学習者の役に立つ日本語会話テストの開発
理論、先行研究、そして実践

Development of Useful Oral Japanese Language Tests: Theory, Research, and Practice

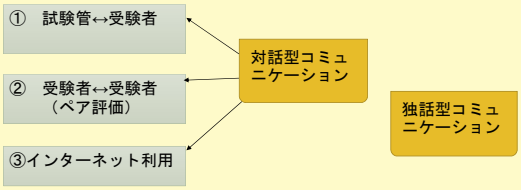
近藤ブラウン妃美（ハワイ大学マノア校）
Kimi Kondo-Brown (University of Hawai'i at Mānoa)

JOPTキックオフシンポジウム@東京外国語大学
2014年3月20日（木）13:00-17:00

1

会話テストとは

「会話とは，話し手と聞き手双方による対話型コミュニケーションと，話者一人が一方的に行う独話型コミュニケーションの両方を意味し，会話テストは両形式のものを指す」（伊東 2008, p.98）



2

対話型コミュニケーションの会話テスト

	✓ 単独形式のテストを行う場合	✓ 複数形式のテストを行う場合
インタビュー形式 interview format	• 個人情報や日常の最も身近な事柄に関する質問（初級） • 専門領域や時事問題に関する抽象的な話題についての質問　（超級）	
視覚的プロンプト形式 visual prompt format	• 単数の写真やイラスト • いくつかのコマでできた絵物語（談話能力）	
ロールプレイ形式 role-play format	• フォーマル vs. インフォーマルな場面 (formal vs. informal) • 社交的なやりとり(interpersonal) 及び商品やサービスを得るためのやりとり (transactional)	

3

パフォーマンス・テストとしての会話テスト

「パフォーマンス・テスト」は，広い意味で，受験者が目標言語を使って「何か」を行うテストのすべてを意味する (Brown, Hudson, Norris, & Bonk, 2002)

実世界で実際に起こりうる、目的を持った言語活動を成し遂げる能力の推測を意図するテスト（Brown et al., 2002; McNamara, 1996; Norris, Brown, Hudson, & Yoshioka, 1998; 近藤ブラウン 2013; Shohamy, 1995）

4

会話テストの目的

学習者を支援する評価という観点からは、形成的評価は非常に大切 (e.g., Popham, 2008; Falsgraf, 2009)

学習者のモチベーションや教師のフィードバックの質を高められる会話テスト

✓ **総括的評価**
(summative assessment)
- **成績判定**
- **資格認定**

✓ **形成的評価**
(formative assessment)
- **モチベーション**
- **フィードバック**

✓ **診断的評価**
(diagnostic assessment)
- **プレースメント**

5

役に立つテスト作りの基本事項

1. 構成概念的妥当性 construct validity	2. 信頼性 reliability	3. 真正性 authenticity
4. 相互作用 interactiveness	5. 実用性 practicality	6. 影響力 impact

Bachman & Palmer (1996)

6

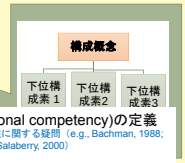
構成概念的妥当性 (construct validity)

意図している能力を実際に評価・測定できているか

「会話能力」という構成概念をどうとらえるか

会話サンプル

会話サンプルの評価と解釈



- ✓ 機能的能力(functional competency)の定義
  ACTFL OPI の構成概念妥当性に関する疑問（e.g., Bachman, 1988; Bachman & Savignon, 1986; Salaberry, 2000）
- ✓ OPIで、現実の日常場面に起こる自然会話での相互行為能力(interactional competence)」が、どの程度評価できているのか（e.g., He & Young, 1998; Johnson, 2001; Johnson & Tyler, 1998; Lazaraton, 2002; Van Lier, 1989）

7

信頼性 (reliability)

テスト結果の一貫性や安定性

テスト結果の一貫性や安定性をできるだけ高める工夫が必要 **（近藤ブラウン 2012)**

8

OPI判定の信頼性

ACTFL OPI試験官によるレベル判定の信頼性

✓１つのOPIにつき２名（もしくはそれ以上）の認定試験官が採点し、得られた２組（もしくはそれ以上）の得点の一致度、つまり評定者間信頼性係数 (inter-rater reliability)を分析 (例.Dandonoli & Henning,1990; Surface & Dierdorff, 2003; Thompson, 1995）

✓これらの調査では、ACTFL OPIの判定には適度な信頼性が見られるという結果が報告されている

9

OPI判定の信頼性(Cont'd)

試験官と受験者間の対応の一貫性

✓試験官の発話サンプルの抽出法（質問の仕方）に、一貫性があるかどうかを調査。一貫性がない場合、試験官の質問の仕方が、どれほど**受験者のパフォーマンスに影響**を与えているのか (Brown , 2003; Ross, 2007; Ross & Berwick, 1992)

✓ 二人の試験官が同一の受験者とOPIを行い、その過程を比べたところ、両試験官の発話サンプルの引き出し方に顕著な違いが見られ、パフォーマンスにも影響を与えた(Brown , 2003)

10

採点ルーブリック使用のメリット

(see Arter & McTighe, 2001 ; Davis & Kondo-Brown, 2012)

- ✓採点の一貫性が高まる(consistency)
- ✓採点の透明性 が高まる (transparency)
- ✓学習者に効果的にフィードバックを与えられる
- ✓成績・能力判定についての説明に役立つ
- ✓同じコメントを何度も書く手間がはぶける

11

評定者訓練

✓採点基準があっても、採点者がその使い方について適切な訓練を受けていないと、妥当性や信頼性のあるテスト結果を得るのが難しくなる (Weigle, 2002)

✓評定者訓練を受けるなどして，評定者としての自分の特性を自覚し、なるべく公平な判定ができるように努力する(McNamara, 1996)

✓評定者訓練を受けることにより、各採点者の自己一貫性（self-consistency）はある程度改善されるが、採点の厳しさに関しては、訓練の効果が出にくい（例. Kondo-Brown, 2002; Lumley & McNamara, 1995; McNamara, 1996; McNamara & Adams, 1991）

➡ 大学入試や資格試験などの重要テスト（high stakes tests) の採点は、少なくとも二人以上の採点者が必要

12

真正性 (authenticity)

現実の場面で起きると予測される言語活動 = テストの内容や形式

場面的真正性（situational authenticity）vs.
相互行為的真正性（interactional authenticity)
(Ellis, 2003, pp. 305-306）

13

対人的な相互作用

テスト結果のOWNERSHIPに関する根本的な問いかけ

✓ 個人の属性(personal attribute) としての言語能力

だれのパフォーマンスか？
Whose performance is it?
(McNamara, 1997, 2000)

相手話者の影響をうけるコミュニケーション能力をどう捉え、どうすれば意味ある、公平な評価ができるか？

やりとり・相互行為能力
(Interactional competence)
(He & Young, 1998: Kasper, 2006)

- ✓ コミュニケーションは二人もしくはそれ以上の参加者によって協働的に構築されるという見方
- ✓ 構成概念としての定義？

14

実用性 (practicality)

✓ 採用するテストのデザインや実施計画は、実際的でなければならない。

実用性に欠けるテスト例：

- o 受験料が高すぎる
- o 他の都市に行かないと受験できない

✓ 実施や採点にかかる時間やコストという点からも、検討すべき

15

影響力 (impact)

波及効果(washback effect)
(Cheng, Watanabe, & Curtis, 2002)

テスト

学習者
カリキュラム
教材内容
指導法

社会やコミュニティーへの影響
(Messick, 1989)

16

到達度テストまたは能力テストとしての会話テスト

日本語を使って何がどの程度できるか？

到達度テスト
Achievement Test
小規模テスト

- 単元テスト
- 期末試験等

✓ カリキュラムに定められた各目標がどの程度達成できたか。

熟達度(能力)テスト
Proficiency Test
大規模テスト

- ACTFL OPI
- Oral Pro… Bilin…
- 日本語会話能力テスト(JOPT)
  http://jopt.jpn.org/

✓どんなことができるか。パフォーマンスの一般的な特徴は何であるか。

(中島 2001)

17

会話能力テストとしてのACTFL OPI

話題にできるトピック？
下限レベルはどの辺か？

（牧野他 2001；鎌田 2008）

導入 Warm up → レベルチェック Level Checks 突き上げ Probes → ロールプレイ Role play → 終結 Wind down

✓挨拶
✓身の回りに関する簡単な質問の受け答え

✓何ができて（下限: floor）、何ができないか（上限: ceiling）を確認するまで質問を繰り返す

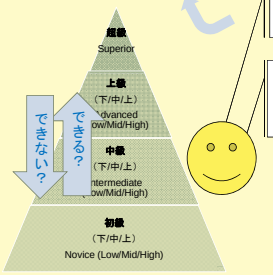
✓ レベル判定の確認

✓簡単な質問

18

会話能力テストとしてのACTFL OPI (cont'd)

受験者の発話に「**言語的足場**」があれば、レベルを下げた質問をする



上級レベルの質問（家庭・学校・余暇活動などに関する話題で描写や叙述を必要とするもの）に、どの程度対応できるか？できないか？

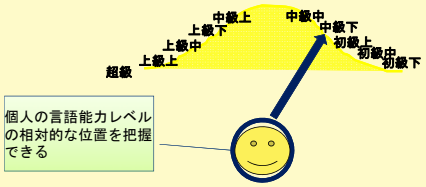
単文や複文を使って、身近な話題に関する質問（中級レベル）に対する応答が、楽に維持できる

先ほど、趣味で時々ピザを作ると言っていましたが、ピザはどんな手順で作るんですか？　作り方を説明してもらえませんか？（上級レベルのプロンプト）

19

会話能力テストとしてのACTFL OPI　(cont'd)

言語能力テストは、受験者全体のテスト結果が、正規分布に近くなるように開発されている (Bachman, 1990; Brown, 2005)



個人の言語能力レベルの相対的な位置を把握できる

20

タスク型会話到達度テスト

実世界タスク は教育用タスクとして調整される
(Norris et al., 1998; Van den Branden, 2006)

実世界タスク
real-world task
(=目標タスク[target task])
何らかの遂行目的を持つ実世界で起こる言語活動

教育用タスク
pedagogic task
学習者のレベルや学習目標に応じて調整される

21

タスク型会話到達度テスト (cont'd)

(近藤ブラウン 2013)

- ✓ 学習者にテストの目的、形式、手順、そして採点方法について、事前に説明しておく
- ✓ クラスで学んだことがテストされる
- ✓ テストで実際に使用されるタスクの指示（prompt）については、学習者は試験開始まで知らされない
- ✓ 採点に使用する評価基準（例. 採点ルーブリック）を事前に手渡しておく

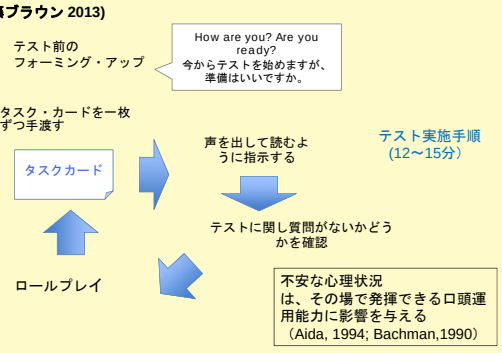
詳細を与えすぎると、現実世界でのパフォーマンスに伴う即興性が失われる　⇒　妥当性に影響

学習者のベストのパフォーマンスを引き出す

22

タスク型会話到達度テスト (cont'd)

(近藤ブラウン 2013)



不安な心理状況は、その場で発揮できる口頭運用能力に影響を与える（Aida, 1994; Bachman,1990）

23

テスト設計図(Design statement)

Bachman & Palmer [1996] のモデルに基づく

テストの目的 (Purpose)	目標言語使用領域とテスト内容・形式の記述 (Description of the TLU domain and test/task types)	テスト受験者の特性 (Characteristics of test takers)
構成概念の定義 (Definition of construct)	テストの有用性検証の計画 (Plan for evaluating the qualities of usefulness)	人材・設備のリストと管理計画 (Inventory of available resources and plan for their allocation & management)

24

1. テストの目的 (Purpos

テスティングとは、推測することである “Testing is about making inferences.” (McNamara, 2000, p. 7)

テストの目的は何か。テスト結果を基に、どのような言語能力の推測・推定をするのか。また、テスト結果をだれが何の目的で使用するのか。

“...specific inferences about language ability...we intend to make on the basis of test results and specific decisions which will be based upon these inferences" (Bachman & Palmer, 1996, p. 88)

JOPT

- ✓ 測定を意図している「会話能力」の定義は？「日本語会話能力」をどうとらえるか？（構成概念の定義にも関連）
- ✓ JOPT及びそのテスト結果を何の目的で使用するのか（診断的・形成的・総括的評価？）

25

2．目標言語使用領域とテスト内容・形式の記述 (Description of the TLU domain and test/task types)

評価・測定を意図している能力をどのようなテストを使って推測するのか。どのような言語使用領域・内容・形式 のテストであるのか。

Statement of "the tasks in the TLU domain to which we want our inferences about language ability to generalize"(Bachman & Palmer, 1996, p. 88)

JOPT

１５分以内で行える対面式テスト（詳細については、現在検討中）

- ✓ 意図している日本語会話能力を、具体的にどのように評価・測定するのか（テストの細目表作成）
- ✓ 会話サンプルの抽出法は、テスト目的や有用性にも深く関係する
- ✓ 超級・上級レベルの会話能力測定をどのようにして１５分以内で行うか
- ✓ テスター資格の条件は？（例. 非母語話者教師の場合、テスターとして最低限必要な言語能力条件は？）

26

3. テスト受験者の特性 (Characteristics of test takers)

どのような日本語使用者を対象にするテストか

“the nature of the population of potential test takers for whom the test is being designed"(Bachman & Palmer, 1996, p. 88)

JOPT

国内外の日本語教育機関のみならず、地域の日本語教室で学ぶ16、17才以上の学習者や定住外国人を対象

- ✓ 測定を意図している会話能力の範囲は？

27

4. 構成概念の定義 (Definition of construct)

テスト目的とも関係するが、言語能力をどうとらえ、どのような能力の推測を意図しているのか。

“the theoretical definition of the construct or particular aspect of language ability...the inferences that can be made from the test scores." (Bachman & Palmer, 1996, p. 88)

JOPT

「会話能力」（JF日本語スタンダード、CEFR, ACTFL Proficiency Guidelines等の基準を踏まえる）

- ✓ JOPTでいう「日本語会話能力」とは何を意味するのか
- ✓ JOPTで測定を意図している「日本語会話能力」は、JF日本語スタンダード（CEFRが土台になっている）で定義されている「コミュニケーション言語能力 (communicative language competence)」及びACTFL能力基準でいう「口頭言語能力 (spoken language ability)」の構成概念をどう踏まえるか

28

5. テストの有用性検証の計画 (Plan for evaluating the qualities of usefulness)

テストの有用性（妥当性・信頼性・真正性・相互作用・影響力・実用性）の検証をだれがどのように行うのか。Considerations for test usefulness & plan for test validation

JOPT

- 日本語教育・評価の専門家がいくつかの班に分かれて、テスト開発及び検証を行う。
- JOPTデータバンクを作成し、検証及びテスター訓練に使用する

- ✓ JOPTで妥当性の高い（意味ある）能力の推測ができているかどうか、また公平で適切な方法でテスト結果が解釈・使用されているかどうかを具体的にどのような方法で検証するのか
- ✓ テスター訓練の具体的計画は？

29

6. 人材・設備のリストと管理計画 (Inventory of available resources and plan for their allocation & management)

計画されたテスト開発や検証を予定通り遂行するに必要な人材・設備は確保できているか。"The resources that will be required and will be available for various activities during test development" (Bachman & Palmer, 1996, p. 89)

JOPT

- テスト開発に必要な人材・設備が整い、その管理計画も考慮されている。

出典：http://jopt.jpn.org/p3.html

30

引用文献

Aida, Y. (1994). Examination of Horowitz, Horowitz, and Cope's construct of foreign language anxiety: The case of students of Japanese. *Modern Language Journal, 78* (2), 155-68.

Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom.* Thousand Oaks, CA: Corwin Press.

Bachman, L. F. (1988). Problems in examining the validity of the ACTFL oral proficiency interview. *Studies in Second Language Acquisition, 10* (2),149-164.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bachman, L. F., & Savignon, S. J. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *The Modern Language Journal, 70* (4), 380-390.

Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing, 20* (1), 1-25.

Brown, J. D. (2005). *Testing in Language Program: A comprehensive guide to English language assessment* (Revised ed.). New York: McGraw-Hill.

Brown, J. D., Hudson, T., Norris, J., & Bonk, W. (2002). *An investigation of second language task-based performance assessments*. Honolulu, HI: National Foreign Language Resource Center, University of Hawai‘i.

Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). (2002). *Washback in language testing: Research contexts and methods.* Mahwah, N.J.: Lawrence Erlbaum and Associates.

Dandonoli, P., & Henning, G. (1990). An investigation of the construct validity of the ACTFL proficiency guidelines and oral interview procedure. *Foreign Language Annals, 23* (1), 11-22.

Davis, L., & Kondo-Brown, K. (2012). Assessing student performance: Types and uses of rubrics. In J.D. Brown (Ed.), *Developing, using, and analyzing rubrics in language assessment with case studies in Asian-Pacific languages* (pp. 33-56). Honolulu, HI: National Foreign Languages Resource Center Publications.

Falsgraf, C. (2009). The ecology of assessment. *Language Teaching, 42* (4), 491-503.

Ellis, R. (2003) *Task-based Language Learning and Teaching.* Oxford: Oxford University Press

He, A. W., & Young, R. (1998). Language proficiency interviews: A discourse approach. In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 1-24). Amsterdam & Philadelphia: John Benjamins.

Johnson, M. (2001). *The art of non-conversation: A re-examination of the validity of the oral proficiency interview.* New Haven, CT: Yale University Press.

Johnson, M., & Tyler, A. (1998). Re-analyzing the OPI: How much does it look like natural conversation? In R. Young & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 27-51). Amsterdam: John Benjamins.

Kasper, G. (2006). Beyond repair: Conversation analysis as an approach to SLA. *AILA Review,* 19 (*1),* 83-99.

Kondo-Brown, K. (2002). An analysis of rater bias with FACETS in measuring Japanese L2 writing performance. *Language Testing, 19*, 1-29.

Lazaraton, A. (2002). *A qualitative approach to the validation of oral language tests.* Cambridge: Cambridge University Press.

Lumley, T., & McNamara, T. F., (1995). Rater characteristics and rater bias: implications for training. *Language Testing, 12*(1), 54-71.

McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.

McNamara, T. F. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics, 18* (4), 446-466.

McNamara, T. F. (2000) *Language testing.* Oxford: Oxford University Press.

McNamara, T. F., & Adams, R. J. (1991, March). *Exploring rater behavior with Rasch Techniques*. Paper presented at the Annual Language Testing Research Colloquium, Princeton, NJ. Eric document ED345 498.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York: Macmillan.

Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments.* Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawai‘i.

Popham, W. J. (2008b). *Transformative Assessment***.** Alexandria, VA: Association for Supervision & Curriculum Development.

Ross, S. J. (2007). A comparative task-in-interaction analysis of OPI backsliding. *Journal of Pragmatics, 39* (11), 2017-2044.

Ross, S. J., & Berwick, R. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition, 14* (2), 159-176.
Salaberry, R. (2000). Revising the revised format of the ACTFL Oral Proficiency Interview. *Language Testing, 17* (3), 289-310.
Shohamy, E. (1995). Performance assessment in language testing. *Annual Review of Applied Linguistics, 15,* 188-211.
Surface, E. A., & Dierdorff, E. C. (2003). Reliability and the ACTFL Oral Proficiency Interview: Reporting indices of interrater consistency and agreement for 19 languages. *Foreign Language Annals, 36* (4), 507-519.
Thompson, I. (1995). A study of interrater reliability of the ACTFL Oral proficiency interview in five European languages: Data from ESL, French, German, Russian, and Spanish. *Foreign Language Annals, 28*(3), 407-422.
Van den Branden, K., Bygate, M., & Norris, J. (2009). Task-based language teaching: Introducing the reader. In K. Van den Branden, M. Bygate, & J. Norris (Eds.), *Task-based language teaching* (pp. 1-13). Amsterdam: John Benjamins.
Van Lier, L., (1989). Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly, 23* (3), 489-508.
Weigle, S. C. (2002). *Assessing writing.* Cambridge: Cambridge University Press.
伊東 祐郎 （2008）『日本語教師のためのテスト作成マニュアル』アルク
鎌田修（2008）「ACTFL-OPIにおける“プロフィシェンシー”の測定」鎌田修・嶋田和子・迫田久美子（編）『プロフィシェンシーを育てる：真の日本語能力をめざして』凡人社, pp. 108-131.
近藤ブラウン(2012)『日本語教師のための評価入門』くろしお出版
近藤ブラウン(2013) 「日本語評価のためのタスク型アチーブメント・テスト」『第二言語としての日本語習得』13, pp. 56-73.
中島和子（2001）「子供を対象とした活用法」牧野成一他（編）『ACTFL OPI 入門』アルク, pp. 152-169.
牧野成一他（2001）（編）『ACTFL OPI 入門』アルク