピックによって変動する。そこで実際は2つの課題から一方を選択してかかれたものであるが、1つの課題を選んだ学習者のみを対象に、その作文の評価とその学習者の聴解・聴読解・読解の得点間の関係をみていくことにする。

2. 1 対象者と試験問題

以下の課題を選んだ学習者 50 名が対象者である。学習者の国籍は中国、韓国の学生で、日本の大学等への進学を希望している。試験は日本留学試験の実施要領に基づき、同様の

時間配分で実施された。

課題：

ここ数年、コンピュータが発達し、Eメール（電子メール）を送ることもできるようになりました。ある人は、近い将来、郵便で配達する手紙は必要でなくなると言います。それに対し、別の人は、将来も郵便の手紙は必要だと言います。二つの意見を比べたとき、あなたはどちらの意見に賛成しますか。どちらかの意見に立ったうえで、理由を挙げて考えを400字程度で書いてください（句読点も含む。）。

2. 2 記述問題の採点

答案は村上(2003)に基づき、4名の教育関係者によって6つの観点から採点された。実際の記述問題の採点は資料にあるように2項目につき6点満点でされるのだが、この基準では詳しい情報がえられないことと、評定者のトレーニングをしたにもかかわらず十分な信頼性が得られなかったことから、以下の基準で行うことにした。

<評価の観点>

項目1) 正確さ

- 5：ほとんど誤りが無い（軽微な誤りが1,2箇所程度見られる）
- 4：中上級レベルの誤りが見られる（軽微な誤りが3箇所以上ある）
- 3：誤りが見られるが、初級レベル、重篤な誤りが1～2箇所
- 2：初級レベル、重篤な誤りが3～4箇所
- 1：初級レベル、重篤な誤りが5箇所以上

項目2) 文体（書き言葉）

- 3：文体が統一されている
- 2：文体が統一されていない（1, 2文混じっている）
- 1：文体が統一されていない（3文以上混じっている）

項目3) 表現の多様性

- 5：話しことばが全く使用されておらず、かつ、レポート・論文に使われるようなアカデミックな語彙や、チャレンジングな文型が使用されている
 - 4：話しことばが全く使用されていないが、レポート・論文の表現としてはやや不適切なところがある
 - 3：縮約形や、終助詞、助詞の省略、俗語、感情表現の使用が1,2箇所見られる
 - 2：縮約形や、終助詞、助詞の省略、俗語、感情表現の使用が3箇所以上見られる
 - 1：縮約形や、終助詞、助詞の省略、俗語、感情表現の使用が頻出する
- * その他減点対象になるもの
ひらがなの多用、体言止め（話しことば的表現とする）

項目4) 文のわかりやすさ

- 5 : 特にわかりにくい文がない
- 4 : わかりにくい文が 1 文ある
- 3 : わかりにくい文が 2 文ある
- 2 : わかりにくい文が 3 文ある
- 1 : わかりにくい文が 4 文以上ある

項目 5) 文章構成、全体構成

- 3 : 段落構成・意味関係が適切
- 2 : つながりの悪い箇所がある
- 1 : 段落構成がなっていない、段落意識がない

項目 6) 内容

- 5 : 根拠に論理的整合性があり、説得的である
- 4 : 根拠に説得力が見られるが、十分ではない
- 3 : 一貫してはいるが根拠が説得的ではない
- 2 : 最初に述べていることと、結論が一致していない (一貫性がない)
- 1 : 意見の根拠が示されていない、または意見が示されていない

2. 3 「記述問題」評価の結果

上の評価項目ごとに 4 名の評価者が独立に評価し、その平均と標準偏差、評定者間一致度 (α 係数) を示したものが表 1 である。

表 1. 「記述問題」評価結果

	正確さ	文体	多様性	文	段落構成	内容	計
平均	3.13	2.83	3.71	3.85	2.07	3.58	19.14
標準偏差	0.78	0.40	0.70	0.70	0.74	0.79	2.39
α 係数	0.89	0.97	0.81	0.73	0.92	0.78	0.89

表 1 のように 4 人の評価者の間の一致度は「文のわかりやすさ」の 0.73、「内容」の 0.78 を除いてすべて 0.8 以上であり、信頼性は高いといえよう。

合計得点の度数分布をあらわしたものが、図 1 である。50 名の学習者の記述の得点間にはある程度の散らばりがみられる。

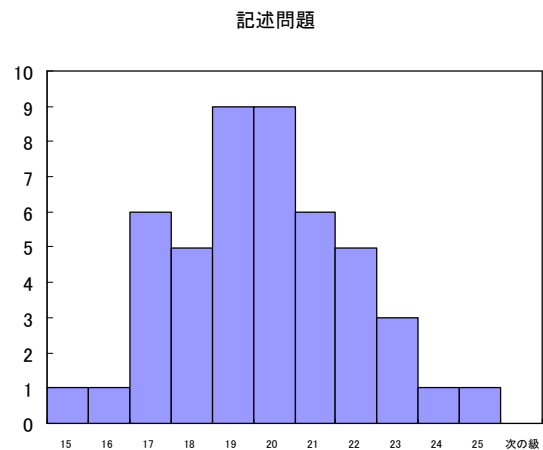


図 1. 記述問題の得点分布

3. 「記述問題」と「聴解・聴読解・読解」の得点間の関係

3. 1 「聴解・聴読解・読解」の得点

同じ学習者に「聴解」・「聴読解」・「読解」の問題を実際と同じ条件で実施した結果、その得点平均と標準偏差を表 2 に示す。

表 2. 聴解・聴読解・読解の平均得点

	聴解	聴読解	読解	合計
平均	18.34	15.96	16.06	50.36
標準偏差	1.51	2.35	1.96	4.48

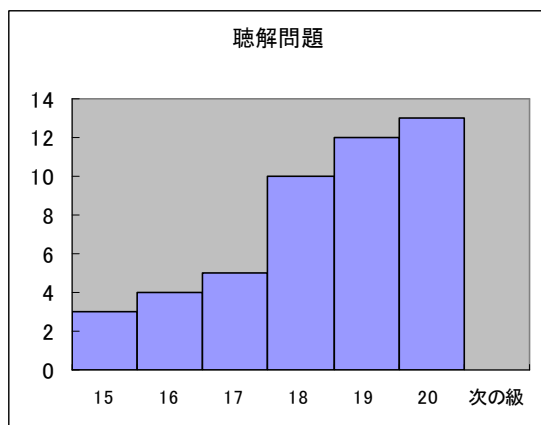


図 2. 聴解の得点分布

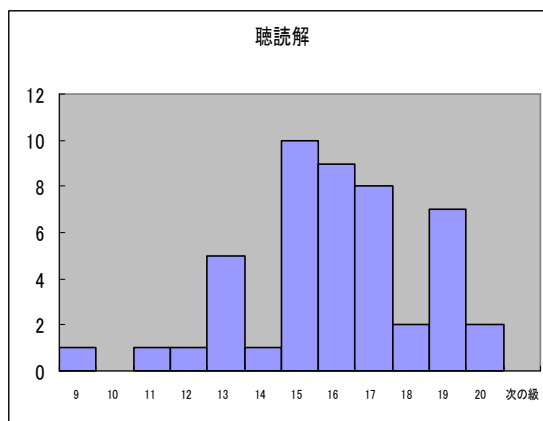


図 3. 聴読解の得点分布

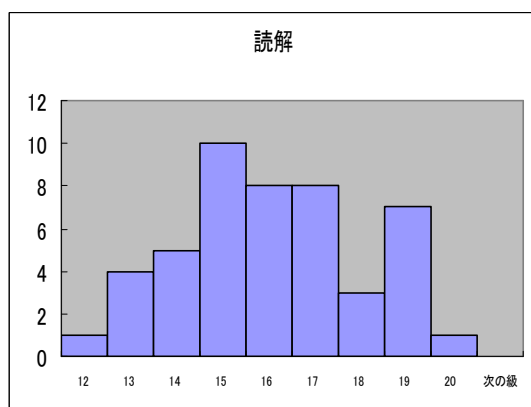


図 4. 読解の得点分布

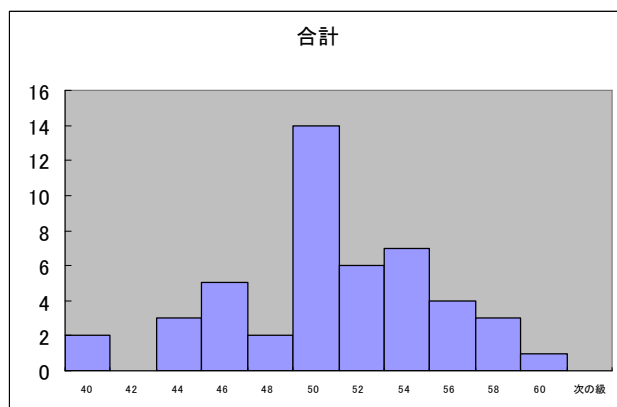


図 5. 合計点の得点分布

3. 2 「記述問題」と「聴解・聴読解・読解」間の相関

50名の学習者の各テスト得点を使ってテスト間相関を出したものが表 3 である。聴解と聴読解では 0.35、聴解と読解では 0.45、聴読解と読解では 0.36 という高くはな

いがそれなりの相関関係がみられた。記述の得点とはいずれも 0.30 程度の相関があったが、やはり低く、それぞれの合計点同士でかろうじて 0.4 程度の相関にとどまる。

表 3. 「記述問題」と「聴解・聴読解・読解」間の相関

	聴解	聴読解	読解	合計点	記述
聴解	1.00				
聴読解	0.35	1.00			
読解	0.45	0.36	1.00		
合計点	0.72	0.80	0.78	1.00	
記述	0.33	0.36	0.29	0.42	1.00

したがって、日本留学試験の各試験はそれぞれ方向の違う側面を測っていることがうかがえる。

4. 結論・考察

日本留学試験「記述問題」に関して、以下の問題点があげられよう。

1. 学習者に対する波及効果

進学を目的とする日本語学習（予備教育）において、これまで日本語能力試験では知識量測定が中心だったため、記憶中心の学習が行われてきた。それに対し、記述問題という産出能力を測る試験が加わったことによる変化は大きい。しかし、学習者が実際にどんな準備をしてきたかが重要である。できる限り実際の大学入学後の日本語運用に役立つ学習ができるような方向付けを入学試験の問題はすべきであろう。

その意味で、現在の問題形式はパターン化しており、多様な文章を学習する機会を奪っていると考えられる。実際の大学での運用を十分配慮した課題設定が必要だと考える。学習者が試験を目指して勉強してきたことが、入学後に有用であることは重要だと思うからである。

2. 選抜という観点から

採点基準が、文法的能力 3 点、論理的能力 3 点、合計 6 点満点であることから、選抜の資料として十分機能しないのではないかという指摘を村上（2003）ではした。ここでは、出題の方法として 2 題の問題の中から 1 題選び、それに関して意見文を書くという方法の信頼性に関して、トピック、文の種類という観点から分析した研究を紹介し、いずれも同じ学習者が書いた作文の間の相関はあまり高くないことを指摘した。

大学入学選抜という繰り返しのきかない重要な役割を担った試験であるため、妥当性・信頼性には十分な配慮が必要なことは言うまでもない。しかし、実際の試験の得点には非常に多くの要因が関わり、1 回の「記述問題」試験で何が測られているのか疑問である。情報を積み上げ、またそれをできるだけ公開することによって広く意見を集め、より信頼性の高い試験にしていく必要があると考える。

文献

- 池田 央 1992 『テストの科学』 日本文化科学社
- 小嶋秀夫、村上 隆 1991 「名古屋大学教育学部における論述式学力検査」『大学入試における実技・面接・小論文の評価に関する研究』平成2年度文部省科学研究費補助金 総合研究(A) 研究報告書 31-60
- 村上京子 2003 「学部留学生の大学生活における日本語運用上の困難と課題」名古屋大学留学生センター紀要 1. 5-17
- 村上京子・小室輝代・三谷閑子 2003 「日本留学試験『記述問題』における採点基準の見直し」『名古屋大学日本語・日本文化論集』11号 107-124
- 菅井英明 2003 「記述テストの特徴と比較」『日本語教育における評価法に関する基礎的資料整備とその分析』平成13~14年度文部科学省科学研究費補助金 基盤(C)(2) 研究報告書 80-95
- 平 直樹 1991 「小論文試験の方法論的諸問題に関する研究の動向について」『大学入試における実技・面接・小論文の評価に関する研究』平成2年度文部省科学研究費補助金 総合研究(A) 研究報告書 77-99
- 田中真理・長阪朱美 2004 「日本語と英語を目標言語とするライティング評価基準の展望：第二言語としての日本語のライティング評価基準作成に向けて」『第二言語としての日本語の習得研究』7号 214-253
- Brown, J. D. 1991 Do English and ESL Faculties Rate Writing Samples Differently? TESOL QUARTERLY 25(4). 587-603.
- Brown, J. D., Hilgers, T., & Marsella, J. 1991 Essay prompts and topics: minimizing the effects of mean differences. Written Communication 8, 533-556
- Carlson, J., Bridgeman, B., Camp, R., Waanders, J. 1985 Relationship of admission test scores to writing performance of native and nonnative speakers of English. TOEFL Research Reports No.19. Princeton, N.J: Educational Testing Service.
- Coffman 1966 On the validity of essay tests of achievement. Journal of education measurement 3(2), 151-166
- Cooper, P. L. 1984 The Assessment of Writing Ability : A Review of Research. GRE Board Research Report GREB.No.82-15R ETS Report 84-12
- Cumming, A. 1990 Expertise in evaluating second language compositions. Language Testing 7(1). pp. 31-51.
- Cumming, A. 2001 ESL/EFL instructors' practices for writing assessment: specific purposes or general purposes? Language Testing 18(2). pp. 207-224.
- Hamp-Lyons, L. and Henning, G. 1991 Communicative Writing profiles: An Investigation of the Transferability of a Multiple-Trait Scoring Instrument Across ESL Writing Assessment Contexts. Language Learning 41(3). pp. 337-373.

- Jacobs, H. I., Zinkgraf, S. A. Wormuth, D. R., Harfiel, V. F. & Hughey, J. 1981
English composition program. Testing ESL composition; a practical approach.
Rowley, Mass: Newbury House.
- Lumley, T. 2002 Assessment criteria in a large-scale writing test: What do they really
mean to the raters? Language Testing 19(3). pp. 246-276.
- Polio, C. G. 1997 Measures of Linguistic Accuracy in Second Language Writing
Research. Language Learning 47(1). pp. 101-143.
- Raimes, A. 1990 The TOEFL Test of Written English: Causes for Concern. TESOL
QUARTERLY 24(3). pp. 427-442.
- Silva, T. 1993 Toward an Understanding of the Distinct Nature of L2 Writing: The
ESL Research and its Implications. TESOL QUARTERLY 27(4). pp. 657-667.
- Weigle, S. C. 1994 Effects of training on raters of ESL compositions.

<資料>

日本学生支援機構（Japan Student Services Organization : JASSO）のホームページから

日本留学試験

1. 目的

外国人留学生として我が国の大学（学部）等に入学を希望する者について、言語の形式に係る能力及び論理を構築する能力を測定することを目的とします。

2. 採点基準

日本語は、「記述」、「聴解」、「聴読解」、「読解」と4領域に分かれています。「記述」以外の3領域についてはマークシート方式の試験ですが、「記述」は筆記試験なので、他の3領域の得点には含まれません。

「記述」の採点に当たっては、文法的能力及び論理的能力のそれぞれについて、以下の基準に基づき採点し、その合計点（0～6点）を表示することとします。

(1) 文法的能力（0～3点）

- 個々の文についても、文章全体についても、執筆者の意図が明快に理解可能であるもの（文法・表記上の軽微な誤りや文体上やや不自然な点は許容する。）……………3点
- 文法・表記上明らかに適切でない点を含むが、文章全体から執筆者の意図は明快に理解可能であるもの……………2点
- 文法・表記上明らかに適切でない点がかかなり目立つが、文章全体から執筆者の意図を想像することは可能であるもの……………1点
- 意味不明の文が多く、文章全体から執筆者の意図を理解することが不可能又は極めて困難なもの……………0点

(2) 論理的能力 (0～3点)

- 主張に根拠が示されており、かつ、主張と根拠との間に十分な論理的関係があり、矛盾が認められないもの 3点
- 主張に根拠が示されており、概ね論理的な関係が認められるが、一部に論理的矛盾や非整合性も存在するもの 2点
- 主張は示されているが、その根拠が示されていない、又は、根拠が示されていても、論理性・客観性を著しく欠いているもの 1点
- 筆者自身の主張が示されていない、又は、何を主張したか曖昧であるもの 0点

3. 成績利用に当たっての注意事項

上記2の基準による採点結果は、文法的能力及び論理的能力に限った実施者側の評価を表したものであり、受験者の意見・主張等についての評価は行っていません。

本会では、大学等に対して答案の写し(受験者自筆のもの)を評価と併せて送付しますので、大学等で定める独自の基準により、直接評価を行うことも可能です。