

教室で行われるテスト問題作成の秘訣

Tips for Item Writing for Classroom-based Tests

真継 左和子

Sawako Matsugu

立教大学

Rikkyo University

本論では、教室で実施される客観テストに使用される問題の作成に必要な3つのステップを紹介する。具体的には、1. カリキュラム全体のゴールや学習行動目標の決定と構成概念の構築を含むテスト準備、2. テスト細目表の作成、3. 問題作成である。評価は授業と密接につながっており、授業は学習行動目標が反映されていなければならない。したがって、学習行動目標なしにテスト作成はできない。さらに、構成概念や授業で学習したユニットやトピックをバランスよく測定するにはテスト細目表が非常に便利である。テスト細目表とは、学習行動目標とトピック等の内容を二方向に1つの表にまとめたものである。最後に、テスト問題を作成するにあたって言語やフォーマット関連等の様々な注意点を紹介する。

This paper introduces three important steps for writing items in objective tests used in classroom-contexts; namely, preparing for the test, which includes writing course goals and objectives and defining the construct, making a table of specifications, and writing items. Because assessment is deeply intertwined with instruction, which reflects pre-determined course goals and objectives, tests cannot be developed without instructional goals and objectives. Furthermore, in order to assess constructs and units/topics covered in the class in a balanced manner, a table of specifications, which is a two-way chart that lists both the instructional objectives and test content, is an important tool. Finally, several tips for developing items are introduced.

はじめに

教室で実施されるテストは、学習者の学習結果だけでなく教員の授業の効果についての情報をも提供できる。さらに、それらのテスト(例えば小テスト、定期試験)の結果は授業の成績の大部分を占めることが多い (Gullickson, 1984)。このようなテストはthe Test of English as a Foreign Language (TOEFL) やthe Test of English for International Communication (TOEIC)、International English Language Test System (IELTS) のように重要な波及効果を持たないかもしれない。しかし、多くの学習者は自身の成績に多大な関心を持っており、例えば留学候補生の審査、インターンシップや奨学金の申し込み等、成績が重要な選考に利用されることもある (Brown, 1992)。ゆえに、たとえ教室で実施される日常のテストであっても高い質であることが求められる。

しかし、英語教員は必ずしもテスト作成、特に問題を作るということに関しての研修を受けているわけではない。Teaching English to Speakers of Other Languages (TESOL) や応用言語学の大学院プログラムにおいても、

テストの授業が開講されていなかったり (Cizek, Fitzgerald, & Rachor, 1995)、開講されていても必修でなかったりするので (Bachman, 2000)、Brown and Bailey (2008) は実際にそれらのプログラムの半分程度しかテストの授業を必修としていないと報告している。その結果、修了してもテストに関しての知識や経験が十分でないまま教壇に立つことがある。さらに、日本においては、英語教員の専門はTESOLからコミュニケーション学、文学、言語学等多岐に渡ることから、先に述べた状況を踏まえると教員のテスト作成に関しての知識も様々だと推測できるのではないだろうか。学習者は公正に評価を受ける権利があるので、英語教員の専門に関わらず、テストの作成知識は必須である。

テスト問題の質が悪い場合、測定しようとしたものが測定できないだけでなく、受験者(学習者)にとっても混乱や苛立ちをもたらす。それに加えて、構成概念 (construct) と無関係の分散 (construct irrelevancy) や構成概念の代表性欠如 (construct underrepresentation) が生じる可能性があり (Downing, 2002, 2006)、どちらもテストの信頼性と妥当性の脅威となる。ここでいうところの構成概念とは測定しようとする能力の詳細な定義である (Bachman & Palmer, 1996)。構成概念と無関係の分散とは、本来測定したいものがそれ以外の要因によって阻害されることを意味する。問いの指示が不明瞭なため不正解になるのはその一例である。また構成概念の代表性欠如とは、測定したい構成概念と関連する問題数が少ないため、受験者の正答が偶然によるものかそうでないのかの判断ができない。仮に正答しても受験者のその構成概念の理解を示すのかが判断ができないことである。そして、そのようなテストは学習者の到達度に関する情報を適切に提供できないばかりか、教員はそのようなテストから得られた結果をティーチングに生かすこともできない。

本論では、テスト作成の初心者进行を想定し、伝統的なペンと紙のテスト(小テスト定期試験)を対象とした問題作成方法の秘訣、特に客観テスト (objective test) の多選択式問題を中心に紹介したい。まず、テスト作成前の重要な準備、すなわち、コースのゴールや学習行動目標、さらに構成概念の設定について述べる。次に、テスト細目表の作成方法、そして利用方法を紹介する。最後に、客観テストでよく使用される問題の種類の紹介と、多肢選択式問題作成における様々な注意点を述べる。

テストの準備

学習行動目標とゴールの設定

テストの詳細を決める前に、初めに考慮すべきことがある。担当授業の詳細が分かり次第、その授業がカリキュラム上どのような位置付けになっているかを理解することである。例えば、必修・選択の別、学年、そしてそのクラス内容である。次に、もし各教員が担当クラスの学習行動目標

やゴールを定める必要があるのならばそれを決め、すでにそれらが決まっている場合はそれらをよく理解する必要がある。この場合、「ゴール (goals)」とはカリキュラムの一般的な目的の記述であるが、「学習行動目標」(objectives)はそのさらに詳細な記述を意味する (Richards, 2001)。Richardsが述べているように、クラスの学習行動目標は学習結果がよく反映されるよう可能な限り詳細に定められていなければならない。そして、その学習結果は授業にも評価にも反映されうるものとなる。すなわち、学習行動目標とは、教科書を選定することや教材を吟味する前に、考えられるべきものであることを忘れてはならない。よってテスト作成の際にはクラスの学習行動目標、そのユニットと学習行動目標との関係、そしてテストが測定すべきもの等をしっかりと考慮することが望ましい。

構成概念の構築

一旦担当クラスの様々な学習行動目標が定められたら、それら学習行動目標をどのように達成すべきか考慮し、適切なテキストや教材、トピックを選ぶ。次に、構成概念をテストのために定義し、決める必要がある。例えばリスニングの構成概念なら、リスニングのどのスキルをどのように測定するのか等を定めなければならない (Bachman & Palmer, 2010)。この場合、具体的には、リスニングのトップダウン処理の一部としてパッセージの要旨の理解を後で述べる多肢選択式問題を使って測定する、等が挙げられる。構成概念の構築後、クラスの学習行動目標とゴール、それに対応するユニットやトピックの関係は以下に述べるテスト細目表 (table of specifications) を作成するにあたって重要になる。

テスト細目表

Miller, Lin, Gronlund (2009)によると、テスト細目表とは、1つの表にそのテストが測定する学習行動目標と、授業で学習したユニットやトピックをそれぞれの目標に呼応するようにまとめたものである。この表を作成することにより、まず学習行動目標が満遍なく測定でき、また、測定にあたり適切なサンプル数の問題を可能な限り出題することができる。さらに、テストで出題されるユニットやトピックの意図しない偏りを選避することができる。テスト作成の際には、出題される問題が学習行動目標を測定するのに十分なサンプル数があり、また、問いがそれぞれの目標を測定するにあたっての代表的なサンプルである、ということがそのテストの妥当性の保証にもつながる。したがって、テスト細目表を作成することはテストを準備する際の大変重要なステップとなる。

付録1では、リーディングのテスト細目表の例が掲載されているので参照されたい。まず表の一番上の行を見るとそのテストで測定される学習行動目標が簡略化されて列挙されている。この例ではリーディングテストのため、main idea (要旨)、detail (詳細)、inference (推測)、prediction (予測)の理解が測定されている。次に、表の一番左の列を見ると出題されているユニットの分野が列挙されており、さらにそれぞれの出題パッセージの語数とリーダビリティのレベルも書かれている。そして、ひとつ右の列を見てみると、各パッセージと学習行動目標に対応する問題番号、問題数、配点が書かれている。このようにテスト細目表は縦横に確認しながら利用する。右の端3列は左から順にパッセージごとの総問題数、合計点、そしてテスト全体に占める割合をパーセントで示している。同様に、一番下の3行も上

から順に測定される目的ごとに総問題数、合計点、そしてテスト全体に占める割合をパーセントで示している。

細目表を埋める際に注意すべきことは、測定したい学習行動目標やユニットを書き込んだ後、テスト全体を占めるそれぞれの大きなパーセントをまず決めるということである。例えば、複数のユニットを学習した場合、かけた時間がそれぞれのユニットに等しくないのであれば、出題数(パーセント)もその時間に呼応して出すべきである。同様に、授業で仮に要旨の理解に最も時間をかけた場合や、それが最も学習者にマスターしてほしい目的の場合は、相応に出題数を多くすればよい (Jamieson, personal communication, September 23, 2008)。このような理由から、ユニットや学習行動目標別に大きなパーセントを先に決めた方がよい。次に、決定したパーセントと対応するようにそれぞれの目標やユニット別に問題数を決め、問題の種類を選び、実際に問題を書いていく。最後に、できるだけ多くのサンプル(情報)を得るため、問題数は多ければ多い方がよいのは言うまでもないが、各目標に対して少なくとも10問出題できれば理想的である。ただし、問題が以下で述べる学習者自らが補充するタイプの場合や、タスクそのものが限られている場合は、5問程度にすることも可能である (Miller et al., 2009)。

問題の種類と作成の注意点

問題の種類: 補充問題と選択問題

問題の種類を選択にあたって、大前提は「測定しようとする学習結果を最も直接的に測定できる問題を選ぶ」ということである (Miller et al., 2009)。筆記試験でよく出題され、本論で扱う問いは大きく2つに分けられる。補充問題 (supply types) と選択問題 (selection types) である。補充問題には短文解答、空欄・空所補充があり、選択問題には組み合わせ法、正誤、多肢選択式等がある。前者は測定しようとする学習結果が筆記や列挙すること、または問われたものの名前を挙げることを必要とする場合に最適である。後者はタスクが正答を選ぶことを必要とする場合に利用するとよい。どちらのタイプの問題を出題すべきか迷う場合は解答のコントロールに優れ、採点の客観性が保証される選択問題が勧められる (Miller et al., 2009)。ただし、選択問題は学習者に言語をアウトプットすることを全く要求しないので、出題する際はその点の考慮が必要である (Brown & Hudson, 2002)。

どちらのタイプの問題も全て客観テストに利用され、知識や事実を理解しているかを測定するのに最適である。また、これらの問題の作成には時間がかかり、先に述べたように多くの問題数が必要である。しかし、学習者の解答をかなりコントロールすることができる上、採点も早く効率的である。また、信頼性にも優れている。特に、多肢選択式問題は、工夫次第で学習した内容の理解から、思考力、そして上位の思考技術 (higher order thinking skills) も測定することが可能である (Miller et al., 2009)。

問題作成にあたっての注意点

問題の表現に関するもの

1. 学習者のレベルの言葉で書かれている
2. 否定や二重否定を含まない
3. 問題や問いの指示が明確である

Poor Ex) "Parents play an important role in life"

although they sometimes complicate matters for their children. This refers to ____.”

- Parents' role in life?
- Complicating matters?

1は問題の内容が学習者のレベルより高い場合、誤答の際の原因が特定できないためである。つまり、問いが理解できなかったからなのか、それとも問われている知識や理解力がなかったからなのか不明である。2は否定や二重否定は学習者に混乱を起しやすく、問いを理解するのも難しくなるからである。3は問われているものが明らかでなければ、複数の解答が正解になりうるからである (Brown & Hudson, 2002)。

問題のフォーマットに関するもの

1. 問題や指示が(解答に)必要な情報のみを含んでいる

Poor Ex) The following eight vocabulary words have been selected from the reading passage in Unit 5 of the course reader. Your teacher discussed these words in the class last week . . .

Good Ex) Choose two words from Unit 5 below.

このように問題が解答に際して不必要な情報を含んでいると、学習者に不必要に負担をかけるからである。

2. 各問題が独立している

Poor Ex 1) What is the square root of 64?

Multiply this by 9.

Poor Ex 2) Who are the three most famous classical music composers in Germany?

When did Beethoven move from Germany to Austria?

これは誤答の際の原因の推測が難しくなるからである。さらに、問題が別の問題の解答にヒントを提供するようなことも避けた方がよい。

3. 問題が見やすく整理できている

Poor Ex) The Olympics will be held ____.

- a. in Tokyo in 2024
- b. in Tokyo in 2016
- c. in Tokyo in 2020

Good Ex) The Olympics will be held in Tokyo in ____.

- a. 2016
- b. 2020
- c. 2024

3については、可能な限り不必要な負担を学習者にかけないよう、テスト作成の際に実践できる事柄ばかりである。例えば、選択肢に重複する語があれば、それを問いの

幹に含めて読みやすく、見やすくできる (Brown, 2004)。また、同じタイプの問題でセクションを分け、解答する際に学習者があちこちに解答の根拠となる箇所を探さなくてよいように整理することもできる。例えばリーディングテストならば出題パッセージに書かれた情報の順番に沿って出題すべきである。また、いうまでもなく適切に余白や行間を使い、テスト全体が美しいレイアウトとなるべきである。見やすいフォントを使うことも忘れてはならない (Brown & Hudson, 2002)。

4. 選択肢が文法的によく整理されている

Poor Ex.1) Which city is the state capital of Illinois?

- a. San Francisco
- b. Boston
- c. New York
- d. The city that is famous for its strong winds

Good Ex.1) Which city is the state capital of Illinois?

- a. San Francisco
- b. Boston
- c. New York
- d. Chicago

この例は、選択肢4つのうち3つは具体的な都市名なのに対して、1つだけ都市の描写になっており、学習者に不必要なヒントを与えてしまう。基本的に、選択肢の種類(品詞等)や長さは同じぐらいにする方が上と同じ理由でよい。

Poor Ex.2) According to the passage, Emily is less than ____ feet tall.

- a. 3
- b. 4
- c. 5
- d. 6

Good Ex. 2) According to the passage, Emily is ____ feet tall.

- a. 3
- b. 4
- c. 5
- d. 6

この例は、問いが身長「～feet以下」となっており、賢い学習者なら、一番大きい数のd(6フィート)を選んでおけば安心、となる。

Poor Ex. 3) Naoki saw a ____.

- a. eagle
- b. elephant
- c. zebra
- d. orangutan

Good Ex. 3) Naoki saw a/an ____.

- a. eagle
- b. elephant
- c. zebra
- d. orangutan

3つ目の例は、下線部の前に冠詞の「a」があるので、選択肢のa, b, dは母音から始まるため、選択肢を読まずに正答にたどり着くことができる (Brown & Hudson, 2002)。

その他避けるべき問題

選択肢に“none of the above,” “A and B, but not C,” “all of the above”のようなものは極力避ける。

これらが選択肢に含まれている場合、誤答の際の理由が判明しないからである。また、学習者が解答にたどり着くまで多くの段階を踏まなければならない、不必要な負担をかけることになる (Brown & Hudson, 2002)。

テスト作成準備について

1. テスト細目表をガイドとして利用する
2. 必要な問題数より多く問題を作成する
3. 日程に余裕を持って問題を作成する

1については、問題作成中にテスト細目表をガイドとして参照することによって、バランスのとれたテストを作成することができる。また測定すべき構成概念を測定し忘れるといったことも防げる。2は、多肢選択問題は必然的に良い問いとそうでないものができるため、必要以上に作成することによって、良い問いを選び出題することができる。3は、どのようなテストにも当てはまるが、一旦問題を作成した後、数日後に見直せば改善すべき問題が見つかることが多い。よって、時間に余裕を持って作成することによってより質の高い問題にすることができる (Miller et al., 2009)。

最後に

本論で紹介した、テスト作成準備、テスト細目表、客観テストでよく出題される問題作成の注意点等は決して難しいものではなく、誰もが今日から始められるものである。特に、テスト細目表は一見表の作成自体に時間がかかるように思えるが、作成に慣れれば簡略化したものを利用することもでき、テスト細目表を利用することによって効率よくテストを作成できる。

テストを作成する教員の少しの心がけや努力でより良い問題を作ることができ、その結果適切に、そして正確に測定したいものを測定することができる。良いテストから得られた情報は、学習者の今後の学習に生かすことができ、教員にとっても今後の授業に役立たせることが可能である。1人でも多くの学習者がより良いテストを受験できることを願ってやまない。

謝辞

本稿の完成にあたり、多くのご助言を下された査読委員の先生方と、夫で研究協力者のジェームズ・カーペンターにはこの場を借りてお礼を申し上げます。

参考文献

- Bachman, L. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17, 1-42.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- Brown, J. D. (1992). Classroom-centered language testing. *TESOL Journal*, 1, 12-15.
- Brown, J. D., & Bailey, K.M. (2008). Language testing courses: What are they in 2007? *Language Testing*, 25, 349-383.
- Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge: Cambridge University Press.
- Cizek, G., Fitzgerald, S., & Rachor, R. (1995). Teachers' assessment practices: Preparation isolation, and the kitchen sink. *Educational Measurement*, 3, 159-179.
- Downing, S. (2002). Construct-irrelevant variance and flawed test questions: Do multiple-choice item-writing principles make any difference? *Academic Medicine*, 77, S103-S104.
- Downing, S. (2006). Twelve steps for effective test development. In S. Downing, & T. Haladyna (Eds.), *Handbook of test development* (pp. 3-25). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.
- Gullickson, A. (1984). Teacher perspectives of their instructional use of tests. *Journal of Educational Research*, 77, 244-248.
- Miller, D. M., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Pearson Education.
- Richards, J. (2001). *Curriculum development in language teaching*. Cambridge: Cambridge University Press.

真継左和子: 北アリゾナ大学博士課程修了。応用言語学博士。専門は言語テストング。特に、スピーキング評価、採点者の経歴がどのように評価に影響を及ぼすかに興味がある。また、プロジェクトベースラーニングの評価も最近研究を開始した。関西地方の大学で非常勤講師、アメリカで博士課程在籍中に大学の付属語学学校で評価コーディネーター・教員として勤務し、現在は立教大学教育講師。



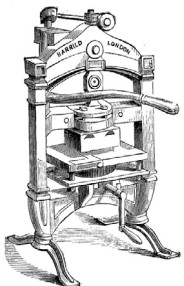
Sawako Matsugu has a PhD in applied linguistics from Northern Arizona University. She has taught EFL in both the Kansai and Kanto areas in Japan. While she was pursuing her PhD, she worked as an assessment coordinator and instructor at an intensive English program. Her research interests include language assessment, in particular speaking assessment and the effects of rater background on assessment, and assessment on project-based learning. She is currently teaching at Rikkyo University.

付録1

Table of Specifications (Reading)

Objective Text type		Main idea	Detail	Inference	Prediction	Total items	Total points	%
Economy 300 words Flesch-Kincaid Grade Level 12.0	Item #	2.1.1	2.1.3, 2.1.4, 2.1.5	2.1.2	-	5	11	22%
	# of items	1	3	1	-			
	points	3	6	2	-			
Science 250 words Flesh-Kincaid Grade Level 14.9	Item #	2.2.1	2.2.2, 2.2.3	2.2.4, 2.2.5	-	5	11	22%
	# of items	1	2	2	-			
	Points	3	4	4	-			
Culture 450 words Flesch-Kincaid Grade Level 11.7	Item #	2.3.1	2.3.3, 2.3.4, 2.3.5	2.3.2, 2.3.6 2.3.7	-	7	15	30%
	# of items	1	3	3	-			
	Points	3	6	6	-			
Refugees 500 words Flesch-Kincaid Grade Level 11.9	Item #	2.4.1,	2.4.3, 2.4.4	2.4.2, 2.4.5	2.4.6	6	13	26%
	# of items	1	2	2	1			
	points	3	4	4	2			
Total items		4	10	8	1	23		
Total points		12	20	16	2		50	
%		24%	40%	32%	4%			100%

The Language Teacher needs you!



If you are interested in writing and editing, have experience in language education in an Asian context, and are a JALT member, we need your help. *TLT* is currently recruiting proofreading and editorial staff.

Learn a new skill, help others, strengthen your résumé, and make a difference! If you would like to join our team, please contact the editors:

<tlt-editor@jalt-publications.org>