

日本語コースにおける「学期・中間テスト」の役割と意義について ―教師・学習者の双方への有効なフィードバックを求める視点に立つて―

森 康眞

1. はじめに―探究の動機

本稿は『学期・中間試験（以下、中間テスト）』の得点結果をどのように受け止めたらいのか」という問いかけに対する「答え」を筆者なりに探ることを目的とする。そもそもこのような「自問自答」に至った背景には、勤務する大学での日本語コース・シラバス／カリキュラムの全面的改定があった。即ち、2003年度・夏学期より段階的に導入された新コース（副専攻課程／4科目・12単位）は、共時的に旧コースの段階的廃止でもあった。これは2003年度以前（副専攻課程／5科目・15単位）に入学した学生（以下、学習者）にも即時適用され、旧コースを新コースに収斂させる移行であった。旧コースを存置させる経過措置を講じなかったため、旧コースの下で履修していた学習者にとって制度の全面的改変は「学習内容の量的・質的増大」を意味した。

この増大が学習者に大きな負担を強いたのは想像に難くないと思われる。事実、それを証左するものの一つは中間テストの「平均得点」が30点台に低下したことであった。これまでの中間テストの平均点は概ね50点台前後を推移していたというのが筆者の実感であるが、この平均点の低下の起因を「コースの制度的改変」に求めるのは容易であろう。また、この教室外的説明によって教師・学習者が仮に納得するとしても、中間テスト以降の教室内の教授・学習過程において如何なる「プラスの波及効果」を期待し得るのかという懸念も呈しよう。

本稿では筆者が担当した2005年度・夏学期（2005年6月―7月／8週間構成）に開講された科目である「JPN334 日本語 IV」の「中間テスト」を事例に考察を進めていく。考察に際しては「言語テスト理論」の科学的知見や「評価法」の教育的知識等に照らし合わせつつ、「中間テスト」の役割及び意義をトータルに捉え直す中で、冒頭の疑問に答えたいと考えている。この「トータルな捉え直し」の視野には言語テスト理論の観点からは「テストの目的・作成・実施・結果・解釈」について、評価法の立場からは「形成的評価」についてそれぞれ含まれるものである。更に、「中間テスト」に対する教師による「トータルな捉え直し」は、教師・学習者がテストに関してテスト得点の良否のみならず、テストの作成・形式・内容・方法・採点・評価に対する全幅的な納得―満足できるか否かは別として―をも得られるものとしなければならないであろう。加えて、今後の教師・学習者の「言語テスト行動」を探究する上で、横断的であれ縦断的であれ、又は質的であれ量的であれ、重要な基礎データになるとともに、教師・学習者双方への適確にして有効なフィードバックの実行にも資するものと思われる。

2. 「中間テスト」の概要

2.1 中間テストの受験者－学習者の属性

中間テストの受験学生は 2005 年度・夏学期に「JPN334 日本語 IV (科目名)」を履修した 1 セクション (以下、クラス)・52 名の学習者で、男女の構成は男性 7 名、女性 45 名である。このクラスの学習者の成績水準を G.P.A. (平均成績値) と呼ばれる指標で見ると、学習者全体の相加平均は 2.65 で、最大が 3.87 で、最少が 1.97 である。尚、これは日本語科目のみの成績値ではなく履修した全科目の成績点から得られた数値を意味する。中間テストの実施は学期・開講日より 4 週目の大学が定めた試験期間中に行われ、試験時間 (「試験」と「テスト」についてはバックマン 1997 : 58-9 を参照) は 3 時間である。

2.2 中間テストの目的－直接的目的と間接的目的

中間テストの「直接的目的」は何よりもコース開始から中間時点に至る迄の学習者の「学習到達度 (学習量)」を測定することであり、これは「到達度テスト (Achievement Test)」と言われる。この「到達度テスト」は学習側から見れば、学習者自身の達成度や弱点及びクラス内での相対的位置を知る機会となる一方で、教師側から見れば、学習者の到達度が客観的に得られ、また、教師が教えた授業内容が学習者にどのくらい学習されたかを確認する機会にもなる。詰まり「形成的テスト」としての教育的役割 (静他 2002 : 103) で、中間テストの「間接的目的」と言える。この「間接的目的」は教師に指導方法の「反省材料」をもたらすとともに、学習者に対してはコース終了時の学期末テストへの「学習の動機付けと方向付け」を与えることである (石田 1996 : 216-7)。しかし同時に「到達度テスト」は「定期テスト」という制度的位置付けの故、「学力テスト」として成績判定の部分的資料をも兼ね、本学では評定 (grading) の 20% を占めている。

2.3 中間テストの出題内容と形式

中間テストが一定期間における学習者の学習到達度を測定することが目的であれば、テストの出題内容はコース・シラバスに基づく既習項目でなければならないのは言うまでもない。

さて、出題内容は授業計画 (Lesson Plan) で策定された範囲に基づいて出題される。従って、文法・文型を中心とする「個別的要素テスト (伊東 2005 : 14-5)」になっている。これは 2.2. で触れた「到達度」という概念と「言語能力 (バックマン他 2000 : 77-91)」との関連性でもあるが、「学習された言語的知識の到達度」を測るという中間テストの目的に鑑みれば、語彙、文法・語法、言語形式、構文・表現等の「言語の各要素」を正確に理解 (わかる) し、且つ、使用する力 (できる) を部分的に測定することに異論はないものと思われる。詰まり、直接テストとしてのパフォーマンス・テスト (口頭表現力テストなど) や統合的テスト (聴解力テストなど) における「言語技能」の測定方法と異なるにせよ、「学習された言語的知識 (文法・文型)」を要素別に作問することで「言語能力」の構成要素となる「言語における手続き的知識 (迫田 2002 : 67)」を間接的に測定しようとする考え方である。

具体的には教師が各種市販教材から抜粋・編集した教科書から出題範囲を指定した上での教師による自作テストである。学習者も出題範囲に関してはコース開始時のコース・アウトライン（講義概要）及びコース・スケジュール（講義予定表）を通して既知の事項となっている。

中間テストは、問題用紙と解答用紙からなる「紙筆テスト」で、これは具体的な言語使用場面の中で行われるテストではないため、「間接テスト」と言われる。また、試験時間内に全項目に解答するべく出題されているので、「力量テスト」という側面も併せ持っている。具体的な出題形式は大問 9、小問 77 から成る合計 100 点満点のテストである。大問というのは出題項目数で、小問というのは各出題項目内の設問数、即ち、問題項目数を意味する。大問となる出題項目間には出題内容の相対的重要性を考慮して、解答の価値に合わせた配点の重み付け（1～2 点）がなされている。

表 1. 出題内容・配点・解答形式

問題Ⅰ	漢字・書取り	1×11	単純再生法	問題Ⅵ	名詞・形容詞の形	1×5	組合せ/変換法記入式
問題Ⅱ	漢字・読取り	1×11	単純再生法	問題Ⅶ	動詞の活用	2×14	変換法記入式
問題Ⅲ	助詞	1×16	完成法穴埋め式	問題Ⅷ	会話・語句	2×6	完成法自由再生式
問題Ⅳ	疑問詞	1×5	自由再生法穴埋め式	問題Ⅸ	条件・短作文	2×3	記述式
問題Ⅴ	待遇・授受表現	1×6	二肢選択法記入式				

※乗法式の左側の数字は配点、右側の数字は問題項目数（小問数）を示す。

解答様式は基本的に「短答構成形式」を採り、即ち、受験する学習者が「学習した言語的知識（わかる）」を手懸りに自ら「正答（できる）」を考えて解答する形式である（垣田 1985:132-7）。それ故、作問・出題した教師が期待する解答（応答）が学習者になされるよう、テスト問題の指示文を明確にする必要がある。指示文は既習の日本語で書かれているが、試験日の 1 週間前に教師が教室で例を示しながら、出題内容、出題方法、出題形式、解答形式を詳細に説明する措置を講じた。

2.4 中間テストの「妥当性」

中間テストに限らず、テストの「妥当性」は 2.5. で扱われる「信頼性」と併せて重要な要件の 1 つである。テスト自体の有用性には、テストの特質という観点から、1) 信頼性、2) 妥当性、3) 真正性、4) 相互性、5) 影響、6) 実用性が列挙されている（バックマン他 2000:20-46）。

テストの「妥当性」については数的基準よりも出題者である教師の作問能力に係っている部分が大きいと言われるが、端的に言えば、「妥当性」は教師によって出題された問題内容が学習者の「到達度＝文法・文型＝部分的言語能力」を的確に測定するのに相応しいものであるか否かのことである。逆に「信頼性」が如何に高くても、出題内容が指導・学習内容と懸け離れていては何の意味も認められないということである。教師は「妥当性」の高い作問に努力する姿勢が求められる。この「妥当性」を検討する幾つかの柱に、1) 構成概念的妥当性、2) 内容的妥当性、3) 経験的妥当性や実証的妥当性とも呼ばれる基準関連的妥当性(併存的妥当性と予測的妥当性)、4)

表面的妥当性が取り上げられている（石田 1992：40-3）。

何れにせよ、テストの得点を解釈する際の重要な前提条件である「妥当性」は、テストの目的と出題された問題項目の内容の適切性をコース・シラバスやカリキュラムに照応させつつ、検討する必要がある。この点については 3.2. 及び 3.3. で具体的に検討することにする。

2.5 中間テストの「信頼性」－採点と得点結果

「信頼性」は採点されたテスト得点が信頼に値する、一貫した結果であるか否かの「得点の安定性」を問うものである。その「信頼性」を検証する方法として、1) 再テスト法、2) 平行テスト法、3) 折半法、4) 内部一貫性（等質性）又は内的整合性がある（三浦 2004：8／石田 1992：43-9）。筆者は手統計で今回の中間テストの「信頼性」を検証してみた。テスト全体の信頼性を求める公式は以下の通りである（ヒートン 1992：243）。

$$\text{信頼性} = \frac{\text{問題項目数}}{\text{問題項目数} - 1} \left[1 - \frac{\text{平均点}(\text{問題項目数} - \text{平均点})}{\text{問題項目数} \times \text{標準偏差の二乗}} \right]$$

問題項目数（小問数）は 77、平均点は 32.98（小数第 3 位四捨五入、以下同様）、標準偏差は 19.25 であった。上の公式に当てはめて計算した結果は 0.96 であった。この数値は「信頼性係数」と呼ばれ、-1.00～+1.00 の範囲をとるが、一般的に 0.90～0.99 の間にあれば「信頼性」は高いと評されるようである（伊東 2005：64）。しかしながら、「信頼性係数」は問題項目数が多くなればなるほどその数値も高くなるという性質が見られるようである（三浦 2004：10）。従って、「信頼性は妥当性の必要条件」と言われる如く、2.4. で述べた「妥当性」の高いテストは必ず「信頼性」も高くなると言えるが、「信頼性」が高いからと言って「妥当性」も高くなると言うことには必ずしもならないことに留意しなければならない（日本語教育学会 1991：63）。

次に「信頼性」は採点された得点（素点）に対しても検討が必要である。即ち、得点そのものが何を意味するのかを理解しておきたい。テストという媒体を手段に間接的に得られた数値は「測定値」であって、可視的な実測値でない点である。これは採点された得点が真値ではなく、誤差を含んでいるという「古典的テスト理論」の考え方である（バックマン 1997：191-3）。「測定値（テスト得点）＝真値（真の言語能力）＋誤差（小池他 2004：280）」で示されるように、誤差の割合が少なくなる程、「信頼性」は向上すると言える。また、測定値に影響を及ぼす背景には問題作成者側の問題、採点上の問題、当て推量の可能性、受験・学習者側の個人的な要因など、様々な理由や原因が挙げられる（日本語教育学会 1991：16-7）。ここでは、得点との関わりが深いと思われる採点方法について若干触れるとする。中間テストは形式・内容的には「客観テスト」の外観を保つものの、解答用紙に記入された「日本語」には 1) 表記の誤り、2) 文法上の誤り、3) 語の使用の適・不適、4) 表現の良し悪し、などが見られ、採点者の主観が入りやすい「主

観テスト」の側面も併せ持っている。しかしながら、正解基準に照らし合わせた採点方法を心掛け、評定者内信頼性（静他 2002：93-4）を高めるよう努める必要があろう。中間テストでは「0, 0.5, 1, 1.5, 2」の配点内で、減点方式による部分採点を採用している。尚、採点者は出題者でもある教師自身で、採点された 52 名の学習者の得点結果を以下に掲げておく。

表 2. 中間テストの採点・得点結果

87	63.5	49	36.5	32.5	27.5	24	21.5	18.5	14	10
83	60.5	48	35.5	32.5	27	23.5	21	16.5	13	9
74	60	42	35	30.5	26	23.5	20.5	16.5	11.5	
69	50.5	40	33.5	29	24	23.5	20	16	11.5	
65	49.5	37	33.5	28.5	24	22	19	15.5	11	

※中間テストの合計得点の最低点は 0 点で、最高点は 100 点である。

3. 中間テストの得点結果の処理と解釈—集団的特性と個人的特性

3.1 記述統計

採点された得点は素点と呼ばれるが、素点だけではテスト結果の解釈には何の意味も有しない。そこでこの素点を基礎に統計的手段による処理を施すことによって、得点結果の解釈が可能となる。そして、この統計的処理によって今回の中間テストの得点結果をどう解釈したらいいかということである。もちろん、統計的処理を行う目的は受験した学習者の到達度の程度や得点結果全体の解釈を有意義に解釈するためである。統計的処理と言うと、SPSS 社の SPSS シリーズと呼ばれる統計パッケージが代表的であり、且つ、最もよく使われているが、これは推計統計の領域を含むものである。以下では基本統計量を中心に得点の解釈を行う、記述統計という処理に拠る。記述統計は Microsoft 社の Excel（表計算ソフト）に組み込まれている「分析ツール」でも可能であるが、筆者は Web 経由でインストールした FreeJSTAT というフリーウェアを使ってみることにした（三浦 2004：153-4）。

得点結果の解釈に際しては集団基準準拠テスト（NRT）的側面と目標基準準拠テスト（CRT）的側面に着目したいと思う。前者は学習者の順位化やグループ化に有用で、言わば集団的特性を「視る」ことであるのに対して、後者は学習者の学習の到達度を把握化することに有益で、個人的特性を「診る」ことである。しかし、統計的に得られた数値や数量の解釈の前提には、あくまでも学習者の得点結果の解釈がクラス集団に依存したもので、且つ、得点の状況がクラスを構成する学習者の能力や性質に左右される性向にあるということである。

さて、今回の中間テストの基本統計量を概観すると、以下の通りである。

3.1.1 代表値

代表値には平均点が最も広く使われているが、平均点は 32.98 点である。他に中央値や中点や最頻値がある。中央値は得点を順位で並べた時に中央に来る値で、27.25 点である。中点は最低

得点と最高得点との中間点で、48 点である。最頻値は最も多数の学習者が取った点であるが、10 点毎の級間（1～10 点）で見ると、21 点から 30 点の級間で 15 名となり、級間の代表値で言えば、25.5 点となる。

3.1.2 得点範囲

代表値だけでは得点全体の状況を見るには十分とは言えない。そこで散布度を求める訳であるが、得点範囲は得点の散らばりの程度を見るためのものである。最小値（最低点）が 9 点で、最大値（最高点）が 87 点である。範囲は 9 点から 87 点までである。

3.1.3 歪度と尖度

歪度は得点の分布が左右対称或いは右ないしは左に片寄っているかを示す一方で、尖度は得点の分布が尖っているか或いは偏平であるかのどちらかを示す。歪度は 1.13 であるから、分布の山は左に寄るとされる。尖度は 0.63 で、正規分布に比して偏平とされる（石田 1992：175）。

3.1.4 標準偏差

平均点（平均値）からの隔たりを示す数値で、求め方は個人の得点と平均の差を自乗した総和を受験者数で割り、その平方根を取った数値で、19.25 (STDEVP)である。尚、分散は標準偏差を二乗した値である。

3.1.5 標準得点と偏差値

標準偏差が集団としての得点の分布傾向を示す指標であるのに対して、標準得点は学習者個人の得点が全体のどこに位置しているのかを相対的に表し、個人の得点が平均点からどのくらい離れているかを示すものである。しかしながら、標準得点はマイナス表示も出現するため、分かりやすい偏差値を使う訳であるが、偏差値は平均値を 50 点と看做し、標準偏差を 10 としたものである。尚、パーセンタイル順位も集団内の相対的位置を表すのに使われる指標である。

表 3. 中間テストの個人別偏差値

78.1	65.9	58.3	51.8	49.8	47.2	45.3	44.0	42.5	40.1	38.1
76.0	64.3	57.8	51.3	49.8	46.9	45.1	43.8	41.4	39.6	37.5
71.3	64.0	54.7	51.0	48.7	46.4	45.1	43.5	41.4	38.8	
68.7	59.1	53.6	50.3	47.9	45.3	45.1	43.3	41.2	38.8	
66.6	58.6	52.1	50.3	47.7	45.3	44.3	42.7	40.9	38.6	

※偏差値＝（標準得点×10）＋50、標準得点は個人の得点から平均点を引き、これを標準偏差で割った数値である。

3.2 正答数・誤答数・無答数

合計された得点のみではテスト結果の内容を個別的に見るには不十分と言えよう。詰まり、問題項目の内容と学習者の得点を関連付けて、解釈する試みである。紙幅の関係上、全項目は割愛せざるを得ないが、配点が相対的に高い問題Ⅶの動詞の活用について検証することとする。

表 4. 問題Ⅶの正答数・誤答数・無答数（人数）

項目	1	2	3	4	5	6	7	8	9	10	11	12	13	14
正答数	17	22	21	28	13	35	16	15	21	35	12	33	19	13
誤答数	33	29	30	23	37	16	34	36	29	15	39	17	31	37
無答数	2	1	1	1	2	1	2	1	2	2	1	2	2	2

※正答には減点された部分的正解も含まれる。項目 3 と 4 の間の点線は同一設問の中に 2 つの項目があることを意味する。

表 4. で正答数が誤答数を上回っている項目は 4、6、10、12 の 4 項目のみであった。更に項目難易度を算出すると、表 5. のようになった。項目難易度は 0～1 の間の値をとるが、1 に接近すれば容易であり、逆に 0 に接近すれば困難であったということを示している。到達度を測定するテストの目的を考慮に入れると、かなり低い数値になってしまったことが窺える。

表 5. 問題Ⅶの項目難易度（通過率）

1	2	3	4	5	6	7	8	9	10	11	12	13	14	平均
0.33	0.42	0.40	0.54	0.25	0.67	0.31	0.29	0.40	0.67	0.23	0.63	0.37	0.25	0.41

※正答数を受験・学習者数全体（52 名）で割った数値である（橋本 1981：88）。

もう 1 つの問題項目の分析に項目識別力（項目弁別力）がある。これは得点の高い学習者と低い学習者を識別しているか否かを調べるものである。

表 6. 問題Ⅶの項目識別力

1	2	3	4	5	6	7	8	9	10	11	12	13	14	平均
0.72	0.72	0.50	0.61	0.33	0.44	0.67	0.61	0.67	0.44	0.38	0.44	0.55	0.33	0.53

※項目識別力＝（得点上位群の項目難易度）－（得点下位群の項目難易度）であるが、得点上位群及び得点下位群をそれぞれ全体の 35% に当たる 18 名とし、得点中位群を全体の 30% に当たる 16 名と便宜上分類した。

項目識別力はマイナス表示になることもある。これは上位群の学習者が誤答し、下位群が正答した場合である。0 の場合は識別なしと看做される。一般的に項目難易度が 0.30 から 0.70 の間で、項目識別力が 0.40 以上の場合に良問と言えるようである（伊東 2005：69-70）。その意味で問題Ⅶ全体を平均で見限りにおいて、項目難易度が 0.41、項目識別力が 0.53 で一応良問であったと判断してもいいのではないと思われる。

3.3 「誤答」のパターンと「無答」

3.2. で得られた統計値とテストの問題内容との関連性を考察すると、項目難易度が 0.30 未満の項目 11 (0.23)、5、14 (各 0.25) と項目識別力が 0.40 未満の項目 5、14 (各 0.33)、11 (0.38) に共通して見られた点は「そうです (様態)」という文型に接続する「動詞の正しい形 (活用形)」

であった。項目 5、11、14 の誤答パターンを表 7. で見ると、3 項目とも「テ形」に変換した誤答が顕著に表れていることである。項目 5 で誤って変換した「テ形」も含めた誤答数は 27 名に上り、項目 5 の全誤答数の 72.97% を占める。そして、第 1 位から第 3 位までに占める「テ形」の出現度を見ると、項目 11 で 86.96%、項目 14 でも 70.97% で高くなっている。一方、項目 11 の「落って」「落つ」から類推して「落ちます」がグループ I の動詞と誤認されていることや項目 14 の誤答例からは「切れます（自動詞）」と「切ります（他動詞）」の混乱も見られ、加えて項目 5 と 11 では「くて」があり、品詞の誤解も見られるなど、誤答パターンの分析の重要性を改めて痛感した形となった（垣田 1983：20-42／ラドー 1971：407-8）。

表 7. 問題Ⅶにおける問題項目 5・11・14 の誤答例（最多順）

項目	第 1 位	第 2 位	第 3 位
5	外れて (23)	外れる (3)	外れんて/外て/外んて/外れくて/外れる/外る/外るん/外れない/外れるない (各 1 名ずつ)
11	落って (13)	落つ(3)/落ちて(3)	落いて(2)/落ちくて(2)
14	切れて (20)	切れる (7)	切って(2)/切る(2)

※括弧内の数字は人数である。

これに関連して、表 5. の項目難易度が 0.50 未満の項目は 1、2、3、5、7、8、9、11、13、14 であるが、この内「マス形」接続の項目は 1、2、5、8、9、11、14 で問題Ⅶの項目総数の半数を占めているにもかかわらず、「正答」に至っていないことに疑問を感じた。因みに項目 1 と 8 は「やすい／にくいです（複合形容詞）」で、項目 2 と 9 は「すぎます（複合動詞）」である。接続する活用形を問う出題形式と内容は学習者が過去これまでに履修した科目の定期テストと同様のものである。確かに過去の定期テストでは文型の中に与えられた動詞の「…ます」を「テ形」「辞書形」「ナイ形」「可能形」「意向形」「条件形」「受身形」等に変換する傾向が多く、「…ます」の「ます」と取った「マス形」を問う活用は少なかったのは事実である。それが今回の中間テストでは「マス形」接続の活用形が半数であったため、学習者は「過去のテスト経験」上からある種の「戸惑い」を感じたのだろうか（問題項目の特性）。或いは、「…ます」の「ます」を取った形式が他の活用形と同じように「マス形」になるという理解が明確になされていなかったのか（学習者の特性）など更に詳しく調べる必要があるかも知れない（大友 2002：94）。

「誤答」に加えて「無答（未答）」に関しても解答記入漏れの不注意、覚えていない、忘れた、思い出せない、時間が足りないなど種々の原因があると考えられるが、刮目する必要がある。

4. 中間テストに対する学習者の受け止め方

中間テスト直後の初回の授業終了時にプリコード・多項選択回答法によるアンケート（14 項目）を実施した（有効回答 38 名）。これは学習者の中間テストに対する「感想」を訊ねたもので、学

習者の「情意反応」がどのように表れているかを見るためである。主な質問項目だけを掲載する。

表 8. 中間テストに関する質問と回答数 (%)

1.試験時間	長過ぎる： 2(5.26)	妥当： 35(92.11)	短過ぎる： 1(2.63)
2.問題項目量	多過ぎる： 9(23.68)	妥当： 29(76.32)	少な過ぎる： 0(0.00)
3.出題内容(全体)	難し過ぎる： 12(31.57)	難しい： 16(42.11)	納得いかない： 10(26.32)
4.7.問題Ⅶの内容	難しい： 21(55.26)	中位： 15(39.48)	易しい： 2(5.26)
5.出題と学習内容との関係	相応： 31(81.58)	納得いかない： 7(18.42)	不相応： 0(0.00)

6. の自由記入欄で目立った事項はテスト問題の中で使われている「漢字含有率」に触れるもので漢字にルビを振って欲しいという要望が若干見られた。

5. 終わりに「形成的評価」との関連性

教師は測定された「得点結果(到達度)」を KR 情報(大友 2002: 12-3)として学習者に伝える一方で、「得点結果の解釈」を通した「評価」をしなければならない。それは中間テストの目的である「到達度の測定評価(佐治・真田 1996: 83)」を意味する。この到達度の評価を学習目標と内容的対応を軸とする質的規準で見るか、分割点のような数値的基準で見るか、又は全体(集団)で見るか、部分(個人)で見るか、更に集団内の内在的基準で見るか、集団外の外在的基準で見るか、などアプローチは様々である。しかし、これらのアプローチは測定結果に対する教師による評価に過ぎない。中間テストが有する形成的評価の役割を考慮に入れるならば、1) 学習目標と到達度、2) 教師による授業活動、3) 学習者の学習活動、4) 学習者の学力的側面、5) 学習者の動機付けと方向付け、6) クラス全体、を対象とした広範囲に及ぶものでなければならない(石田 1992: 7-8)。

即ち、集団・集団内個人・個人別レベルのそれぞれの「結果」から「過程」への関心で、また、近年の学習者中心主義に立つ完全習得学習理論とを併せた視点からは学習者の「何ができて、何ができなかったか」という到達度結果を受けて、教師・学習者が今後の指導・学習プロセスの中で何をしなければならないかという「教育・学習活動」への取り組みである。逆接的に言えば、テストから得られた「結果」が「中間テスト」を媒介に「教師の指導と学習者の学習への双方向的評価」を促進・形成しているということである。この「双方向的評価」は常に「教師→学習者→教師」という相互的循環サイクルの中で有効的なフィードバックを可能にさせるものとなろう。この有効的なフィードバックの可能性を生む「双方向的評価」を行うためには、評価の判断材料となる「精確な情報」が必要となる。この点で、筆者が本稿で試みた記述はその出発点にあることは言うまでもない。

最後に筆者自身の問いかけに対する「正答」は暫定的なもので、「完答」ではないかも知れないが、上述の「双方向的評価」の具体的枠組み作りと質的向上を目指して、且つ、教師・学習者が

共に指導・学習目標に到達できるよう、努力を重ねていくことを誓って、結びに代えたい。

参考文献

- 石田敏子 (1992) 『入門日本語テスト法』大修館書店
- _____ (1996) 『改訂新版 日本語教授法』大修館書店、215-255 頁。
- 伊東祐郎 (2005) 『NAFL 日本語教師養成プログラム 22 日本語教育評価法』アルク
- 大友賢二 (1996) 『言語テスト・データの新しい分析法 項目応答理論入門』大修館書店
- _____監修 (2002) 『テストで言語能力は測れるか～言語テストデータ分析入門～』桐原書店
- 垣田直巳監修 (1983) 『英語教育学モノグラフ・シリーズ 英語の誤答分析』大修館書店
- _____監修 (1985) 『英語教育学モノグラフ・シリーズ 英語の評価論』大修館書店
- 小池生夫編集主幹 (2004) 『第二言語習得研究の現在－これからの外国語教育への視点』大修館書店、275-294 頁。
- 小林ミナ (1998) 『日本語教師・分野別マスターシリーズ よくわかる教授法 日本語教育能力検定試験対応』アルク、125-154 頁。
- 静哲人・竹内理・吉澤清美 (2002) 『外国語教育リサーチとテストの基礎概念』関西大学出版部、3-195 頁。
- 迫田久美子 (2002) 『日本語教育に生かす第二言語習得研究』アルク、194-220 頁。
- 佐治圭三・真田信治監修 (1996) 『日本語教師養成シリーズ 日本語教授法』東京法令出版、82-104 頁。
- 高見澤孟 (1996) 『はじめての日本語教育・2 日本語教授法入門』アスク講談社、12-13 頁及び 136-146 頁。
- 日本語教育学会編 (1991) 『日本語テストハンドブック』大修館書店
- 橋本重治 (1981) 『到達度の研究－その方法と技術－』図書文化
- ライル・F・バックマン [池田央・大友賢二監修] (1997) 『言語テスト法の基礎』C.S.L 学習評価研究所
- L.F. バックマン/A.S. パーマー [大友賢二/ランドルフ・スラッシャー監訳] (2000) 『〈実践〉言語テスト作成法』大修館書店
- J.B. ヒートン [土屋澄男・斎藤誠毅監修] (1992) 『コミュニカティブ・テストイングー英語テストの作り方』研究社
- 三浦省五監修 (2004) 『英語教師のための教育データ分析入門－授業が変わるテスト・評価・研究－』大修館書店、4-27 頁及び 153-157 頁。
- ロバート・ラドー [門司勝・本田漠・吉田一衛・松田熙一訳] (1971) 『言語テスト 外国語テストの作成とその利用』大修館書店