

# 現職日本語教師研修のための聴解授業の評価

## 聴解評価基準について

西村 よしみ

### 要 旨

2005年9月、京畿道外国語教育研修院（以下研修院）との初めての京畿道日本語教師の再教育が始まった。留学生センターと研修院との学术交流協定の締結に至る交渉期間を経て、教材開発から続く研修実施への行程には多くの困難があり、研修が終わった時には達成感があった。しかし、同時に研修実施の過程で多くの課題が残されているのも事実である。

筆者は、聴解担当講師として2週間の研修を行った。まず、今回の研修で再考を要すると感じた課題は「評価」である。果たして、妥当性と信頼性のある評価が出来たのであろうかという疑問が残る。一つは、英語・中国語・日本語の現職教師を対象とした再教育を実施している研修院側の評価の共通枠についてであり、二つめは、評価の64%をしめる「必須科目」の評価尺度の記述についての問いである。特に、この研修評価が研修生にとって、高い社会的責任を持つものだけに、次年度に向け、検討と改善が必要だと考える。本報告では、研修院の作成した評価の共通枠について概観し、聴解の授業における評価尺度の記述について検討を行う。

【キーワード】社会的責任 high-stakes 評価 評価尺度 ヨーロッパ共通参照枠

## A Report on the Overall Evaluation and Evaluation of Listening Classroom

NISHIMURA yoshimi

【Abstract】 In September 2005, the first In-service Japanese Language Training Program run by the International Student Center, University of Tsukuba, and Gyeonggi-do Institute for Foreign Language Education (hereafter called GIFLE). This program achieved some notable results, but at the same time many questions remain. Whether the evaluation of this program was appropriate still remains an outstanding question. There are two important topics to be discussed, one is the common framework of total evaluation which was made by GIFLE, and the second one is the evaluation rates of Listening Classroom which the author made. This evaluation has high social responsibility for the trainee, so we should re-examine and improve this evaluation.

In this report, the author will overview the general framework and discuss the rate of evaluation.

【Keywords】 social responsibility, high-stake, evaluation, evaluation rates, Common European Framework of Reference for Languages

## 1. はじめに

評価は、「社会的責任」を負うものである。今回の研修の総合評価は、まず成績上位者は教頭資格試験に有利に働くだけでなく、この得点が職場での評価においても参考にされると聞く。さらに、成績上位者の12人は、筑波大学留学生センターで行われる1ヶ月の研修に参加できる権利を有する。研修2週目の後半に、クラスリーダーと筑波からの講師、研修院の日本語担当者間で、評価、最終テスト、宿題についての改善要求の話合いが持たれた。リーダーからは、点数での評価の代わりにA,B,Cなどの評価、最終テストの意義を問う意見が出てきた。また、日本語だけでなくITの研修などの宿題も負担が多すぎるという状況も知った。その場での、我々の回答は、評価は研修院の方針に従うこと、最終テストは計画通り行うこと、宿題については軽減を検討する等を伝えた。その話合いや筆者が滞在していた2週間を通じ、每晚1時2時まで必死に授業の予習・復習を行い、宿題を提出するという状況を見聞きし、改めて今回の研修評価は、「社会的責任」を果たす妥当性や信頼性を持ったものであったのかという疑問が生じた。

筑波大学留学生センターでは、プレースメントテストや、期末の成績評価など比較的low-stakesな評価が中心である。多少の配置の間違いに対しても、採点開始後の調整が可能である等、柔軟に対応している。また、指導教官に成績を報告しても、それが大学院入試結果を左右するという事は聞いていない。

マクナマラ(2004)は、アセスメントは、個人的な立場で行うものではなく、機関の役割においてその目的を果たすため行うものであるとし、「教育においても、他の分野においても、社会的な政策目標の達成を目指している。そして、社会的に実践することにより社会的責任という問題も生じる。」と記している。研修院が、京畿道内の現職日本語教師を5年間で再教育し、道内の中等教育の外国語教育の水準を上げようという政策目標の下にこのプログラムは始まったのである。したがって、研修の評価は、研修生のその後の教育現場の中での昇進や査定に大きな影響を与えることになる。そのために授業だけでなく教室外の予習復習、課題提出を必死に頑張ったのであろう。病人を出すほど厳しい研修になった理由は、単に評価だけの問題ではない。教材の量、各科目間における宿題のバランスの調整がなかったこと、更には研修コース全体のカリキュラムの重さなど、次年度に向け見直しが必要となるところである。しかし、評価は、研修のカリキュラム、シラバス、課題などと複雑に絡み合っている。

そこで、本報告では、最初に研修院の評価システムを概観し、その後筆者の担当である聴解の評価の能力記述について報告する。

## 2. 研修の評価システム

最初に、研修院の総合評価について検討する。全体の評価を表1に示し、各領域について簡単に説明する。

表1. 評価領域と配点

領域	配点	評価内容	備考
必須教科	64%	講師授業内容	観察・発表及び課題
プロジェクト	10%	模擬授業	発表及び資料
選択科目	9%	選択科目 授業内容	修行評価(グループ別)
会話能力評価	10%	ビデオ活用を通じた学習内容	講師とのインタビュー
出席率	6%	出席状況	欠席ごとに減点
研修協力貢献度	1%	研修協力貢献度(リーダーシップ)	全体代表(1点) クラスリーダー(0.5点) 選択科目班長(0.3点)

「必須教科」は、筑波大学で作成した教材による授業科目で、その内容は「文法と誤用分析」「会話・作文」「特集 日本語を楽しむ」「読解」「聴解」「教授法」の5科目である。各評価は、授業担当者が行った。ただ、受容技能については主観的評価だけに頼るのを避け、「文法・誤用分析」「読解」「聴解」の3つは、最後に客観テスト(3科目で60分)を行い評価の参考にした。

「プロジェクト」というのは、「教授法」の中での模擬授業である。これは、当初日本語で模擬授業を行って欲しいという希望が研修院からあったが、京畿道の高校では韓国語を使っている授業がほとんどであり、日本語での授業についてはその必要性がないことから、韓国語での模擬授業を行い、その評価は、教授法の担当者が行った。

「選択科目」は、各講師が「ドラマ(桃太郎)」「新聞作成」「ニュースステーション」「会話教材作成」を担当・助言を行い、研修生が自主的に参加・作成を行った。「選択科目」についての評価は、合格あるいは不合格のいずれかで、合格の場合には9%の点数を得る。表2が選択科目の評価基準で、今回は全員が合格した。

「会話能力評価」は、総合評価全体の10%を占めている。これは、自習教材という筑波側が選択したビデオ教材を5本を授業外に視聴し、その中から1本を選び内容理解に関する質疑応答によって評価するものと、評価者が作成した問に答えるというインタビュー形式で会話能力を評価するというのが研修院の意図であった。

しかし、ビデオも聴解の授業で学習するものより内容は難しく、また授業外に5本見て内容を把握する時間的余裕もないことから、文字化資料が用意されている1本を選び、その内容理解を問うことと、インタビュー形式のもので評価を行った。

表 2 . 選 択 科 目 の 評 価 基 準

区 分		内 容	点数
Pass	三項の内二項を満たしている	1) 授業活動に参加 2) 発表に参加 3) 意見共有及び討議に参加	9
Fail	三項の内二項を満たしている	1) 授業活動に参加不足 2) 同僚研修生との協同性に欠ける 3) 選択教科授業に 2 時間以上不参加	5.4
研修協力貢献度		選択科目班長加算点	0.3

日本語教師に課せられた評価は、表 1 における必須科目と会話能力の評価である。研修生の必須教科と会話能力の評価点数を研修院に提出し、研修院の日本語担当者が総得点を出すという流れで評価が行われた。そして、その評価点数を 80 点から 100 点の範囲で正規分布するように配列しなおす (表 3)。なお、この。総合評価で 1 位のものには、100,000 ウォンの賞金が与えられ、上位 3 名には教頭昇任のためには 3 つの研修を受けなければならないという条件の内 1 つが免除されるということである。

表 3 . 総 得 点 の 点 数 分 布

点数	100	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80
人数	1	1	1	1	2	2	2	3	3	4	8	4	3	3	2	2	2	1	1	1	1

表 4 同 点 処 理

なお、総得点の点数分布を行う場合の同点処理として表 4 のような順位決定を行うという基準も示されている。

1 順位	必須教科の点数上位者
2 順位	会話能力評価の点数上位者
3 順位	プロジェクトの点数上位者
4 順位	選択科目の点数上位者
5 順位	出席率
6 順位	研修協力貢献度 (リーダーシップ)
7 順位	年配者

このように、総合評価については非常に緻密な評価枠が出来上がっている。これは、研修院が京畿道の英語教師を対象とする研修の総合評価と同じもので、今回の日本語研修と同時期に行われた中国語現職者教育においても同じ評価枠を使っている。

だが、ここで問題になるのは、まず総合評価の 64%を占める「必須教科」の評価基準の記述である。英語、日本語、中国語教師の言語的能力記述、学習参加度等の評価を行うのであれば、コンテキストの制約を受けない共通の尺度の作成が必要であろう。しかし同時に、個別的コンテキストも反映できる記述が求められる。つまり、コンテキストに依存しないものとコンテキストと関連しているものを統合する評価の記述が必要となるのである。3つの言語に対応する共通の記述尺度を検討するのは、我々の筑波の日本語担当講師の仕事ではない。だが、日本語に限って考えても、教材内容や評価者などの制約を受けず、なおかつ個別のコンテキストを反映する評価記述を作成するのは非常に困難なことである。しかし、high-stakes な評価を受ける研修生のためにも、評価の信頼性、妥当性を高める努力を怠ってはならないのではないかと考える。次章では、筆者が担当した聴解授業の評価記述について検討したい。

### 3. 聴解授業の評価

「必須教科」の64%の内訳は、「文法と誤用分析」「会話・作文」「読解」「聴解」を各11%ずつ、「教授法」と「ことば遊び」をそれぞれ10%とした。11%というのは評価しにくいので110点を満点として表5のような評価の記述を授業担当者である筆者が授業を行う前に作成し研修院の評価担当者と検討したものである。

表5．聴解評価項目

評価項目	点数	使用言語領域
言語的内容	41-50 (A)	<ul style="list-style-type: none"> <li>母語話者にかなり速いスピードで話されていても、生であれ、放送であれどんな種類の話し言葉も難なく理解できる。</li> <li>事柄に対する話し手の言外の意図や話し手の間の関係など細かい点が聞き取れる。</li> <li>省略された文や日常言語の縮約形など十分に理解できる。</li> <li>聴解テストの合格基準に達しており、成績が優秀な人。</li> </ul>
	16-40 (B)	<ul style="list-style-type: none"> <li>短い物語や日常出会う場面での事柄について、明瞭で標準的に話された物であれば要点を理解できる。</li> <li>繰り返し音声素材を聞けば、事柄に対する話し手の言外の意図、人間関係などを理解できる。</li> <li>教師の説明、ワークシートなどにより、必要な情報、主要な点を聞き取り理解できる。</li> <li>省略された文や日常会話の縮約形など教師の説明を聞けば理解できる。</li> <li>聴解テストの合格点に達している。</li> </ul>

	0-15 (C)	<ul style="list-style-type: none"> <li>録音された音声素材を繰り返し聞けば、大部分の情報主要な点は理解できる。</li> <li>文字化資料を参考にして音声素材を聞くことで、必要な情報を理解することができる。</li> <li>聴解テストの合格基準に達していない。</li> </ul>
	Total in this section: 50	
授業における参加度	21-30	<ul style="list-style-type: none"> <li>日本語を使って、授業の中で積極的に発言し、自主的に参加することができる。</li> <li>グループ作業において積極的に参加することができる。</li> <li>個人作業の取り組みが積極的である。</li> </ul>
	11-20	<ul style="list-style-type: none"> <li>日本語で時々、発言することができる。</li> <li>グループ作業において、発言を求められれば参加する。</li> <li>教師の発言を理解し、推測して行動することができる。</li> <li>授業内で理解できなかったことや出来なかったことを時間外にやり遂げることができる。</li> </ul>
	0-10	<ul style="list-style-type: none"> <li>授業の中で、ほとんど発表することがない。</li> <li>グループ作業や個人作業において消極的である。</li> <li>日本語で話された言葉を理解し、行動することがほとんどできない。</li> </ul>
	Total in this section: 30	
課題達成度	21-30	<ul style="list-style-type: none"> <li>予習や宿題や授業での課題を積極的かつ完璧に行うことができる。</li> <li>学習したこと、教師の意図を理解し、発表することができる。</li> <li>与えられた課題を期限内に提出できる。</li> </ul>
	11-21	<ul style="list-style-type: none"> <li>予習や課題などを出さないこともある。</li> <li>予習・復習や課題を教師の指示があれば提出する。</li> <li>授業での課題は、教師の指示がなければ発表することがない。</li> </ul>
	0-10	<ul style="list-style-type: none"> <li>発表することに興味を示さない。</li> <li>授業でやったことを発表することまでに至らない。</li> <li>宿題や課題を充分に行うことができない。</li> </ul>
	Total in this section: 30	
Total of all sections: 110		

「言語的内容」(50点満点)という領域に関して、聴解は聴解特有の評価の難しさがあることに気づいた。いわゆる、受容技能といわれる「読解」「聴解」は、信頼性、妥当性の高いテスト項目さえ作成できれば、一応客観的テストで「読解力」「聴解力」を測定することはできるであろう。しかし、授業内では、彼らの能力を果たして表5のような能力記述で評価できるのであろうか。授業の流れを細かいタスク課題のように区切り、各タスクの理解を確認していけば、可能かもしれない。また、授業を行いながら研修生の中で、特に聴解能力が優れている人、逆に聴解能力が劣っている人は推測できるが、その中間に位置する研修生の能力を見極めることは非常に難しい。筆者は、授業中の研修生を観察するために便宜上、50-41までを3段階に分けてA+、A、A-とし、40-16までと15-0までを同様に3段階に分け、各

授業時間の終了後に記録した。しかし、結果を見ると、B の領域に集中し、その中での細分化は不可能であった。つまり、B の使用言語領域の項目である「繰り返し音声素材を聞けば、事柄に対する話し手の言外の意図、人間関係などを理解できる。」「教師の説明、ワークシートなどにより、必要な情報、主要な点を聞き取り理解できる。」「省略された文や日常会話の縮約形など教師の説明を聞けば理解できる。」などは、授業の目標そのものである。授業中に、小さくパートに分けられたビデオを少なくとも2回は視聴し、ワークシートを見ながら講師は説明をし、その過程で明らかに分かるのは、A+ のものと C、C- のものであり、約 80% の研修生は B として採点する他はなかった。

聴解の評価項目を作成する時、参考にしたのはヨーロッパ共通参照枠（2004）である。これは、the Council of Europe がヨーロッパの言語教育のシラバス、テスト、テキスト等を共通化するために開発したもので、学習者が言語伝達的手段として言語を用いるとき、学ぶべき知識、場面、技能、タスク、方略などを記述しているものである。「学習行程」区間を以下の図のように6つのレベルに分け、柔軟性、一貫性も持たし、評価者にとっては評価しやすいものを作成している。

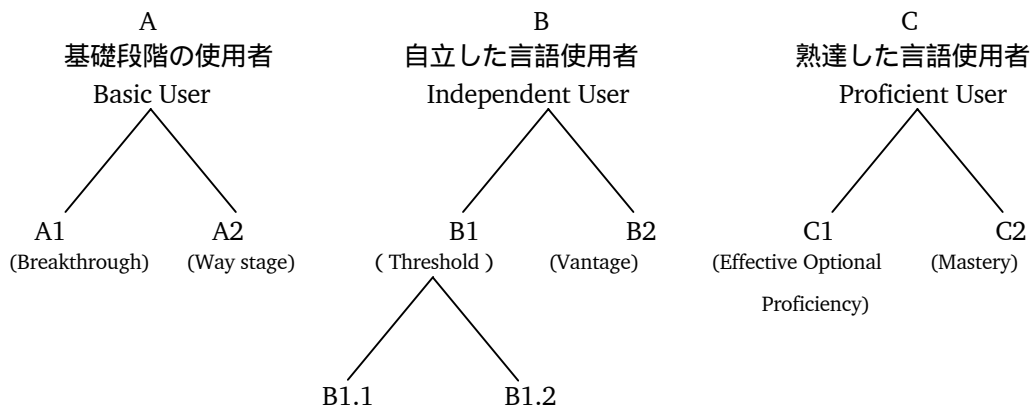


図1 . ヨーロッパ共通参照枠の6つのレベル

B レベルの判定が困難になったようなときには、それをさらに枝分かれさせる「枝分かれ方式 (hypertext)」を採用し、柔軟性を確保している。各基礎段階の言語使用者、自立した言語使用者、熟達した言語使用者についての全体的尺度として詳細な記述が行われているが、本報告では省略するが、ツリーの下位分類について簡単に触れる。

表6. レベルA2.1とA2.2 聴解

A2	A2.1	もしはっきり、ゆっくり発音されるならば、具体的な用件に充分応じることが出来る理解力がある。
	A2.2	もしはっきり、ゆっくり発音されるならば、直接に重要な領域(例:ごく基本的な個人や家族の情報、買い物、地元の地理、仕事)に関する語句や表現が理解できる。

レベル別の分割点(cut-point)の設定は主観的である。ヨーロッパ共通参照枠では、このような枝分かれ方式により言語能力の記述を下位分類し、数字を付けることでそれぞれの機関、組織が重要とする枝分かれを作成し、適切なきめ細かい評価を行うことができるとしている。

だが、枝分かれ方式により、言語能力の下位分類を行っても、4週間という短期間のうちに受容技能の評価ができるのかという疑問はまだ残る。

共通参照枠の聴解の評価利用者のために、以下のようなアドバイスが記されている。

#### 聞き取りについて

- どのようなインプットの処理を学習者は学習する必要があるか。
- どのようなインプットの処理を学習者に身につけさせるか。
- どのようなインプットの処理が出来れば学習者は評価できるのか。

学習者はどのような目的でインプットを聞くのか。

学習者はどのような聞き取りの仕方をするのか。

今回、聴解の評価基準は曖昧で、B1-1のレベルまで降りて記述することはしなかったことは大きな反省点である。そのために、授業で学習した内容に対する客観テストを行い、評価の参考にせざるを得なかった。上述のアドバイスを教材内容と関連づけて検討し、研修生の努力が報いられるような評価システムの構築が反省とともに必要だと考える。

#### 4. 今後の課題

評価領域の3つの内、上述の「言語内容」だけでなく「授業参加度」「課題達成度」についても、課題が多く残されている。

まず、「授業参加度」であるが、今回の教材作成については、統合シラバスを基本軸としているので、クラスによっては「読解」の授業でトピック内容に関する知識が既に学習されていることがあり、「グループ作業において積極的に参加することができる。」など評価する必要がないものである。発言に関しても、研修生は皆日本語を使い、積極的に参加している人が多いのである。

次に、「課題達成度」は、前述したように必須科目間の宿題に対する調整ができていなか



ったり、選択科目、自習教材の課題も多くあったりして、研修生に多くの負担を与えた。そのためか、聴解の宿題の内容は何人かで同じものが提出されていた。共同で宿題を行うということは決して悪いことではないのだが、評価の対象にはできないものがほとんどであった。

また、聴解の最終テストの作成と実施にも問題があった。筆者が2週間で帰国し、その後の聴解のテスト作成は、帰国後で細かな点についての確認が出来なかったことで、指示の問題、受験会場の環境など多くの問題を残している。

以上のように、多くの問題点と反省を残した評価であった。今後は、各科目ごとに厳密な記述を試みるための研究と検討会が必要である。

### 【参考文献】

Council of Europe (1997) *Common European Framework of Reference for Languages:*

*Learning, teaching, assessment*, Goethe-Institute Japan 朝日出版社

ティム・マクナマラ(2004)伊東祐郎他訳『言語テストング概論』スリーエーネットワーク