

測定・評価部門

学力低下論争, 目標標準拠評価の定着, 学力テストブームの狭間で

山 森 光 陽

(国立教育政策研究所)

はじめに

本稿では、2004年7月から2005年6月末を中心に発表された著書、論文等、および2005年度の日本教育心理学会総会発表論文ならびに隣接領域の諸学会において発表された論文等の中から、初等中等教育における教育評価に関わるものを中心にレビューを行う。

従来、『教育心理学年報』におけるレビューは、細分化された分野ごとに研究成果をまとめた総覧的なものや、あるひとつの分析・統計手法に特化して研究成果をまとめ、それらの問題点と今後の課題を論じるものが多かった。このようなレビューは研究者の関心を強く引くものではあるが、教育実践に直接関わる者にとっては、強い関心を抱くに至らないという問題があった。教育評価・測定は教育実践において日常的になされている営みであることを考慮すると、教育実践の側から見たレビューもまた必要である。

さらにいえば、教師をはじめとした実践者にとっては、自らが普段行っている評価のあり方を検討するための研究成果のレビューが望まれる。同時に、研究者にとっては、今後の研究の方向を検討するための研究成果のレビューが必要である。筆者は、これらを両立させることが、研究と実践をつなぐ、一つの方法であると考えた。

そこで、この1年の間に、学校を中心とした教育実践の場で起こった(降りかかってきた)問題を整理し、それらの問題解決に教育心理学がどのように寄与してきたのかを提示する。そのうえで、今後の教育心理学における教育評価研究の方向を見定めたい。

1. 学力低下論争, 目標標準拠評価の定着, 学力テストブームの狭間で

この1年の初等中等教育における教育評価の動きをまとめると、学力低下論争が2つの国際学力調査の結果の公表とともに一気に再燃した中で、学校においては、ようやく目標標準拠評価が定着しつつありながらも、学力テストブームに呑み込まれた1年であったといえよう。

1. 国際比較調査と教育課程実施状況調査の結果公表と学力低下論争

1998年の現行学習指導要領の告示から現在に至るまで、学力問題の論争が続いている。特に2004年12月のOECD生徒の学習到達度調査(以下、PISA2003)ならびに国際数学・理科教育動向調査(以下、TIMSS2003)といった国際調査の結果の発表をきっかけとして、学力問題が従前以上にクローズアップされるようになった。

これらの調査結果の概要をまとめると、TABLE 1のとおりとなる。これらの調査においては、結果の経年比較が可能となるように、得点が標準化されている。前回調査と比較して平均得点が下がったのは、PISA2003における読解力およびTIMSS2003における小4理科、中2数学であった。しかし、PISA2003における数学的リテラシーおよびTIMSS2003における中2理科の分野において、平均得点には変化がないにもかかわらず、参加国中における順位が下がっていた。また、平均得点以外にも問題となった点として、学習内容に興味をもつ児童生徒が少ないことや、家庭学習の時間が少ないことなどが挙げられる。

これらの結果から、わが国の学力が低下傾向にあることをはっきりと認識すべきであるという意見や、わが国の学力は世界トップレベルとはいえない状況であるとする見解が、新聞紙上をにぎわせることとなった。

しかし、2005年4月に公表された、平成15年度小・中学校教育課程実施状況調査の結果は、TABLE 2に示したとおり、前回調査と比べて正答率の上昇がみられた問題が多いほか、勉強が好きだと答えた児童生徒が増加傾向であった。だが、PISA2003の結果と同様に、全体的な傾向として読解力に難があるという結果であった。

このように、国際調査では「学力低下」傾向がみられたのに対し、国内調査では「学力向上」傾向がみられた。これらの結果の解釈は様々であるが、調査実施時期に特に留意すべきである。

TABLE 3は各種学力調査と学習指導要領との関係を示したものである。PISA2003、TIMSS2003はいずれも、学習指導要領の移行措置期間中ないしはその直後に実施

TABLE 1 国際学力調査の結果（国立教育政策研究所（2004, 2005a, 2005b）を参考に作表）

調査	対象学年 (実施時期)	分野	前回順位 (2000年)	今回順位	日本の結果の特徴
PISA2003	高1 (2003年7月)	数学的リテラシー	1位/31カ国中	6位/40カ国中	2000年調査と同様に1位グループ OECD平均に比べて上位の習熟度レベルに位置する生徒の割合が多い 2000年調査の平均得点と有意差なし
		読解力	8位/31カ国中	14位/40カ国中	8位の国（スウェーデン）以上と有意差あり OECD平均と比べて同程度 2000年調査の平均得点の方が有意に高い
		科学的リテラシー	2位/31カ国中	2位/40カ国中	1位グループ 2000年調査の平均得点と有意差なし
		問題解決能力	—	4位/40カ国中	1位グループ
		質問紙調査	—	—	数学で学ぶ内容に興味のある生徒の割合が少ない 数学を得意だと感じている生徒の割合が少ない 数学を日常生活にどう応用できるかを考えている生徒の割合が少ない 保護者の学歴や職業等が生徒の数学的リテラシー得点に与える影響は弱い 学級の雰囲気は良好な状態にある 数学の授業の雰囲気は良好である 宿題や自分の勉強をする時間がOECD平均より低い 平均得点は前回調査と有意差なし
TIMSS2003	小4 (2003年2月)	算数	3位/26カ国中	3位/25カ国中	算数の勉強が楽しいと「強くそう思う」児童の割合は国際的に見て低いレベルにあるが、前回調査と比べて有意に高くなっている
		理科	2位/26カ国中	3位/25カ国中	平均得点は前回調査と比べて有意に低い 理科の勉強が楽しいと「強くそう思う」児童の割合は国際平均を10ポイント下回るが、前回調査と比べて有意に高くなっている
	中2 (2003年2月)	数学	3位/39カ国中	5位/46カ国中	平均得点は前回調査と比べて有意に低い 数学の勉強が楽しいと「強くそう思う」生徒の割合は国際的に見て低いレベルにあるが、前回調査と比べて有意に高くなっている
		理科	4位/39カ国中	6位/46カ国中	平均得点は前回調査と有意差なし 理科の勉強が楽しいと「強くそう思う」生徒の割合は国際的に見て低いレベルにあるが、前回調査と比べて有意に高くなっている

TABLE 2 平成15年度小・中学校教育課程実施状況調査結果の概要

前回調査との同一問題の正答率の比較	有意に上回る問題数は全体の約43% 有意差がない問題数は全体の約39% 有意に下回る問題数は全体の約17% 特に、国語の記述式問題の正答率が前回調査を下回っている
前々回調査、前回調査との同一問題の正答率の比較	前回と比較すると上回る傾向 前々回と比較すると一部で下回る
大都市、都市、町村における児童生徒の平均正答率	大都市、都市、町村の間にほとんど差はない
質問紙調査の結果	勉強が大切だ、好きだと答えた児童生徒の割合が増加傾向 授業がわかると答えた児童生徒の割合が増加傾向 学校の授業以外でほとんど勉強しない児童生徒の割合が減少傾向

されたことからわかるように、これらの結果が現行指導要領による教育課程の影響を受けたものであると解釈するのは妥当ではない。同様に、平成15年度教育課程実施状況調査は現行学習指導要領の完全実施から2年後に実

施されているため、これをもって現行学習指導要領の効果と解釈するのもまた拙速であろう。さらに、1～2年の間で学力の回復が図られたとする解釈も適切ではない。また、国際調査の結果を解釈する際には、以下の点に留意すべきである。

第1に、国際調査における質問紙調査結果の解釈に関わる問題である。OECD（2004）は、質問紙調査に対する回答は、対象生徒の内省報告であるため、実際の学習における行動や態度を直接測定してはいないこと、回答の傾向にも文化差があることなどを挙げている。そして、参加国間における結果の比較を行うことには問題があるとともに、結果の国内比較は有益であることを指摘している。

第2に、国際調査結果の国別順位を問題とすることに関してである。PISA、TIMSSともに国別のランキングを得ること（相対評価）を目標としているのではなく、学習到達度を捉えることを目的としている。そのため、これらの結果を順位のみ注目して解釈することは慎まなければならない。特にPISAにおいては、得点と習熟度

TABLE 3 学習指導要領と学力調査実施時期との関係

		平成6年度	平成7年度	平成8年度	平成9年度	平成10年度	平成11年度	平成12年度	平成13年度	平成14年度	平成15年度
小学校学習指導要領 中学校学習指導要領		告示						移行措置	移行措置	完全実施	
テスト名称	対象学年										
PISA2000	高1	小4	小5	小6	中1	中2	中3	調査実施			
PISA2003	高1	小1	小2	小3	小4	小5	小6	中1	中2	中3	調査実施
TIMSS1995 (算数・理科)	小4	調査実施									
TIMSS2003 (算数・理科)	小4						小1	小2	小3	調査実施	
TIMSS1999 (数学・理科)	中2	小4	小5	小6	中1	調査実施					
TIMSS2003 (数学・理科)	中2		小1	小2	小3	小4	小5	小6	中1	調査実施	
平成13年度 教育課程実施状況調査	小5				小1	小2	小3	小4	調査実施		
	小6			小1	小2	小3	小4	小5	調査実施		
	中1		小1	小2	小3	小4	小5	小6	調査実施		
	中2	小1	小2	小3	小4	小5	小6	中1	調査実施		
	中3	小2	小3	小4	小5	小6	中1	中2	調査実施		
平成15年度 教育課程実施状況調査	小5					小1	小1	小2	小3	小4	調査実施
	小6					小1	小2	小3	小4	小5	調査実施
	中1				小1	小2	小3	小4	小5	小6	調査実施
	中2				小2	小3	小4	小5	小6	中1	調査実施
	中3				小3	小4	小5	小6	中1	中2	調査実施

注：TIMSSは各国における学年末に行うこととなっている。

レベルの対応表が作成されており、何点をとれば、具体的にどのようなことができるのかといった記述がなされている。

したがって、各種学力調査の結果を論じる際には、単純に得点や順位だけに焦点を当てるのではなく、これらの背景を理解するとともに、また関係するデータや各学校における現状等もふまえながら、「解釈」を行う必要があるといえよう。いいかえれば、測定即評価という態度ではなく、測定と評価を別物として捉える必要がある。なお、PISA2003の結果の解釈の要点は、村山(2005a)にまとめられている。

2. 学校における目標準拠評価の定着

指導要録の改訂が行われ、目標準拠評価(いわゆる絶対評価)が本格的に導入されて4年が経過した。当初は多くの学校での試行錯誤が行われたが、現在では、評価方法に対する理解が徐々に浸透してきていると考えられる。目標準拠評価によって児童生徒の学習状況の評価を行うこととなった理由をたどると、平成12年の教課審答申にいきつく。この答申のポイントは、以下の4つに整理することができる。

第1に、自ら学び自ら考える力などの「生きる力」の育成状況を総合的に評価する工夫の必要が述べられている。第2に、現行の学習指導要領においては基礎的・基本的な内容の確実な習得を図ることを重視しているため、学習指導要領に示された目標に照らし合わせて、その実現状況を見る評価を重視するべきであると述べられている。第3に、評価を行うにあたっては、知識や技能の到達度を的確に評価することはもとより、自ら学ぶ意欲や

思考力、判断力、表現力などの資質や能力までを含めた学習の到達度を適切に評価することが大切であると述べられている。第4に、日常の指導の中で評価を児童生徒の学習の改善に生かすという、いわゆる指導と評価の一体化が求められている。

知識や技能の到達度はもとより、自ら学ぶ意欲や思考力、判断力、表現力などの資質や能力までを含めた学習の到達度を評価するといった思想は、真正の評価(authentic assessment; Wiggins, 1989)という考え方を一にしている。真正の評価という考え方が出てきた背景として、テストだけで学力を捉えようとしても、子どもがどこまで真に理解し、どこからができないのかといった、発達の真の様相を捉えることができないという問題が認識されてきたということが挙げられる(山森, 2005a)。

そして、現在行われている目標準拠評価においては、分析的、記述的な評価ならびに多様な評価資料の活用などの工夫が求められている。このような状況において、学習行為そのものを対象としたパフォーマンス評価が広く実施されるようになっている(北尾, 2004)。

この、パフォーマンス評価を行うにあたって必要不可欠なのがルーブリックである。これは、学習活動レベルにまで落とし込まれた評価規準(達成目標)に対して、どのような行動が見られれば、どの程度の実現状況かを判断し、得点化するための基準である。

現在、ルーブリックを用いた評価の実践は数多く取り組まれている。具体的には、桑田(2004)による小学校社会科における実践のほか、高浦(2005a)をはじめとした小・中学校の全教科における実践、国立教育政策研究所

(2005c)による総合的な学習の時間における実践の報告などが挙げられる。

パフォーマンス評価においてルーブリックを用いる際には、これをあらかじめ設定することが必要(北尾, 2004)であり、ひいては目標標準評価の客観性を保障することにつながる(田中, 2005)ことが指摘されている。さらに、いわゆる指導と評価の一体化が適切に図られるという利点が強調されている(高浦, 2004)。このように、ルーブリックを用いた評価についての研究は、主に教育方法学において取り組まれてきている。

3. 地方における学力テストブーム

先に挙げた各種調査の結果とともに今年度注目を集めたのは、学力の格差問題である。苅谷・志水(2005)は、親の社会経済的地位が子どもの学力(ペーパーテストで測定された学習到達度)に加えて、学習時間や学習意欲にも影響を与えるばかりか、これらの階層差が拡大していることを主張している。さらに、学力低下問題が国民の関心を集めるようになった現在、わが国は「学力調査の時代」に入ったことを指摘している。

この指摘を裏付けるかのように、2004年度には、47都道府県および13政令指定都市中、39都道府県および11指定都市が独自に学力調査を行った。そのうち、31の自治体が悉皆による調査を行っている。

さらに、東京大学大学院教育学研究科基礎学力研究開発センター(2005)が、指定都市を除く市区町村、全2,949自治体を対象とし、1,335自治体から回答を得た調査によれば、2000年度以降に独自に学力調査を実施した自治体は390であった。これらの自治体のうち、9割以上が悉皆調査を行っている。

II. 教育評価をめぐる教育政策の展開

このような流れの中で、文部科学省も学力テスト実施の必要があるとの見方を示すようになり、悉皆による全国学力調査を行うことを決定した。2006年度には問題作成等が行われ、2007年度には小学6年生(国語, 算数)、中学3年生(国語, 数学)を対象に全児童生徒が参加できる規模で実施される予定である。

また、各学校においては、学校運営の状況についての自己点検・評価を行うことが定められた2002年度の小学校設置基準等の改正を機に、ほとんどすべての学校で自己点検・評価が行われている。さらに、自己評価に加えて第三者評価も求められるようになってきており、文部科学省においては、学校評価システムの構築に向けた研究や、第三者評価に関する研究などにも着手されようとしている。

III. 教育評価研究を概観する視点

先に述べたように、学校は、学力低下論争、目標標準評価の定着、学力テストブームの狭間の中でもがき苦しむことを余儀なくされているのが現状である。このような現状において、学校に求められていることを整理すると、以下のとおりとなろう。

第1に、学力向上を目標とした日々の指導における工夫と改善を行いながら実践を行うとともに、子どもの学習状況の評価を適切に行い、説明責任を果たすことが求められている。

第2に、授業時数の捻出が困難な中、国や地方による学力テストを実施することとともに、学校運営に対する自己評価と他者評価にも取り組むことが求められている。

教師や学校に対するこれらの要求をふまえ、教育心理学における教育評価研究を分類するならば、以下3つのカテゴリーに分けられよう。

第1に、子どもの学習とその評価についての研究である。具体的には、学習到達度の評価方法、評価方法が学習行動等に及ぼす影響についての研究などが挙げられよう。

第2に、子どもを取りまく環境に対する評価についての研究である。具体的には、授業評価や教員評価、学校評価の方法についての研究などが挙げられよう。

第3に、評価理論や手法についての研究である。具体的には、学習到達度の経年変化や発達的变化を捉える方法や、大規模データの分析方法、さらには評価そのもののあり方についての研究などが挙げられよう。

また、現在の教育評価研究の主たる関心事の一つである、全国レベル、地方レベルの学力テストについての研究は、子どもの学習とその評価についての研究としてもカテゴライズ可能ではあるが、独立したセクションを設けて検討したい。

以下では、これら3つの分類に沿って、この1年を中心とした期間に公表された研究を概観する。

IV. 子どもの学習とその評価

1. 評価と学習行動

村山(2004)は、中学生を対象に実際の学校における学習場面に近い場面における実験を行った。その結果、授業後の確認テストが毎回記述式で行われる群では、空所補充型テストが行われる群と比べてマイクロ理解方略やマクロ理解方略といった「深い処理」の方略を多く用いるようになるとともに、暗記方略などの「浅い処理」の方略の使用頻度が少なくなることを示した。

さらに、村山(2005b)は、大学生を対象とした実験を行い、テスト形式の予期が方略を変容させるメカニズムに

ついて検討を行った。その結果、学習者はテストの課題要求を能動的に分析し、その結果に基づいて合理的に学習方略を変容させているのではないことが示された。そして、テストの経験に照らし合わせて、テストで困難を感じ、かつ困難の原因を学習方略に帰属させることを経て、学習方略を変容させていることが示唆された。この結果は、単純にテスト形式を変えたからといって学習者の学習方略は容易に変容しないことを意味している。

関連して、村山(2005c)は、中学生を対象とした実験を行い、学習者に暗記方略を促進してしまうと考えられている空所補充型テストであっても、その実施前に意味理解方略を用いることがテスト得点を高めるというように、学習者のもつテスト形式スキーマへ介入することが、暗記方略使用を抑制させることを示した。

これら一連の研究は、テスト形式が学習方略の使用に影響を与えるものの、それだけが学習方略使用の変化を促すことはないことを示しているとともに、学習方略そのものもまた学習させることが可能であることを示している。

岡・松山(2005)は、小学4年生の算数においてドリル学習を行った際の評価が計算力や自己効力感に及ぼす影響について検討を行った。その結果、解決方法(繰り上がりに気をつけて計算することができた、など)について自己評価を行いながらドリル学習を行った場合に計算力の伸びが大きいことを示している。この結果は、先に示した、目標準拠評価を適切に実施するための方法の一つとしてのルーブリックを利用した自己評価を行うことが、学習成果を高めることにつながることを示唆している。

ルーブリックに関しては、2005年度の日本心理学会公開シンポジウム(鹿毛・大村・志水・市川・松下・無藤・北尾, 2005)において、北尾が、ルーブリックを用いた評価についての教育心理学的研究を主張している。

ルーブリックが学習行動に及ぼす影響については、現在のところ、ルーブリックを用いて実際の指導および評価を行った際に、子どもの学習意欲が高まったと思われる事例が報告されている(四日市市立教育センター, 2003; 山森, 2005b)。今後は、ルーブリックを用いた評価が学習行動に及ぼす影響について実証的に検討した研究が求められるだろう。

自己評価、相互評価活動についての研究では、廣瀬(2005)は、大学の情報教育においてプレゼンテーションスライドを学生につくらせる過程において、学生同士の相互評価を行い、他者からの評価をふまえて作品を作成させた。その結果、最終作品のレベルアップが図られたことを示した。

また、Inuzuka(2005)は、中学生を対象に、他者が書

いた文章を評価する活動を取り入れることによって作文のパフォーマンスが上がるかどうかを検討した。その結果、個人で評価活動を行った後にグループで話し合うという条件で、作文のパフォーマンスが向上するという結果が得られた。

わが国の学校教育の特徴の一つに、生活集団と学習集団が同一である学級を単位として授業が行われることが多いため、子どもどうしの学び合いが成立しやすいことが挙げられる。そのため、子どもどうしの相互評価といった活動は取り入れやすい。日頃の学習活動において、自己評価を適切に行わせながら、相互評価を取り入れることは、学習成果を高めるための方略としても有効に機能するだろう。

さらに、学習指導における「ことばがけ」もまた評価活動の一つであるといえよう。名取(2005)は、少年サッカーチームに所属する児童を対象に調査を行った結果、指導者が否定的なことばがけをしても、肯定的な理由からそのような言葉が発されていると子どもが認知したときには、サッカーに対するやる気が高まっていることを示している。この結果は、評価する側とされる側のラポールの形成が重要であることを示唆しているといえよう。

2. 評価方法・項目

吉川・岸(2005)は、意見文を評価する際の評価項目について検討を行った。その結果、適切に評価を行う際には、学習指導要領を反映させた評価項目だけではなく、その子らしさ、伝わりやすさを評価する項目を入れることによって適切に評価されることを示した。この結果から教師が学ぶべきことは、意見文の適切な評価を行い、指導に活かすためには、学習指導要領の内容だけにこだわった評価を行うのではなく、題材に応じた評価項目を適切に設定する必要があるということだろう。

笹山・岸(2005)は、説明文読解テストの出題形式が問題の難易度に与える影響について検討を行った。項目応答理論を用い、学年差を反映する問題項目の特性および学年ごとの能力推定値の高い出題形式について分析を行った。その結果、文章全体を視野に入れた回答を要求することが学年差を生むこと、中心語句の抽出の出題を行った場合、学年を通して多肢選択式で出題することが最も能力推定値の精度が高いことが明らかになった。

この結果は、テストの作成にあたっては、子どもの発達段階に応じた形式で出題を行う必要があることを示唆している。この研究は説明文の読解テストだけを対象としているが、今後、同様の研究が各教科を対象になされれば、学年ごとの最も適切な出題形式のインベントリーが開発されよう。そして、教師が自らテストをつくる際、

また大規模な学力テストの開発の際に役立つ資料が得られる可能性がある。さらに、テスト作成者の出題形式に対する意識を高めることにもつながるだろう。

ところで、2006年1月実施の大学入試センター試験からリスニングテストが導入される。わが国ではリスニングテストについての知見が少ない現状があったが、大学入試センターではこの新しい取り組みに対して、精力的に研究が進められている。

内田・中畝・荘島(2005)では、リスニングテスト実施時に騒音が与える影響について検討が行われた。さらに、騒音によって生じた試験成績への影響の除去を目的として、項目応答理論を用いて得点の調整を行った場合、良好な補正結果が得られることが明らかになった。また、内田・中畝・石塚(2004)では、リスニング問題の出題順序や話者の連続性について検討が行われた。その結果、これらの要因と試験成績の間に関連はみられなかった。

V. 子どもを取りまく環境に対する評価

1. 授業評価

各大学での教育内容・授業方法改善の動きの高まりを背景に、特に大学において行われた授業評価の研究が多い。服部・上殿・宇都宮・清水・徳田・本橋(2005)は、学生による授業評価における評定の個人差と評定平均値を、マルチレベルアナリシスを用いて検討した。その結果、授業の評定平均値を解釈する際には、評定の個人差を見逃すことができないことを指摘している。

この結果は、授業評価の評定平均値だけを一律にならべ、他の授業との比較を行うだけでは不十分であることを示している。マルチレベルアナリシス(または、階層線型モデル)は、集団単位の変動要因と個人単位の変動要因の両者を一度に扱うことができる方法であり、諸外国ではこの方法を用いた研究が多くみられる。

学生の自己評価の結果をそのまま授業評価として用いることに対する問題提起を行った研究として、牧野・西浦(2005)が挙げられる。この研究では、授業にほとんど出ていない一部の学生は、授業に出ている学生と比べて授業を低く評価することを明らかにした。また、遠藤(2005)も、学生の自己評価のみに依拠した授業評価には注意を要することを指摘している。

そのほかに、大学における授業評価の研究としては、学生による授業評価と自己効力感・テスト不安の関係を検討した西浦(2005)や、授業態度および授業評価と授業の満足度との関係を検討した中嶋(2005)、学生が授業評価をどのように捉えているかを調査した増田(2005)などが挙げられる。

ここで筆者は、大学における授業評価の研究に対して

問題提起をしておきたい。先に述べたように、多くの大学が学生による授業評価を実施している現状を反映して、教育心理学会総会でもこの分野の研究発表が多い。これは、データが比較的容易に得られるという事情にも起因しているのだろう。

しかし、その多くは授業評価結果の集計に簡単な考察を加えただけのことが多い。このような研究の結果は、教育心理学の理論等に照らし合わせた考察がなされていないため、結果から教育実践への含意を導き出すことができない。そのため、結果を見た第三者が、他文脈において知見を援用することができないだろう。

また、大学において制度化された授業評価に対して、教育心理学的な手法を手続きに入れることはできなくても、授業評価の結果以外の変数を扱うなどして、教育心理学の知見を取り入れた結果の解釈を行う必要があるだろう。

最近では、小・中学校、高等学校においても児童・生徒による授業評価に取り組んでいることが多い。野村(2004)は、高知県の高等学校における授業評価の取り組みを紹介し、授業評価の結果が指導の改善に結びついていない実態や、取り組み自体が不十分であることを指摘している。

ところで、授業を評価する際には、児童・生徒・学生だけが評価を行うのでは不十分であり、多様な評価資料を収集し、結果を分析することによって総合的に評価がなされなければならないだろう。教育心理学においては、授業や学級の構成もまた研究対象とされることが多いが、そのために開発された尺度を利用することや、参与観察の仕方などは、授業の第三者評価にも利用できるだろう。

たとえば、大道・野嶋(2005)では、授業風景を撮影したビデオ映像を第三者評定させることで、クラス間における目標構造の差を検討している。このような研究の手続きは、第三者の観察による授業評価を行う上で参考となろう。また、角谷・無藤(2005)は、小学5年生から中学3年生までを対象としたアンケート調査を行い、科学的興味や追求意欲を高める理科授業の意義を検討した。このような研究で得られる因果モデルから、具体的にどのような項目で授業評価を行えばよいかのヒントが得られるだろう。

2. 教員評価・学校評価

教員評価や学校評価についての研究は、教育心理学ではほとんど取り込まれていないのが現状である。棚上・淵上・田中(2005)による教師集団の効力感尺度作成に関する研究は、学校を教師がチームを組んで職務を遂行する場として捉え、集団効力感に着目して尺度作成を行い、分析を行った。その結果、教師集団の効力感の高い教師が、学校組織コミットメントも高いことが明らかとなっ

た。このような研究は、教員評価や学校評価に適用可能であると考えられる。

現在、学校評価がほとんどの学校で行われるようになったが、その評価資料を収集するために、学校では頻繁にアンケート調査が実施されている現状がある。そして、調査が多すぎるために結果の解釈が困難という状況を生み出している。

評価を実施する際には、できるだけ少ない労力で実施できるように工夫する必要がある。また、改善のポイントとなる項目に絞って評価を行い、評価結果を組織や業務の改善に結びつけることが必要である。なお、教員評価の現状と課題については、八尾坂(2005)に詳しい。

VI. 評価理論や手法についての研究

1. 日本テスト学会とテストスタンダード

2005年3月に、『日本テスト学会誌』が創刊された。池田(2005)は巻頭言の中で、現在のテスト研究を取りまく現状を概観し、自己理解を助けるテスト、目標指向型テスト、パフォーマンスアセスメントなどの開発の必要を主張している。

また、この学会では日本版テストスタンダードの作成に取り組んでおり、テストの作成と実施、結果の利用と保管、テスト実施者の責任などについての基本的条項と、テスト実施の具体的な手順等を示したガイドラインが近々策定される予定である。

2. 教育評価をめぐる問題と指導と評価の一体化

鹿毛(2005)は、教育評価研究における問題として、測定と評価が混同されてきたこと、評価研究が評価方法の開発とその正しい実施の仕方ばかり注目してきたこと、教育実践的視座が欠如されてきたことを挙げながら、これまでの教育評価研究のレビューを行った。そして、評価を、問題解決過程を貫く思考のサイクルとしてモデル化することを提案した。さらに、このモデルによる評価活動を実施することが、学習と指導と評価の三位一体化の成立につながり、カリキュラムに埋め込まれたアセスメントの実現につながることを主張した。

ここで、「指導と評価の一体化」という用語について、筆者の見解を述べておきたい。

そもそも、「指導」と「評価」は意図的に一体化させなければならないのだろうか。筆者は、「指導と評価の一体化」なる用語がスローガンになること自体が、これまで指導と評価を別物と捉えてきたことの現れではないかと考えている。いいかえると、評価活動は期末考査や小テスト、また入学試験や資格試験といった、学習活動から切り離された時間においてのみ実施されるものといった考え方が根強いのではないだろうか。

Glaser(1976)に代表される教授学習過程のモデルでは、教授目標の設定を行うとともに、教授開始時の学習者の状態を把握した上で教授手続きを行い、次いで教授内容が学習者の身についたかどうかの評価を行い、その結果が次の教授目標の設定に反映されると考えられている。つまり、教授学習過程においては常に評価が実施されながら教授活動が続けられているのである。

たとえば、ある教室で1時間の授業を行う場合を考えてみよう。授業開始時に教師は子どもの様子を把握する。そして、実際に授業を行っている間も、常に子どもに発問し、また質問などを受け、机間指導を行いながら、子どもが指導内容をきちんと理解しているかどうかを適宜考えながら授業を進行させる。また、定期テストを行えば、指導の改善の方向を結果から探ろうとするだろう。このように、教師は常に評価を行いながら学習指導を行っているのである。

仮に、子どもの実態の把握もせずに一方通行の授業をしたら、その授業は効果的ではなくなることは想像に難くない。また、指導を適切に行うには、指導内容に基づいた適切な評価を行う必要が生じる。

このように考えると、指導と評価は一体化させるべきものではなく、常に一体であると捉えるべきであることが理解できよう。そして、指導と評価を一体化させようとするのではなく、すでに一体である指導と評価の両方が適切になされることによって、教育実践が首尾よく行われると考えるべきであろう。

ところで、三上(2004)は、教育評価と教育行政ないしは教育政策との関連について論じ、教育評価を子どもの学習状況の評価にとどめるのではなく、教育評価と社会評価とを総合的に捉える必要が生じていることを指摘している。教師評価、授業評価、学校評価などがクローズアップされている現状は、教育心理学における教育評価研究に対しても、従来の枠組みからの脱却を求めているのだろう。

2005年度の日本心理学会公開シンポジウム「教育のアカウンタビリティを問う」(鹿毛他, 2005)などが開催されたのも、教育心理学研究者がこのような問題意識をもつようになったことの現れであろう。なお、藤岡(2004)は、最近の教育評価の動向と課題を教育政策の展開を軸に論じている。

3. 適応型テストの研究

項目反応理論を用いた適応型テストの開発は従来から取り組まれているところではあるが、e-learningが実用化され、また情報技術の進歩により一層盛んになっている。

従来の項目反応理論を用いたテストにおいては、正解

または不正解といった2値によって能力推定値を求め、結果のレポートが行われていた。しかし、学校などで多用されるテストは、2値ではなく多値の反応（部分点なども与えるなど）を仮定したことが多い。

このような形式のテストに対して適用可能なモデルとして、段階反応モデルが挙げられる。菊地（2005）はこのモデルに基づく適応型テストのシステムを開発した。このシステムは、複雑なプログラミングの知識を必要とせず、一般に用いられている文書作成ソフトと表計算ソフトがあれば適応型テストが実施できるよう工夫されている。

また、学校などで多用されるテストにおいては、大問形式のものが多く用いられるが、項目反応理論は独立した項目からなるテストに対してのみ適用可能である。大森・繁樹（2005）は、大問形式のテスト構造をベイジアンネットワークモデルにより表現し、適応型テストへの適用を試みている。

従来の項目反応理論を用いたテストは、包括的に能力を捉えようとするテストとなるため、その結果を、ある特定の過程で学習者が学んだことと直接的に結びつけることはできないという問題があった。この研究で提案された方法は、そういった問題を解決する一方法である。

4. 項目反応理論による等化を用いた学力の経年比較

斉田（2003）による、全県規模の英語学力テストに対して項目反応理論を用いて共通受験者デザインによる等化を行い、数年間にわたる学力テストの結果について同一尺度上で検討を行った研究は、学力低下の実態を実証的に示すデータとしてセンセーショナルなものであった。

以後、同様のデザインを用いて、異なる項目からなる複数のテストを等化し、学力の経年変化を検討した研究が増えつつある。

斉田（2005）は、全県規模の英語学力テストの等化を行い、各年度の高校1年生の入学時の英語学力の経年変化を検討した。その結果、現行学習指導要領によって中学校教育を受けた年数の多い生徒ほど英語学力の落ち込みが大きいことを示した。

また、熊谷・山口・小林・脇田・野口（2005）は、1995年度から2004年度に実施された大学受験予備校における英語学力テストの等化を行った。その結果、この10年間に明らかな学力の推移が生じているという結論には至らなかった。

さらに、吉村・荘島・杉野・野澤・清水・齋藤・根岸・岡部・フレイザー（2005）は、1990年から2004年に出題された大学入試センター試験英語問題の第2問の問題項目から識別の高い100項目を抽出し、共通受験者デザインによる等化を行い、英語学力の経年変化を検討した。そ

の結果、1997年以降のセンター試験受験者の英語学力が、センター試験で把握可能な英語学力に限っていえば低下していることが明らかとなった。

このように、共通受験者デザインと呼ばれる方法による等化を行うことで、項目反応理論を用いた等化を行う前提のない試験（アンカー項目のない試験）を、同一尺度上に扱い、結果の変化を検討することが可能となる。しかし、等化に用いられた過去に実施されたテストにおける項目に代表される指導内容が、等化に用いられたデータの収集対象となった時期の集団にとっても同様の重みで指導されたかどうかで等化後の能力推定値が左右されかねない点は、結果の解釈を行う上で十分留意すべきである。

また、過去に行われたテストの結果を時系列に並べて「学力」の経年変化を検討するという場合には、比較対象となるテストの実施された期間においては学力観も変わらないという前提が必要となる。この前提が満たされない場合においては、以前の受験者にとってはできたことが、ある時期からの受験者にとってはできなくなっているとか、以前はできなかったことが、ある時期からできるようになっている、または昔も今もできることには変わりがないというような結論しか導けない。

学力観は時代によって変化するものであるため、学力の経年比較を行う場合には以上のような点に十分留意するとともに、1つのテストから学力を論じるのではなく、多様な資料を用いて総合的に判断することが必要である。

VII. 学力テストをめぐって

先に述べたように多くの地方自治体で学力テストが実施され、学力テストブームの様相を呈しているのが現状である。また、悉皆による全国学力テストに向けた準備も進められており、学力テストは多くの国民の関心事となっている。

小寺（2004）は、東京都が悉皆で行った中学校数学の学力調査について、結果が政治的に用いられている実態に触れ、また調査問題自体の不適切性、結果分析の一面性、学力調査結果に基づく「授業改善推進プラン」強要の危険性について検討を行っている。

特に、東京都教育委員会が学力調査の結果をふまえて提示した5教科共通の授業改善の視点が、楽しい授業の創造、体験的な学習や問題解決的な学習の充実、繰り返し学習の必要性、個に応じた指導の充実であったことを紹介している。さらに数学については、いろいろな解き方をする活動を取り入れるという視点が提示されたことも紹介し、これらの視点を数学の授業にどう具体化するかが問題であると指摘している。

同様の問題提起は他にもなされている。たとえば、高

浦(2005b)は、学力を育て、高めるための方法や内容を探るためには、単元を特定した事例研究的な手法による研究調査が不可欠であることを述べ、今後は学力の測定よりも、学力を育て、高める評価へとシフトした研究調査が重視されるべきであることを主張している。

また、山森(2005c)は、学力テストの結果が地域や学校の順位づけにつながっている実態や、特に順位の低かった学校に対しては学力テストの得点を高めるべく、休業期間中の補習授業の実施などが強いられている実態があることを指摘している。さらに、このような学力テストの結果からは、どのような子どもに、どのような指導を行えばよいのかといった、指導方法の具体的な改善案が提示されない問題があることが指摘されている。

なお、梶田(2005)は、学力調査や統一学力テストは実証的基盤に立って教育のあり方を改善・向上させていくために不可欠の役割を果たすと主張するとともに、それらの結果をそのまま学力とみなしてしまうことや、過度な競争意識を生む可能性があるといった危険性を十分認識する必要を指摘している。

2005年度の教育心理学会総会では、学力評価法(森・村山・瀬尾・松下・深谷・村井, 2005)や、学力をどのように捉えるべきか(荘島・山森・柳井・青木・盛永・伊藤・無藤, 2005)をテーマとしたシンポジウムが企画されるなど、学力問題を視野に入れたシンポジウムがいくつか企画された。このように、学力問題や学力テストについての問題に対する関心が教育心理学においても高まりつつあるといえる。これらの関心が、学力テストに関わる問題の解決を意図した研究が増えることにつながることを期待したい。

Ⅷ. 教育心理学における教育評価研究の方向

1. 意志決定志向的研究の充実に向けて

Cronbach and Suppes (1969) は、教育研究を意志決定志向的研究と、結論志向的研究の2つのスタイルに分類した。前者は、さしせまった教育上の問題の解決という社会的要請に応えるための研究である。一方、後者は、研究者の学問的興味などに支えられながら、正しい結論に向かって掘り下げる研究である。

先に筆者は、学力向上を目標とした日々の指導における工夫と改善を行いながら実践を行うとともに、子どもの学習状況の評価を適切に行い、説明責任を果たすことが学校に求められていることを指摘した。また、授業時数の捻出が困難な中、国や地方による学力テストを実施することとともに、学校運営に対する自己評価と他者評価にも取り組むことが、学校に対して求められていることも指摘した。

このレビューを行ってわかったことは、学校が抱える

問題に対して、直接的、間接的に寄与できるような研究成果が多いことである。しかし、現在学校で行われている目標準拠評価についての研究が手薄であり、また、学力テストについての研究に対して、ようやく関心が寄せられるようになっている状況であることも、また事実である。いいかえると、結論志向的研究の充実が図られる一方で、意志決定志向的研究を行おうとする意識が、教育心理学研究者の中に芽生え始めているといえよう。

現在、国や各地方で教育改革の動きは急である。また、学校は目の前の子どもに対して働きかけなければならない。そのため、これらの問題に対応した意志決定志向的研究に取り組む際には迅速性が求められよう。また同時に、小手先の対応に終始しない研究を行うためには、結論志向的研究の充実も図られなければならないだろう。そのため、教育心理学における教育評価研究においては、意志決定志向的研究と結論志向的研究の両者が、バランスよく進められることを期待したい。

2. 教育心理学における教育評価研究

筆者は、このレビューを執筆しながら、ある違和感を始終もち続けていた。

その違和感の一つが、教授・学習過程における評価の研究が、なぜ「評価」の研究という狭いカテゴリーに分類されてしまうのだろうかという思いである。特に、「子どもの学習とその評価」という節において取り上げた研究の多くは、教授・学習過程の研究としてもカテゴライズできるのではないかという思いをもっていた。さらに、なぜ指導と評価がそれぞれ別物として扱われているのだろうかという疑問を抱き続けるとともに、従来の分類による教授・学習過程研究と評価研究を一緒くたにするのも、また問題であろうと、おぼろげに感じていた。

もう一つの違和感は、教育評価研究における教育心理学の立つ位置についてである。教育評価研究のレビューを行うには、教育心理学でなされた研究だけを取り上げるのでは不十分なものにならざるを得ないと判断し、このようなレビューを行うに至った。その中で、教育心理学と隣接諸領域との境界が曖昧となっていることに気づくに至った。

これらの点に関して、鹿毛(2005)は、評価実践が教育実践に埋め込まれている発想に立つことの必要を主張し、教育評価研究においては「教育実践としての評価」という営為を明らかにすることが課題であると述べている。また、心理学における教育評価研究は、テスト開発、データ解析などに特化した閉じられた研究領域から、学習、思考、動機づけ、教授・学習過程といった他の教育心理学分野、さらには教育方法学、教育工学などの隣接分野と融合した開かれた研究領域へと発展していく可能性を

指摘している。

このような論考に触れつつ筆者が感じたのは、教育評価研究のうち、教授・学習過程における評価の研究にあっては、評価だけを単独に扱うのでは不十分だということである。そして、評価を切り口として教授・学習過程を論じるという態度をもつことが、今後必要となるだろう。

さらに、隣接諸領域と融合した開かれた研究領域へと発展する可能性があるからこそ、教育評価研究における教育心理学の役割と、教育心理学における評価研究の位置づけを問い直すことが求められるだろう。

引用文献

- Cronbach, L. J., & Suppes, P. 1969 *Research for tomorrow's schools: Disciplined inquiry for education*. Toronto: Macmillan.
- 遠藤健治 2005 授業評価項目による学生の受講態度形成に関する検討 日本心理学会第69回大会発表論文集, 1280.
- 藤岡秀樹 2004 教育評価の動向と課題 教育目標・評価学会紀要, **14**, 3-10.
- Glaser, R. 1976 Components of a psychology of instruction: Toward a science of design. *Review of Educational Research*, **46**, 1-24.
- 服部 環・上殿明良・宇都宮公訓・清水一彦・徳田克己・本橋信義 2005 学生による授業評価における評定の個人差と評定平均値 日本教育心理学会第47回総会発表論文集, 579.
- 廣瀬英子 2005 他者評価を取り入れたときの大学生の発表スライド作成力の変化 日本教育心理学会第47回総会発表論文集, 379.
- 池田 央 2005 「日本テスト学会」に期待される課題と役割—「日本テスト学会誌」の創刊に寄せて— 日本テスト学会誌, **1**, 3-14.
- Inuzuka, M. 2005 Learning how to write through encouraging metacognitive monitoring: The effect of evaluating essays written by others. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Pp. 1018-1023.
- 鹿毛雅治 2005 教育評価再考—実践的視座からの展望— 心理学評論, **47**, 300-317.
- 鹿毛雅治・大村彰道・志水宏吉・市川伸一・松下佳代・無藤 隆・北尾倫彦 2005 教育のアカウンタビリティを問う 日本心理学会第69回大会発表論文集, S2.
- 梶田叡一 2005 学力調査・学力テストの論理とその活用 人間教育研究協議会(編), 教育評価の課題を問い直す(教育フォーラム35) 金子書房. Pp. 121-128.
- 荻谷剛彦・志水宏吉 2005 学力の社会学 岩波書店.
- 菊地賢一 2005 段階反応モデルにもとづく汎用的適応型テストシステムの開発 日本行動計量学会第33回大会発表論文抄録集, 362-363.
- 北尾倫彦 2004 目標準拠評価三年目の現状 指導と評価, **50**(8), 4-8.
- 小寺隆幸 2004 中学校数学の「学力調査」を分析する視点—東京都「児童・生徒の学力向上を図るための調査」の検討を通して— 教育目標・評価学会紀要, **14**, 11-21.
- 国立教育政策研究所 2004 生きるための知識と技能2—OECD 生徒の学習到達度調査(PISA) 2003年調査国際結果報告書— ぎょうせい.
- 国立教育政策研究所 2005a 算数・数学教育の国際比較—国際数学・理科教育動向調査の2003年調査報告書— ぎょうせい.
- 国立教育政策研究所 2005b 理科教育の国際比較—国際数学・理科教育動向調査の2003年調査報告書— ぎょうせい.
- 国立教育政策研究所 2005c 総合的な学習の時間の授業と評価の工夫—評価規準及び評価基準を介した指導の改善, 自己学習力の向上及び外部への説明責任に向けて(第三次・最終報告書)—.
- 熊谷龍一・山口大輔・小林万里子・脇田貴文・野口裕之 2005 大規模英語学力テストにおける年度間比較—大学受験生の英語学力の推移— 日本教育心理学会第47回総会発表論文集, 572.
- 桑田隆男 2004 子どもの作品をどう評価するか—小学校社会の「観察・資料活用の技能・表現」でルーブリック・スタンダード準拠評価— 指導と評価, **50**(8), 26-29.
- 牧野幸志・西浦和樹 2005 学生による授業評価と出席率との関係(1)—授業に出ていない学生は授業を悪く評価するのか?— 日本心理学会第69回大会発表論文集, 1240.
- 増田公男 2005 大学生における「授業評価の評価」の変化 日本教育心理学会第47回総会発表論文集, 397.
- 三上和夫 2004 評価と国家—「評価と社会」の視点から— 教育目標・評価学会紀要, **14**, 29-37.
- 森 敏昭・村山 航・瀬尾美紀子・松下佳代・深谷優子・村井潤一郎 2005 学力評価法の新たな可能性 日本教育心理学会第47回総会発表論文集, S18-S19.
- 村山 航 2004 テスト形式の違いによる学習方略と有効性の認知の変容 心理学研究, **75**, 262-268.
- 村山 航 2005a PISAの“基本”を確認する—調査の

- 枠組みや得点解釈の方法に関する素朴な疑問について— 指導と評価, **51**(6), 13-17.
- 村山 航 2005b テスト形式の予期による方略変容メカニズムの検討 教育心理学研究, **53**, 172-184.
- 村山 航 2005c 空所補充型テストと学習方略—テスト形式スキーマの介入が及ぼす影響— 日本教育心理学会第47回総会発表論文集, 371.
- 中嶋順子 2005 授業態度および授業評価と授業の満足度との関係 日本心理学会第69回大会発表論文集, 1279.
- 名取洋典 2005 指導者のことばがけが競技者のやる気におよぼす影響(3)—ことばがけの捉え方とやる気— 日本教育心理学会第47回総会発表論文集, 470.
- 西浦和樹 2005 学生による授業評価と自己効力感・テスト不安との関係 日本教育心理学会第47回総会発表論文集, 384.
- 野村幸司 2004 高知における「授業評価」—高等学校での取り組みを中心に— 教育目標・評価学会紀要, **14**, 22-28.
- OECD 2004 *Learning for tomorrow's world: First results from PISA 2003*. Paris: Author.
- 岡 直樹・松山美香 2005 ドリル学習における評価が計算力や自己効力感に及ぼす影響 日本教育心理学会第47回総会発表論文集, 450.
- 大道一弘・野嶋栄一郎 2005 異なるクラス間における目標構造の差の検討 日本教育心理学会第47回総会発表論文集, 494.
- 大森拓哉・繁桝算男 2005 テストレット形式問題のネットワーク型アダプティブテスト 日本行動計量学会第33回大会発表論文抄録集, 264-265.
- 斉田智里 2003 高校入学時の英語能力値の年次推移—項目応答理論を用いた県規模英語学力テストの共通尺度化— *Step Bulletin*, **15**, 12-24.
- 斉田智里 2005 ゆとり教育と英語学力低下(2002-2005) 日本テスト学会第3回大会発表論文抄録集, 74-75.
- 笹山萩子・岸 学 2005 説明文読解テストの出題形式が問題の難易度に与える影響—項目応答理論における項目困難度による検討— 日本教育心理学会第47回総会発表論文集, 565.
- 荘島宏二郎・山森光陽・柳井晴夫・青木栄一・盛永俊弘・伊藤 圭・無藤 隆 2005 学力のとらえ方 日本教育心理学会第47回総会発表論文集, S20-S21.
- 角谷詩織・無藤 隆 2005 科学的興味追求意欲を高める理科の授業の意義 日本教育心理学会第47回総会発表論文集, 121.
- 高浦勝義 2004 絶対評価とルーブリックの理論と実際 黎明書房.
- 高浦勝義 2005a ポートフォリオ評価を活用した指導の改善, 自己学習力の向上及び外部への説明責任に向けた評価の工夫—生活, 国語, 社会, 理科, 音楽, 図画工作・美術, 家庭, 保健体育, 英語, 道徳, 特別活動を事例にして(第三次・最終報告書)— 科学研究費補助金報告書.
- 高浦勝義 2005b 学力の測定から学力を育てる評価へ 論文集編集委員会(編), 学力の総合的研究 黎明書房. Pp. 8-20.
- 田中耕治 2005 入学試験の過去・現在・未来 指導と評価, **51**(1), 24-27.
- 棚上奈緒・淵上克義・田中宏二 2005 教師集団の効力感尺度作成に関する研究 日本教育心理学会第47回総会発表論文集, 328.
- 東京大学大学院教育学研究科基礎学力研究開発センター 2005 市区町村の学力向上施策に関する調査—集計結果報告書—.
- 内田照久・中畝菜穂子・石塚智一 2004 大学入試センターの英語問題で測られる学力とリスニング・テストで測定される言語運用能力の能力推定値を介した関係性の検討—試験問題の事後標準化とリスニング・テストでの問題提示順序の検討と共に— 大学入試センター研究紀要, **33**, 29-63.
- 内田照久・中畝菜穂子・荘島宏二郎 2005 英語リスニングテスト実施時に各種騒音が与える影響 日本テスト学会誌, **1**, 117-127.
- Wiggins, G. 1989 A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, **70**, 703-713.
- 山森光陽 2005a 目標標準評価を支える評価理論と測定理論 指導と評価, **51**(11), 22-25.
- 山森光陽 2005b 内容系列表, 単元指導計画, 評価計画の作成の意義をめぐって—「授業及び評価に関する開発的研究」研究協力教員を対象としたアンケート調査から— 日本生活科・総合的学習教育学会第14回全国大会大会要項, 29.
- 山森光陽 2005c 学力テストの教育心理学的考察 論文集編集委員会(編), 学力の総合的研究 黎明書房. Pp. 205-215.
- 八尾坂修 2005 人事考課—基本は人材育成— 指導と評価, **51**(4), 33-37.
- 四日市市立教育センター 2003 中学校での社会科学習における絶対評価の在り方に関する研究—評価規準の作成と評価基準の設定による授業実践を通して—.

吉川愛弓・岸 学 2005 作文の評価項目に関する検討—意見文の評価は何に影響を受けるのか— 日本教育心理学会第47回総会発表論文集, 337.
吉村 幸・荘島宏二郎・杉野直樹・野澤 健・清水裕子・

齋藤栄二・根岸雅史・岡部純子・サイモンフレイザー 2005 大学入試センター試験既出問題を利用した共通受験者計画による英語学力の経年変化の調査 日本テスト学会誌, 1, 51-58.