

平成 17 年度 校内研修会

主題：「授業の質的向上のために」

期日 平成 17 年 9 月 14 日（水） 13：00～

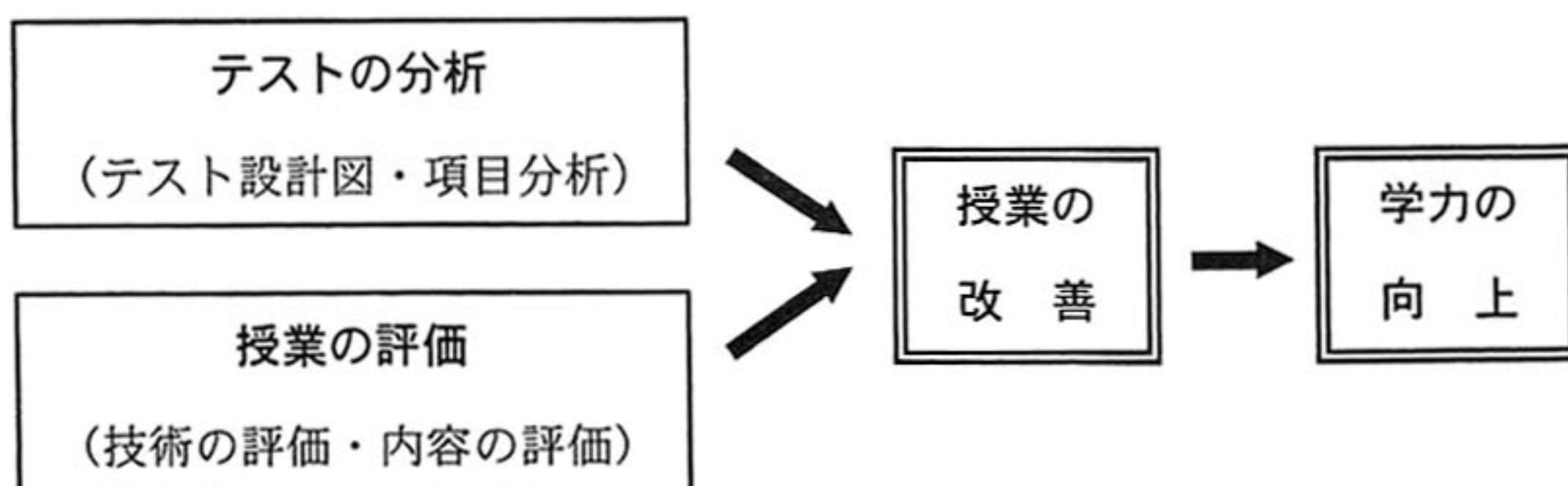
場所 北海道札幌北陵高等学校 会議室

「テスト分析と授業評価」

<説明のポイント>

- (1) 生徒の学力向上要因の 1 つに、授業改善がある
- (2) 授業改善の目的は生徒の学力を向上させることにある
- (3) 授業改善の方法には、テストの分析や授業の評価などがある
- (4) テストの分析には、テスト設計図や項目分析などがある
- (5) 授業の評価には、授業「技術」の評価と授業「内容」の評価がある。

<説明のモデル>



1. テストの分析

(1) テスト設計図 (test blueprint/具体化表 table of specifications ともいう)

ア. 設計図はテストに不可欠である テストは生徒の授業理解度や学力の定着度をみるだけでなく、テスト結果を利用して5段階に評定し、追認考査の対象者を特定したりもする。当然のことである。

だが、資料1のように「テスト設計図」を用意して、目標と内容のバランスを点検し、個々の項目の難易度に配慮しながら実施されるテストが、ふだん、いったいどれくらいあるだろうか。「設計図なしで作った飛行機をパイロットが操縦することはありえないが、設計図なしで作ったテストを教師が実施することはよくある」とは、ある専門家 (Mehrens & Lehman) の見解だが、一方、われわれ教師が大学時代に習った記憶がないのも事実である。残念ながら「測定と評価」の基礎的教養は研修会でも教えられていない。今、はじめてお聞きになった方は「そんなものいちいち作ってられないよ」とおっしゃるかもしれない。もっともな話である。

イ. 設計図はテストの内容的妥当性を保証する 資料1のテスト設計図は、昨年度前期末に . . . の3人が共同で指導した第1学年「現代社会」のそれである。縦軸に単元内容、横軸に目標 (= 評価の観点) を設定し、内容的にも・目標水準でも・難易度の点でもバランスの取れたテストを設計するために開発したものである (評価の観点のうち、「関心・態度」は筆記テストに馴染まないのを除外してある。また、「知識・理解」は近年軽視されがちであるが、基礎学力として重要なものであるから、「知識」と「理解」に分割して表示してある)。言い換えれば、テスト設計図とは、テスト全体が内容と目標をうまく代表する項目 (テストの1つ1つの小問を「項目」という) の集合から成り立つようにするための用具である。

このような性質をテストが備えたとき、「テストの(内容的)妥当性」が確保された、という。反対に、このバランスを確保しなければ、生徒の学力を妥当に測定したことにはならない。ちなみに、大学入試センター試験では、毎年この設計図に基づいて本試験と追試験の問題を作成しており、2つの妥当性がほぼ等価になるよう設計している。

ここでは公民科目を例示したが、指導要録に添付された「評価の観点」一覧を参照すれば、このような設計図は、どの教科でも開発可能である。その効用をいくつか説明したい。

ウ. 設計図はより高度な目標を測定する項目の開発を教師に促す 資料1のテストは多肢選択式項目56個からなる期末テストで、地歴・公民科目にありがちな、知識偏重型テストになっていない点にご注意いただきたい。どの教科・科目でもそうだが、テスト設計図なしの「カンと経験」だけでテストを作成すると、特定の目標 (= 観点)、特に知識に偏ったテストになってしまうものである。知識を測定する項目は教師が「作成しやすい」からである。

テスト設計図を用意した上で、項目を開発するたびに、その番号をこの図に書き入れていくと、知識以外の高度な目標を測定するための項目開発が、想像以上に困難な作業であることにすぐ気づく。当然、産みの苦しきも味わう。だが、そこでの産みの苦しきは、教師のテスト問題開発の技術力を着実に向上させていく。ひとたび技術が

向上すると、レベルダウンすることはなくなる。開発技術が向上すると、他人の開発した問題の良否を鑑定する力も向上する。目標を持つ重大な意義の認識も深まる。要するに、われわれの「カンと経験」がどんどん洗練されていくのである。

エ. 設計図は過去のテストを点検させ、学習習慣の改善をも促す また、この図を用いると、過去に実施したテストにどの程度の妥当性があったのかを分析し評価することもできる。「ウチの生徒は考えないで、すぐ暗記に走るからダメだ」などと言っているとしても、生徒を暗記に走らせる、知識偏重型テストを実施していたかもしれない。それが本当かどうかはテスト設計図でただちに点検できる。このとき、テストの形態が、生徒に悪しき学習習慣や学習態度を生んでいた、ということを発見させてくれる場合が多いし、同時に、テスト改善の指針をも与えてくれる。テストの改善が学習習慣を改善することとなり、いわば「指導と評価の一体化」に貢献していく。

オ. 設計図は協議しながらテスト項目を洗練してゆく土俵である 56個の項目案をすべて開発し終えてから、われわれ3人は、内容と目標に照らして出題が妥当かどうかを1つ1つの項目について協議し、改訂作業を進めた。この協議の段階で、難易度を調整したり、選択肢の言い回しや誤字脱字を手直ししたり、1つの項目全部を作り直したりしながら、テスト全体を洗練していった。項目案の改訂に必要なと感じたら、遠慮してはいけない。意見を出し合えば出し合うほど、よりよいテストが出来上がっていくからである。3人が協議するときの共通の土俵、それがこのテスト設計図である。

カ. 設計図作成上の注意とヒント 実は、テスト設計図は、個々の項目を開発する前に、表の右側の合計欄と下段の合計欄を作成しておかなければならない。このとき、

①単元内容は、授業で割いた時間に比例して配点を決めていく。

②目標(=評価の観点)は、基礎的目標で50%、発展的目標で50%とか、そのうちの「知識」は20%、「理解」は30%などと、観点別に目安を決めておく。

資料1を例に取れば、内容の13~14は、テスト範囲全体の20%、その領域で「知識」が20%、「理解」が30%とすると、それぞれ

③知識： $100 \text{点} \times 0.2 \times 0.2 = 4 \text{点}$ 、理解： $100 \text{点} \times 0.2 \times 0.3 = 6 \text{点}$ 、の配点になる。

これら①~③を目安に1つ1つの項目を開発していく。仮に、「知識」項目は1つ1点、「理解」項目は2点とすれば、それぞれ4項目、3項目を開発すべきであることが自動的に決まるし、開発に要する時間もおよそ見当がつくのである(実際には、「知識」2項目、「理解」4項目でズレが1つあるが、それはわれわれ3人が協議を進めていったときに、改訂したためである。このような微調整はむろん差し支えない)。

以上の説明は、分担してテストを作成するような場合、特に有効である。複数の教師が同一科目を指導し、作成するテスト項目を分担し合うことはきわめて多い。だが、ふつうは、たとえば「テスト範囲の前半と後半を2人で分担しあうという取り決め」に留まっているのではないだろうか。テスト範囲とはあくまで「内容」であり、「目標」を指していない。教師たちがテスト全体で「理解」は何%、「思考力」は何%とあらかじめ設計図を用いて決めておかなければ、作問を引き受けた教師にとって開発すべき項目数は決まってこない。この作業を省けば、いくら分担した項目案を寄せ集めても、テスト全体の妥当性は保証されない。このままでは、出題する担当者の癖や好みに合わせて、生徒がテスト勉強をするという、悪しき学習習慣・態度を助長することにも

なる。

また、単なる知識よりも、理解・思考・判断など上位の目標を代表する項目を多く出題できるようになると、われわれ教師は、自然に、これらの学力が身につくような授業展開を工夫するようになる。要するにテストの改善が授業の改善を促すのである。

キ. 設計図は教師の作問習慣を改善する 資料1に例示したテストは、実施2週間前から作成を開始し、10日前に原案が完成した。その後何度かわれわれ3人が協議しながら改訂を続け、1週間前に最終原稿が完成している。このため、実施当日に問題文中の誤字脱字や解答欄不足などの訂正のために教室を巡回するという、ごくふつうにありがちな事態は生じなかった。

1週間前の完成というのは、別の研究目的があったから早かったのだが、少なくとも3日前の完成でもよい。それだけでも「訂正のための教室巡回」はほぼ避けることができよう。

また、1項目ずつ原稿を作成しながら設計図に書き込む生活を始めると、项目开发のアイディアは、授業を実施している最中または実施直後に得られることが多くなる。授業が進行していけば、開発した項目数もそれとともに増えて記録されていくため、未開発の内容項目や、未開発の目標項目は自動的に絞り込まれ、時間を割いて作成すべき項目に集中できるようにもなる。開発済みの項目でも、選択肢の配列を変更したり、正解番号の位置を工夫したり、あるいは微妙な表現やレイアウトを工夫するなどのゆとりも生まれる。精神衛生上、非常によい。

かく言う私も、4、5年前からようやくこの生活習慣が確立したにすぎない。それ以前は、「推敲」している満足感に浸りながら、テスト実施前日までかかって作成していた。そして推敲し何度も点検したはずなのに、必ずと言ってよいほど印刷後に訂正箇所が見つかった。そのたびに自分の至らなさにがっかりしていた。

(2) 項目分析 (item analysis)

テスト設計図は、テストの実施前に、設計と開発に必要な用具であったが、テストの実施後に、当のテストがどの程度妥当に生徒の学力を測定したのかを点検する作業が「項目分析」である。これは、テスト全体の平均点や得点分布を調べるものではなく、1つ1つの項目の通過率（項目に正解した生徒の百分率）などを調べて、項目の適否や改善点、さらに授業の改善点までも探り、テスト作成技能を向上させていくことに目的がある。なお、項目分析の手法はこれまで約70年間に種々開発されてきたが、どれも大同小異と言われ、特に「学校教師にとっては古典的な項目分析で十分」とも言われている。この古典的な手法を紹介する。

ア. 項目分析原票を作成する テストが終了すると、われわれは採点を急ぎ、できるだけ早く答案を返却してやりたいと考える。生徒の気持ちを考えれば当然であろう。しかし、このときただ返却するのではなく、(忘れがちだが) 答案のコピーを取っておくだけで、大きな収穫が得られる。答案は項目分析のための豊富なデータなのである。

資料2は、資料1に示したテスト結果をまとめたもので、当時私が指導していた1年8組のものである。名前がないので、仮に「項目分析原票」(item-analysis raw data)

と名づけた。この表は縦軸に生徒番号、横軸に項目の通し番号（正解番号と配点も）を並べ、正答なら空欄、誤答ならその番号、無答なら「*」を升目に記入していく。1人1人について横方向にこれを集計すると、正答率（正答百分率）と得点が得られる。得点は上位者から下位者に並べ直し、上下に折半する（表の中では二重線で区切っている）。

項目を縦方向に見て空欄の数を集計すれば、その正解数と正解率がわかる。これを通過数、通過率という（通過率に配点をかけたものを仮に「得点率」と名づけたが、ここでは空欄のままにした。得点率を全項目について集計すれば、クラス全体の平均点と一致する）。通過率は項目について言い、正答率は生徒について言う。

この表に現れた誤答や無答を、上位群と下位群ごとに集計しておく。あとで述べるが、「まよわし分析」で利用する基礎データとなる。Excelに不慣れな私でも、集計作業は比較的短時間でできる。項目分析原票の作成作業で時間を要するのは、集計作業ではなく、答案を見ながら行なう入力・点検作業である。入力・点検だけで1クラス当り約4時間かかる。サンプルだけで3クラス実施したからざっとこれだけで12時間である。もし、大型のマークカードリーダーがあれば、入力・点検・集計から、これから述べる分析まで、5分程度で終了できる。約60万円で購入できるのだが当面は無理だろう。だから、ここでは手作業で作成している。また、手作業であるが故に項目分析は一般の先生方に敬遠されてしまいがちである（入力・点検・集計・分析が一度にでき、専用の解答シートが不要、価格は15万円程度という格安のソフトもある。ただし、1クラスに25分位かかる）。

イ. 通過率を計算し、特徴的な項目を抽出する ここからが本番である。項目の通過率は困難度とも呼ばれるが、項目反応理論以外の通常の場合では、特に区別されていない。通過率は百分率で示せば0~100の値を取るが、大きいほど正解者が多く、平易であり、小さければ正解者が少なく、難しかったことになる。サンプル組の1年8組についてみると、通過率が高いのは28番(97%)の1個であり、低いのは4番(24%)、14番(13%)、16番(11%)、31番(24%)、47番(24%)の5個である。われわれ学校教師が実施するテストは、平易な項目から困難な項目までまんべんなく出題するは

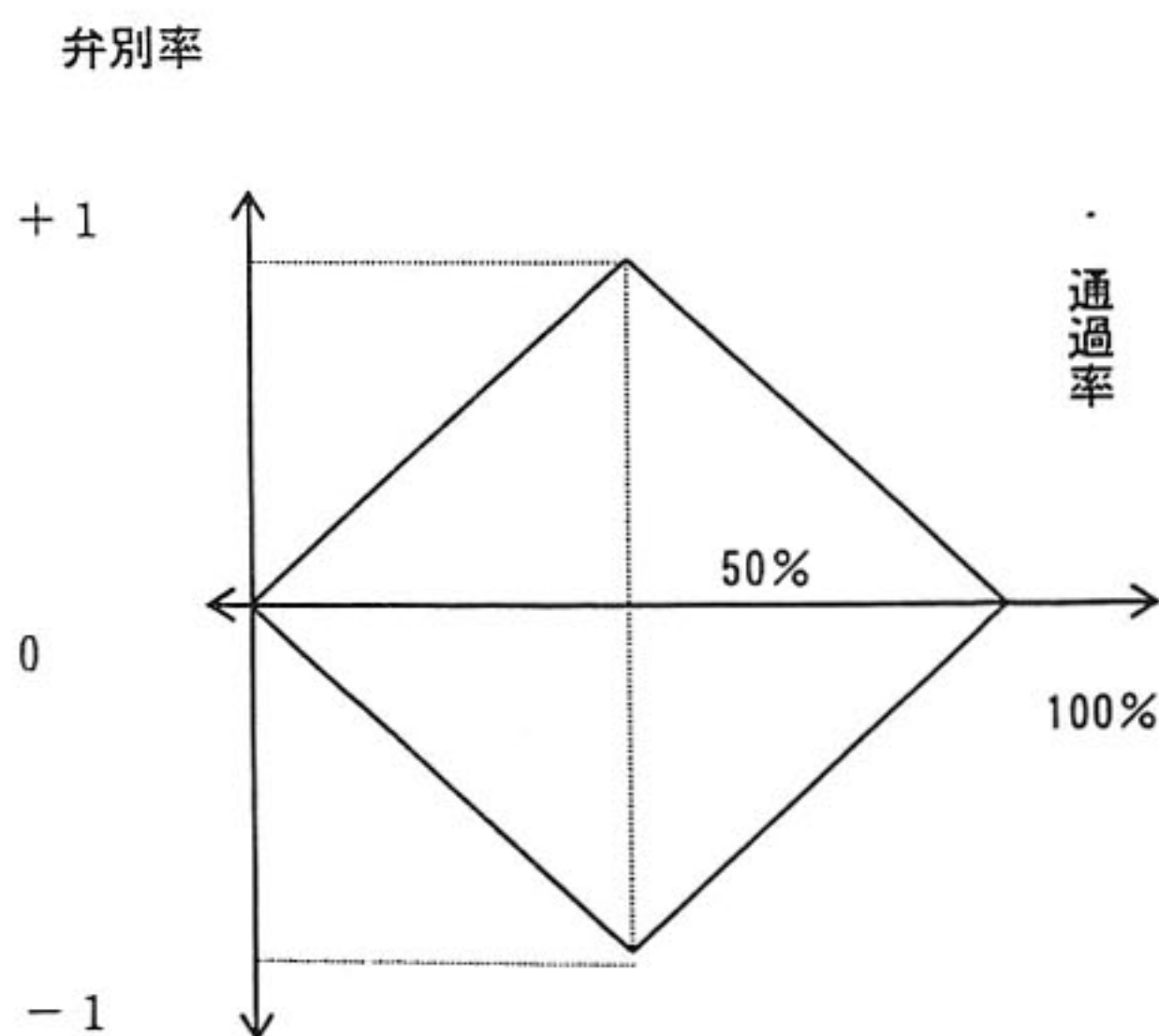


図1 通過率と弁別率の関係

ずであるから、通過率に多少とも高低があるのは構わない。だが、たとえば通過率が0または100なら、出題しなくても同じである。そのような項目を、他のサンプル組（1年5組・7組）と合わせ3クラス全体で拾えば、28番（98%）、14番（4%）、22番（9%）が特に浮かび上がる。項目の（あるいは授業や生徒の学力の）どこに欠陥があるのか、の詳しい検討は後で行なうが、各項目の通過率はこのような事態を手短かに教えてくれる。

ウ. 弁別率を計算し、負の値を持つ項目を分析する 弁別率（識別率と呼ぶこともある）とは聞きなれない用語であるが、有用である。これは「上位群の通過数－下位群の通過数」を全体人数の半分（ここでは19人）で割った値である。弁別率の変域は、－1から＋1までである。上位の成績と下位の成績を弁別する程度を示す指標という意味を込めてこの名称がついている。

通過率と弁別率の関係について一言すれば、通過率50%のとき弁別率＝＋1で最大、－1で最小となり、通過率0または100%のとき弁別率＝0である。さらに言えば、各項目の通過率を x 、弁別率を y として平面上にプロットすると、図1のひし形で囲んだ図形の中に点在することになる。実際の分布は平面の第1象限に多く、第4象限に少ない。弁別率＝－1というのは、成績上位群の正解者がゼロで、下位群が全員正解したことを意味するから、理論的には可能でも実際にはありえない。だが、もし弁別率が負の値を取れば、上位群の正解者が下位群よりも少なかったことを意味するから、項目として望ましいものとは言えず、次のような原因が考えられる。すなわち、

- i. 指定した正解番号に誤りがある
- ii. 正解選択肢が複数存在する
- iii. 正解選択肢の言い回しが曖昧である
- iv. まよわし（不正解の選択肢）の表現に何らかの欠陥がある
- v. 幹（問題内容を指示する文）の表現に何らかの欠陥がある
- vi. 出題項目を授業で指導しなかった、または指導が不十分だった
- vii. 授業で間違えて指導した
- viii. 生徒の学習（習慣や態度）が不十分だった、または内容を誤解した

このうち、iからvまではテストの改善で克服できる問題点であり、vi～viiiは授業や生徒の学習の改善によって克服可能な問題点である。弁別率の数値だけで、そのどれに該当するのかを特定できる場合は少ないが、次の「まよわし分析」と組み合わせればおよその見当はつけられる。手短かに言えば、項目分析の結果は授業の改善を促す場合がある、ということである。

資料2の1年8組を例にとると、1番の項目のみが負の値になっている。実際、この項目の内容（国家の3要素）について、私はほとんど授業で触れていなかったことにテスト終了後に気づいた（上のviに該当する）。（1年5組）（7組）が担当したクラスでは、同一項目の弁別率がいずれも正の値だから、私と他の2人との違いが顕著であり、弁別率を見るだけで私の授業の改善すべき点を具体的に把握できたのである。このように、負の弁別率を有する項目は、指導者によって差が出やすい。

負の弁別率を有する項目をテスト実施前に排除できないのか、と言えばそれは「やってみなければわからない」以上、無理だとしか答えようがない。ただ、テスト設計図に基づいて、担当教師たちが入念に事前協議を進めれば、限りなくゼロに近づける

ことはできる。直前になってあわててテストを作成している場合は、協議の余裕もないから、ほとんど発見できない。ある程度ゆとりがあっても、過去 20 年間の私の狭い経験では、定期テスト 1 回につき弁別率が負の値を取る項目は 1～3% 発生してしまう。だからこそ、そこにテストと授業の改善点を発見することもできるのである。

エ. まよわし分析を行い、改訂すべき選択肢や授業の改善点を特定する 項目が含む不正解の選択肢をまよわし(distracters)という。各項目について、まよわしの選択肢が選ばれる状況を把握し、高い通過率や負の弁別率と組み合わせて、選択肢の改善点や授業の改善点などを調べることをまよわし分析(distracter analysis)という。具体例に即していくつか紹介したい。次の(1)～(3)の例をご覧いただきたい。

例(1) 項目番号 14						目標：内容の理解					
日本国憲法の特徴に関する次の説明のうち、正しいものを 1 つ選べ。(正解④)											
① 憲法には、「国民主権」「基本的人権の尊重」「平和主義」がその三大原理であると書いてある。											
② 憲法において、「戦争の放棄」と「恒久の平和」とは全く同一の意味内容をもっている。											
③ 憲法では、国会・内閣・裁判所の三権分立の原則が明確かつ厳格に規定してある。											
④ 憲法が前文や第 1 条で定めた「国民主権」は、制定後、数年間実現することはなかった。											
まよわし分析表 (*は正解番号)											
1 年 8 組 通過率 13% 弁別率 0.1						1 年 5 組 通過率 0% 弁別率 0					
	①	②	③	*④	無答		①	②	③	*④	無答
上位群	2	8	6	3	0	上位群	16	1	2	0	0
下位群	7	3	7	2	0	下位群	18	0	1	0	0
1 年 7 組 通過率 0% 弁別率 0						3 組全体 通過率 4% 弁別率 0.01					
	①	②	③	*④	無答		①	②	③	*④	無答
上位群	12	2	5	0	0	上位群	30	11	13	3	0
下位群	13	1	5	0	0	下位群	38	4	13	2	0

(1)の例は、通過率が低く、このため弁別率は上下の群で差がなく、いずれの群もまよわしを多く選んでしまった項目である(項目番号 22 も同様の傾向を示している)。このような場合、上の i～viiiの原因が考えられるが、生徒が集中して選択したまよわしを調べると原因を特定できる場合が多い。この例の「まよわし分析表」を見ると、5 組() 7 組()では正解者がゼロであり、2 クラスとも①のまよわしに誤答が集中していることがわかる。

「日本国憲法の三大原理」とは教科書には書いてあるが、憲法に書いてあるわけではない!ふだんの授業は条文に即して展開しているから、そのことにすぐ気づくはずである。ちなみに、憲法制定後 1956 年までの約 10 年間は「三大原理説」は実在しなかった。①を選んだ生徒たちは、条文に即した地味な授業をおろそかにし、教科書の本文だけを読み、かつ中学校時代の「公民」の予備知識だけを頼りに選んだのかもしれない(つまりviii)。あるいは/また、授業できちんと指摘しなかったことも十分に考えられる(vi)。

この項目 14 は、日本国憲法の成立から独立の回復までを主題としており、「1952 年

のサンフランシスコ講和条約の発効によって始めて、日本は独立すなわち主権を回復した」(だからそれまで国民「主権」といくら騒いでも、それはただの絵に描いた餅にすぎなかった)という、基本的事実の理解度を試している。だが、上の表の結果は、中学時代までの、そして「現代社会」の教科書も継続している「憲法の三大原理説」という刷り込みから脱皮させる授業に3人とも失敗したことを如実に示した。そして、ここに授業の重大な改善点があることも、まよわし分析から知ることができる。

例(2) 項目番号 17		目標：思考力・判断力			
大日本帝国憲法と日本国憲法における天皇について説明した次の文のうち、誤っているものを次の中から1つ選べ。(正解①)					
① 大日本帝国憲法では天皇が統治権を総攬するとともに軍の統帥権を有したため、統治を各大臣に輔弼させることはあっても、軍の権力を輔弼させることはなかった。					
② 天皇の地位は日本国憲法において日本国および日本国民統合の象徴に変更されたため、国事行為を除いて国政に関する権能を有しないことになった。					
③ 天皇の「人間宣言」とは正しくは「新日本建設に関する詔書」であって、大日本帝国憲法においても天皇自らも現人神と宣言したことはなかった。					
④ 日本国憲法において天皇を「象徴」としたのは、GHQスタッフの提案であり、バジヨットの『英国憲政論』にヒントを得たものであった。					
まよわし分析表 (*は正解番号)					
1年8組 通過率 42% 弁別率 0.2		1年5組 通過率 21% 弁別率 0			
*①	②	③	④	無答	
上位群	9	0	2	8	0
下位群	7	0	10	2	0
1年7組 通過率 13% 弁別率 0.1		3組全体 通過率 25% 弁別率 0.1			
*①	②	③	④	無答	
上位群	3	0	9	5	0
下位群	2	1	11	5	0

(2)の例では、まず8組を見ると、通過率・弁別率とも特に問題はないが、1人も選んでいないまよわし(②)を含む項目になっていることに気づく。これは、まよわしが有効に働いていないことを意味し、一般的には、そのまよわしを削除するか改訂する必要があることを示唆している。まよわしの言い回しが不適切だったり、受験者にとってもっともらしく見えなかったりするとき、よく現れる現象である。しかし、このことは、当のまよわしが絶対的に無価値であるということの意味しない。全く別の項目にまよわしとして挿入すれば有効に働く可能性は残っているし、当のクラスで選択されなかっただけで、他のクラスでは有効に働く可能性もわずかに残っている。だから速断は禁物である。実際に調べてみると、②を選んだ生徒は5組と7組に1人ずついる。いるにはいるが少ない。つまり、②を削除または改訂したほうがよいという、8組のまよわし分析表が示唆していた方向がおよそ正しかったことになる。

今回の研究では8クラス全体の3クラスをていねいに調べたからここまでわかるが、通常ならサンプルは1クラスだから、ここまでうまくはいかない。しかし、それでもこの程度はわかる、という典型例である。

項目番号 17 の②は、象徴としての天皇に関する、あまりにも基本的な理解事項（正しい説明）だったために、目標の「思考力」を求めるほど生徒をひきつけなかったわけである。だから、②はたとえば次のように改訂するとよいかもかもしれない。

- ② 大日本帝国憲法において、天皇は元首であったが、日本国憲法においても、外国大使及び公使の接受が定められているため、元首としての地位は失われていない。

この改訂した②を挿入したテストは実施していないが、項目全体としてはかなり難しくなるため、やってみる価値はある（改訂した②を挿入しても正解は①）。

なお、③について付言すれば、「現人神」は加藤玄智（1873～1960）が昭和 16 年（1941 年）に言い出したもので、それ以前には実在しない表現である。大日本帝国憲法第 3 条にいう「天皇は神聖にして侵すべからず」というのは、天皇が神のごとき尊い存在であるとか、現人神であるという意味ではなく、いくつもの欧州憲法が示すような、国際法上の「無答責」の原則（政治上の責任を負わないという原則）の表明であって、伊藤博文が最も腐心した点でもあり、授業で十分に配慮すべき事柄であろう。

例(3) 項目番号 8		目標：思考力・判断力				
国民主権に関する次の説明のうち、正しいものを 1 つ選べ。（正解②）						
① アメリカは、独立宣言に表明した人権思想に基づいて、国民主権を基調とする民主主義的な憲法を制定した。						
② イギリスの混合政体をモデルとして三権分立を唱えたモンテスキューは、国民主権を全面的に否定した。						
③ 法の支配とコモン＝ローの伝統を受け継ぐイギリスは、実権のない英国国王のもとで国民主権を憲法に定めている。						
④ 国民主権と議会制民主主義は、日本国憲法前文にも明らかにされているとおり、「人類普遍の原理」である。						
まよわし分析表（*は正解番号）						
1 年 8 組	通過率 53%	弁別率 0.3				
	①	*②	③	④	無答	
上位群	2	13	4	0	0	
下位群	3	7	3	6	0	
1 年 7 組	通過率 8%	弁別率 -0.2				
	①	*②	③	④	無答	
上位群	9	0	2	8	0	
下位群	5	3	5	6	0	
1 年 5 組	通過率 8%	弁別率 0.1				
	①	*②	③	④	無答	
上位群	3	2	5	8	1	
下位群	9	1	4	5	0	
3 組全体	通過率 35%	弁別率 0.1				
	①	*②	③	④	無答	
上位群	14	15	11	16	1	
下位群	17	11	12	17	0	

(3)の例では、8組では通過率・弁別率とも標準的であり、しかもすべてのまよわしがバランスよく選択されており、無駄なまよわしが少ない。しかも、まよわしは上位群よりも下位群のほうで多く選んでおり、どのまよわしもほぼ有効に働いている、と言ってよい。サンプルクラスが1つならこれで分析作業は終了する。

だが、7組に目を転じると、通過率が極端に低くだけでなく、弁別率が負の値を取っており、さらに上位群の正解者がゼロであることに気づく。5組を見ると弁別率こそかろうじて正の値になっているが、通過率の低さなどは7組に共通している。これ

をどのように解釈したらよいのだろうか。

項目番号8の主題である「国民主権」は、イギリス・アメリカ・フランス・日本と比較すると、非常に誤解を招きやすい、その点で慎重に授業を行なうことが求められる内容なのだが、3人ともばらばらに指導したようである。国民主権に関する基本命題をいくつかあげると、

- i. 法の支配とコモンローを基調とするイギリスは、国民主権を否定している（だから③は誤り）。法の支配を伝統としている以上、国民主権を認めることはありえない[なぜなら、国民主権は「法の支配」を否定する「人の支配」を意味しているから。この点で、市販の資料集にある法の支配の説明は、ほとんどいいかげんである]。
- ii. フランスが国民主権(=人民主権)を唱えて革命を起こしたが、そこでは国民主権が守られたことは一度もなく、待ち受けていたのは恐怖政治だった。
- iii. アメリカには国民主権という概念がない。合衆国憲法は、ジェファーソンの独立宣言とジョンロックの社会契約説を完全に排除し、法の支配(コーク)とコモンロー(ブラックストーン)という英国憲法思想のみから成立した。独立宣言と合衆国憲法には何ら連続性がない（だから①は誤り）[にもかかわらず、資料集などでは連続性があるかのような年表を掲げている]。
- iv. 以上から、国民主権が「人類普遍」（いつでもどこの国にでもあてはまる）の原理になれるわけがない（だから④は誤り）。
- v. モンテスキューは、『法の精神』で国民主権を否定し、イギリス流の権力分立体制（立憲君主制）を支持して三権分立を唱えた。合衆国憲法を起草したハミルトンらの『ザ・フェデラリスト』は『法の精神』を下敷きにして書かれた（だから②が正しい）[モンテスキューを読めばすぐわかるこれらの趣旨を明記してある資料集は、残念ながら皆無である]。

このi～vのうち、太字部分は特に誤解されやすい部分である。われわれ3人は共通の授業プリントを用いて授業を展開し、クラス間の成績傾斜を防止しようとしていたが、今触れたような点に深く踏み込んだ打ち合わせはできていなかった。というよりも、共通の授業プリントがあるから、打ち合わせを要しないほどの指導ができるはずだとの暗黙の了解があった。だから、この項目の分析結果は、「国民主権」の用語に限定しても、この概念から育成する思考力について、複数教師による指導体制の改善を具体的に促している、と言えよう。

この点は、シラバスを作成しても未然に防止できるわけではない。なぜなら、未然防止すべき点は山ほどあり、ほぼすべてを未然防止できるほどのシラバスを作成するとすれば、膨大な時間と労力を注いだ授業との違いが見出せなくなるからである。それなら先に授業をやって、テスト後の項目分析で点検・改善したほうが能率的なのだ。

(3) その他

ここまで、テスト設計図と項目分析の活用法などを個別に説明してきたが、この2つを組み合わせた場合の活用法の事例を2つ紹介しておきたい。授業の改善に直接結びつくわけではないので、あくまで参考にしていただきたい。

ア. テストの長さの妥当性 テストの出題項目数をテストの長さという。テストを実施するとき、あるいは実施する前に、生徒が時間を持て余したり、反対に問題が多す

ぎて時間不足になったりしないか、と心配になるものである。テスト制限時間で最適な問題の量はいったいどれくらいなのか。つまり、テストの長さの妥当性とは何か。

このような疑問に答える実践的な研究は、ほとんどない。私の数少ない研究では、多肢選択項目において、「知識問題は、読み取り後に正解と思う選択肢を選び、解答用紙にマークを完了するまで、一般に、1項目当り20秒、理解や思考力などの説明文問題は、同様に、1項目当り約55秒」と判定している。教室場面で練習問題を解答させたときの（北陵高校で得た）経験的な数値である（この数値が学校によって多少変わるかどうかは研究していない）。この数値を積み重ねると、テストの長さの妥当性が測れるのである。

テスト設計図を見ると（問題冊子でもよいが）、知識問題は14個、説明文問題は42個あるから、56項目全体の見込み所要時間は

$$20 \text{ 秒} \times 14 + \text{約} 55 \text{ 秒} \times 42 = 4 \text{ 分} 40 \text{ 秒} + \text{約} 38 \text{ 分} 30 \text{ 秒} = \text{約} 43 \text{ 分} 10 \text{ 秒}$$

と計算できる。北陵高校のテスト制限時間は、正味50分と厳格（すばらしい！）なので、見直しの時間を含めれば、5分強のゆとりがある、と事前予測できる。読書スピードには個人差があるので、経験的にはこれくらいのゆとりを持たせたテストの長さが妥当と思われる。それが本当かどうかは、テスト実施直後に試験監督の先生に質問し、生徒の受験の様子を確認すればよい（実際に確認すると、「ほぼ良好」「一部を除いてちょうどよい問題量だったと思う」との回答をいただいた）。だが、もう1つ確認の仕方がある。

項目分析原票を見ていただきたい。「*」の無答の状況を調べるのである。8組の例を見ると、上位群には、833番の生徒に無答（10個）が多く、下位群には826番と824番（同7個、10個）に多い。設問がすべてマークシート方式だから、あて推量で全項目に回答していてもよさそうだが、「正直者の多い北陵高校」ではそうになっていない（もっとずるくなってほしいものだ！）。

上位群で無答が多い生徒がいる、というのは珍しい。一般には、下位群で2～3名である（サンプルクラスの5組と7組では無答は下位群2名に集中している）。経験的には、この程度ならばテストの長さがほぼ妥当と判断してよい。これ以上短くすると、上位群の生徒たちはテスト時間を持て余すようになる。

無答分析はこれに尽きない。無答の多い生徒の学習習慣や学習態度、テストの受け方などについての改善策を指導できるようになるのだ。個別指導対象者をピックアップできるということである。上位群にしながら無答の多い生徒は、読書スピードが遅く、物事に慎重すぎる場合が多い。迷ってばかりでなかなか決断が下せない、内気な生徒でもよく見られる。練習問題を数多くこなすことが直接の改善策だが、時間配分を考えてテストを受けよ、という指導でもある程度改善できる。

下位群で無答の多い場合は、読書スピードに難があるのは当然だが、それ以上に日頃の学習態度に問題のある場合がほとんどで、他教科の成績も低迷しているはずである。だとすれば、科目指導以前の生活指導から再出発しなければならない。生活リズム・ノートの取り方・予習をして授業に臨む態度など、改善策の「入り口」はどこにでもある。そしてどこから入ってもよい。

イ. テスト平均点予測の方法 テスト設計図と項目分析原票を活用して、テスト実施前に、そのテストの学年平均点をほぼ正確に予測することができる、という珍しい研究を昨年度われわれ3人は行った。Nedelsky法、Jaeger法、Angoff法、Ebel法などのうち、代表的な後二者を徹底研究し、米国の学者たちの見解に反して、Ebel法が最も優れていることをわれわれは見出した。この予測が可能になれば、テストの難易度や得点分布をある程度事前にコントロールできることになる。中間考査と期末考査で平均点を調整したり、選択科目同士で平均点が揃うように調整したりできるようになるため、きわめて実用的な知見と言えよう。長くなるので説明は省略し、別稿に譲る(「平均点予測の可能性」とのタイトルでレポートをまとめ、昨年度末に学力向上委員会に報告してあるが、先生方全員への報告はまだである)。

2. 授業の評価

(1) 授業評価の目的

ア. 授業評価は動的なものである 授業評価を通じて授業を改善する、などと聞いたら先生方は驚いたり怒り出したりするかもしれない。冗談じゃない、オレの授業を査定されたり値踏みされたりしてはかなわない、と。だが、ここには誤解と誤用がある。この風土が授業評価を困難にしている。授業評価は教師の査定を目的とするものではない。授業評価活動を通じて、授業の技術や内容の改善点を調べ、良き技術や内容を財産として蓄積する一方、改善すべき点を特定して、それを克服する道を模索することに目的がある。学者でないわれわれにとって、テストや授業の改善という実践に結びついたとき、初めてテスト設計図や項目分析の研究が意味を持つと同様に、管理職でないわれわれにとって、授業改善に結びついたとき、初めて授業評価は意味を持つ。授業や教師の現状を固定して見てはならない。授業は変化し、向上し、改善されていく、という動的な見方をしない限り、授業評価は不可能である。

イ. 授業はテープ音声や手料理に似ている ところで、テープレコーダから流れてくる自分の声を初めて聞いたとき、「どうも違うな」というショックを受けた経験はないだろうか。「俺の声はもっといいはずだ」と。授業もこれに似ている。授業を行う教師の「つもり」と授業の受け手の生徒の「印象」との間にはズレがある。このズレは家庭料理のそれにも似ている。授業も手料理も、作り手の意図が必ずしも消費者に届かないのである。めったに「まずい」とは口に出せないし、料理にこちらの舌を合わせようとするし、公開されないまま食べ終わったら証拠が消えるので問題が大きくなり、ひとまず安心できるのである。残るのは「味の記憶」であり、母親の愛情である。

ウ. 授業評価はショックから始まる テープの自分の音声と同様、自分の授業をビデオで初めて観たとき、大きなショックを受けるものである。私の体験で言えば、自分とそっくりな人間が画面の向こうで授業をしているのだが、「どうも違う」のだ。要するに授業があまりうまくないし、やってはいけないことをいろいろやっている様子が露骨に映し出されているという実感なのである。そして最後に自覚する。「ああ、情けないが、やっぱりこれはオレだな」と。これは録画体験した者にしかわからない。

だから、授業評価の理屈はとりあえず横において、まず録画し、それを自分で観ることをお勧めしたい。かなり自信のある内容の授業でも、たいていの人はがっかりす

る。そして、このショックから授業評価・授業改善は始まる。何度か眺めるうち、自分の授業の欠点をカバーする対策が浮かぶ。長所も見えてくる。明日からはここを改善してみよう、という方針の立つ箇所もわかってくる。まだある。

エ. 授業録画のメリット かつては忙しい合間を縫って授業観察をし、記録を取り、まとめる授業評価が主流だった。教室の後ろで多人数が観察するのも大変だった。「単に仕事を増やすだけ」だったと言う人もいる。「公開授業週間」が嫌われがちな理由にもなっている。だが、ビデオ録画し、DVD コピーの配付が容易になった今日、ほぼすべての心配や不満から解放された。もう、狭い教室の後ろに立ち並ぶ必要はない。授業変更の依頼をしてまで観察に出かける必要はない。見たかったのにと悔しい思いをすることもない。

ビデオ録画した授業の場合、①自宅で自分の授業を観察できる（自己評価）、②同僚数名と、寝転がってでも酒を飲みながらでも、自宅で相互評価ができる、③時間や場所に制約されない、④同じ場面を何度でも巻き戻して観察できるし、一時停止も容易だ、巻き戻しても一時停止しても、授業者は文句を言わない、⑤数年、十数年後にでも観察しなおすことができる、などメリットが山ほどある。デメリットはむしろ、記録が残ってしまうことぐらいだ。教育実習生にもやってみてはどうか。

私の体験としては、自分の授業を観察したとき、その欠点をいやというほど自覚したため（自己評価）、二度と同じ過ちを犯さなくなったのが最大のメリットだが、授業評価表を用いて同僚教師に評価してもらうメリットも大きい。自分よりも客観的な評価資料を提供してもらえらるからである。しかし、それは最初のショックから立ち直ってからにしたほうがいい。

ともかく、授業は変わる。新卒なら、3日・3週間・3ヶ月・3年くらいのサイクルで質的に転換する（向上する）。北陵高校では、進学指導や受験講習の体験が、質的転換に大きく貢献する。大学受験に通用するレベルの高さと授業の知的な面白さの両立を工夫するようになる。さらに経験を積むと、40代後半にマンネリから抜けられない自分の壁に気づく。そのまま50代に入り、脱皮できなければ地獄である。マンネリ授業を自覚したとき、質的転換のチャンスである。授業の自己評価というショック療法は特に効く。授業に満足したとき、転落は始まっている。自戒を込めて言う。

（2）授業「技術」の評価と授業「内容」の評価の区別

ア. 授業評価の方法を確立する 授業評価を困難にするもう1つの理由は、評価方法が必ずしも確立していない点にある。授業評価を実施しようとする場合、「授業の評価はそもそも客観的に行なうことが不可能であるし、たとえ可能でも、教科にはそれぞれ専門性があるから、教科の壁を越えて授業評価をすることなど不可能だ」という反対が寄せられるかと思う。この反対は、授業を評価されたくないという第1の理由と連動してもいる。

なるほど理論家が提唱する方法は、「今すぐにでも役に立つ」具体性・実用性に欠けるものが多い。中には、「主役の子供が授業評価する」と称して、評価の訓練を受けていない生徒を評価の主体とってしまう悪質なケースも見られる。特に後者のやり方が主流になれば、教師は人気取りの「芸者」になり下がり、生徒の興味をある程度犠牲にしてでも教える事項を抱えた教師の立場は切り捨てられる恐れがある。

イ. 技術と内容を区別する このため、授業評価の方法は、授業技術の評価と授業内容の評価に分けて論じるとさまざまな誤解から解放される。たとえば、板書が整理されているとか、話術に長けているなどの授業技術の場合、教科の専門性は不要である。授業技術に関する限り、教科の垣根を越えて相互に評価が可能である。実際、他教科の授業観察のほうが、技術を学び取るチャンスは高く、教えられる点が多い。また、授業技術は、スポーツの技術のように、ひとたび身につけると、レベルが低下しにくくなる。私の場合、「黒板を見ながら説明する」という悪しき習慣（生徒に声が聞こえない、生徒の様子を見落とす）から抜け出し、「生徒を見渡しながらか説明する」ことができるようになった。

ウ. 教科の専門性も重視する これに対して、授業内容の評価はできるだけ同一教科に任せたほうがいい。教科の専門性を尊重するのである。ただし、内容の評価には、目標（＝評価の観点）ごとに論じるとか、内容と目標の関連づけを論じるという、新たな視点が必要である。このことによって初めて、授業の「目標」が空念仏でないことに気づくし、これがテスト設計図とテスト問題の作成作業に直結する事柄であることにも気づくようになる。授業と評価と目標は三位一体のものなのである。

エ. 評価の様式を確立する 以上の観点から、授業技術と授業内容の評価様式（書式）を工夫しておく必要がある。別添の様式2種類（資料3、4）は、以上の観点到配慮して作成したものである。このうち、授業技術の評価表は、20年近く前に私が作成し、利用してきた（同僚に評価していただいたこともある）もので、幾度か様式を改訂しながら今日に至った。各項目の立て方にはある程度経験的な裏づけがある。一方、授業内容の評価表は、今回の提言のために私が開発したもので、未使用である。だから自信がない。うまくつかえるかどうかはやってみなければわからない。添付した様式には補足説明を入れてあるが、実際に活用するのであれば、説明を省き、記入欄を拡張して記入しやすくするとよい。内容であれ技術であれ、開発済みの別様式があるのなら、それを用いてもよい。

ただ、どんな様式にしる、自分たちで少しずつ工夫を加えながら、自分の授業のこういう点を改善するのに役立った、という実感をつかんでいただければ、授業評価は大成功であろう。1年や2年で完結するものではなく、授業評価は数年間のサイクルを要する息の長いものである（テスト分析も同様である）。そんな風土が育ってくれることを願っている。