

総括的評価による形成的評価への阻害要因とその解消事例の検討

—Harlenの評価機能の緊張関係と相乗効果議論を手掛かりに—

教職開発コース 岩 堀 翔 太

Investigating the Barriers to Formative Assessment by Summative Assessment and the Case for its Resolution: Using Harlen's Discussion of Tensions and Synergies of Assessment Functions

Shota IWAHORI

This study examined teachers' narratives to identify the barriers to formative assessment by summative assessment and explore the case for its resolution. Data were collected through semi-structured interviews with 11 public middle school teachers and analyzed using a modified grounded theory approach (M-GTA). The results of the analysis were summarized into 20 concepts and 6 categories, and the relationship between the 20 concepts and 6 categories was shown in the resultant diagram. In addition, the ways of resolving the obstacles to formative assessment through summative assessment extracted from the teachers' narratives were organized in the framework of 3 approaches categorized from previous studies, and how each case corresponds to the 20 concepts generated by M-GTA was interpretatively examined to explore the possibility of overcoming the obstacles.

目 次

1. 問題と目的
 - A. 問題の所在
 - B. 先行研究
 - C. 本研究の目的
2. 方法
 - A. 研究協力者と手続き
 - B. データの分析
3. 結果
 - A. 総括的評価による形成的評価への阻害要因
 - 1 授業スタイルの硬直化
 - 2 評価方法の限定化
 - 3 フィードバック効果の弱体化
 - 4 評価意識による制約
 - 5 評価リテラシーの停滞
 - 6 学習者の総括的評価志向
 - 7 カテゴリー・概念間の関係
 - B. 総括的評価による形成的評価への阻害要因の解消事例
 - 1 形成的評価の総括的活用（総括的活用型）
 - 2 総括的評価の形成的活用（形成的活用型）
 - 3 評価機能の分離（分離型）
4. 考察

1. 問題と目的

A. 問題の所在

学習評価は機能の違いにより大きく2つに分類できる。1つ目は、「総括的評価」と呼ばれ学習成果の成績や証明のためになされ、説明責任や順位決め、クラス分け、成績付けと密接に関係している評価である（Black, Harrison, Lee, Marshall, & Wiliam, 2004）¹⁾。2つ目は、「形成的評価」と呼ばれ学習の促進や次の手立ての判断を示すためになされ、学習ニーズを把握し指導を調整するための頻繁かつ対話的な学習評価である（OECD, 2005）²⁾。授業の質を改善したり学習成果を高めたりする上での、形成的評価の効果的な活用の重要性が言われているが（Black & Wiliam, 1998）³⁾、形成的評価を効果的に活用するための課題も指摘されている。ベネット（2011）⁴⁾が整理した形成的評価の課題の中では形成的評価と総括的評価が相乗的に働いていないという「評価の内的一貫性」の問題が指摘されている。同様に、佐貫（2012）⁵⁾は「評価の二重性議論」として形成的評価に総括的評価としての機能が不可分に存在することで純粋な形成的評価過程が歪められることの問題に言及している。

このことから、形成的評価の効果的な活用のためには、総括的評価による障壁を克服する方略を考えるこ

とが必要になる。そこで、本研究では、総括的評価による形成的評価への阻害要因とその解消方法について教師の語りを基に明らかにする。以下、本研究に関連する先行研究をレビューし研究の目的を記述する。

B. 先行研究

1 総括的評価が教師の形成的評価に与える影響

1990年代以降欧米で始まった「スタンダード」に基づく教育改革は総括的評価と形成的評価の関係についての問い直しを迫っているという(西岡ら, 2015)⁶⁾。「スタンダード」に基づく教育改革の下では、教師や学校はアカウンタビリティを果たすために統一テストの結果をスタンダードの到達人数によって示すなど、数値目標の達成という自己目的的な活動へと学校教育を向かわせている。こうしたマクロな教育システムが働いている状況下では形成的評価がテストのための学習点検へと矮小化され、質の高い学習を保障する教師の草の根的な評価という機能が断たれることとなる。

ハーレン(2004; 2005)⁷⁾⁸⁾のシステマティックレビューによると、評価が証明や選抜など総括的に機能するとき、評価はハイスティクス化¹⁾し教授や学習の在り方にも影響する。例えば、評価が総括的に機能するとき、教師はテスト演習の繰り返しや頻出問題への解法指導に注力し一方通行型の授業に陥る(Stiggins, 1999; Linn, 2000)^{9) 10)}。また、総括的評価が支配的な環境では、客観性や公平性、信頼性のある評価方法が重要であるため、ペーパーテスト型の評価が重用される。このことは、教師が授業において、子どもの学習状況を見取りフィードバックを与えるといった質的な評価の価値を相対的に低下させ得る。また、形成的評価がハイスティクスなテストに向けた頻繁な小テストの実施などに矮小化されて理解されてきたことや(石井, 2020)¹¹⁾、形成的評価に総括的評価としての機能が不十分に存在してしまっていること(佐貫, 2012)¹²⁾などの課題も多い。

2 総括的評価による形成的評価への阻害要因の解消方法

これまでの研究を踏まえると、総括的評価による形成的評価への阻害要因の解消方法は次の3つのアプローチに分類できる。1つ目は、「形成的評価の総括的活用」である(以下、総括的活用型)。これは、形成的評価を第1目的として収集した評価情報を総括的評価としても利用する方略である。ハーレン(2005)¹³⁾は、形成的目的の教師の質的な評価を総括的

に活用することの利点として、教師が日常の対話の中から妥当性の高い見取りができることや、学習者と評価を共有することで学習者が自身の学習プロセスに積極的に関与できることなどを挙げている。

2つ目は、「総括的評価の形成的活用」である(以下、形成的活用型)。これは、総括的活用型とは反対に、総括的評価を第1目的として収集した評価情報を形成的評価としても利用する方略である。この方略は、ブラックら(2004)¹⁴⁾が形成的評価の4つの活用方略を示した中の1つに位置づけられており、テスト場面を想定した具体的な実践例を示している。

3つ目は、「評価機能の分離」である(以下、分離型)。ハーレン(2005)¹⁵⁾は、クイーンズランドでの取り組み事例を基に授業における形成的目的の教師の評価を説明責任の手続きから完全に切り離すことで、総括的評価による形成的評価への負の影響を緩和する方略を示している。評価の目的を明確に分類することで形成的評価過程に総括的評価が無制限に侵食することを防ぎ、学習改善・指導改善のための形成的評価を実現することを構造的に支えることが期待される。

C. 本研究の目的

本研究は、半構造化インタビューによる教師の語りの検討から、総括的評価による形成的評価への阻害要因とその解消方法を明らかにすることを目的とする。やまだ(2007)¹⁶⁾によると、語りの研究は語る人の意味付けを明らかにしようとする点に意義があり、経験をどのように解釈するかは、個々の教師の行動や考え方を規定する深層レベルの価値観に影響される。これまで、総括的評価と形成的評価の関係について教育システムなどマクロな視点から論じられてきた一方で、教師の経験や認識から両者の関係を問い直す研究は十分に行われてこなかった。そのため、本研究において教師の具体的な語りを検討することで、教師の信念や教師が位置付けられている文脈に則して、ボトムアップに総括的評価による形成的評価への阻害要因に迫ることができる。また、先行研究で示した3つのアプローチ(総括的活用型、形成的活用型、分離型)の有用性を、教師の語りから明らかにする阻害要因と対応させて検討することで、実際の教授・学習プロセスに即した効果的な方略を検討でき実践的な知見としての意義もある。以上の問題意識をもとに本研究を構想した。

2. 方法

A. 研究協力者と手続き

本研究は、機縁法により集めた公立中学校の11名の教師に協力を依頼した。研究協力者の11名の教師は、研究会で実践発表をしたり、校内研修での研究授業を担当したりした経験を持ち形成的評価という言葉がある程度知っているとして自己認識している教師である。11名の教師の概要を表1に記す。

半構造化インタビューは、Skypeによるオンライン形式で1人につき約75分間1対1で実施した。D, I, J教諭には、1回目のインタビューから約1か月後に2回目のインタビューをそれぞれ約30分実施した。インタビューでは、「評価をする際に評定はどの程度意識しますか。」や「フィードバックする際に困ることはありますか。」「テストは授業にどのように影響していますか。」等について質問を行った。インタビューの様子は、各教師の了承を得たうえでICレコーダーに記録し全てトランスクリプトを作成した。

B. データの分析

分析方法は、分析対象者の語りを帰納的に分析する質的研究法である木下(2003, 2007)^{17) 18)}の修正版グラウンデッドセオリーアプローチ(以下、M-GTA)を用いた。総括的評価が教師の形成的評価の活用の障壁となる背景には、教師が日常の教授場面で受けている総括的評価による影響の経験プロセスがあるため、プロセス的な特性を背景にもつ現象に対し独自の分析テーマをたて検討する研究に適したM-GTAが本研究に適していると考えられる。

また、木下はM-GTAの応用可能性として、分析結果が実践の改善のために現場で活用されることの重要性を指摘している。この点について、総括的評価による形成的評価への阻害要因を明らかにすることは、どのように教師がその障壁を克服し形成的評価の効果的

な活用を実現できるのかを検討する上での基礎的な知見となるため、教師への実践的な支援を期待する本研究の意義とも一致するものである。

本研究では、まず1人分のトランスクリプトを検討し、設定した分析テーマ(総括的評価が教師の形成的評価の活用に与える影響)と関連性のある概念の生成を行った。1人分の分析が終了したら2人目のトランスクリプトの検討を行い、新しい具体例が出現する度に既存の概念との相違を検討した上で、新たな概念を生成した。また、既存の概念に新たな具体例を追加するには、定義が具体例を包括できているかどうか検討し、必要に応じて定義を修正した。5人の分析が終了したところで統合または分化可能な概念がないか検討した。木下(2003)¹⁹⁾は、個々の概念の有望さはヴァリエーションによると述べており、具体例が少なすぎる場合は他の近接する概念との統合を検討し、反対に、具体例が多すぎる場合には概念の分化を検討した。そこで新たに概念が生成された場合には、コーディングを終えた分のデータも再検討した。研究協力者全員分のコーディングを行い、これ以上新たな概念が生成されず、新しいデータの収集の必要がなくなった時点で概念生成に関する理論的飽和化に達したと判断した。

生成された各概念は、概念間の関係性を検討した上で上位のカテゴリーにまとめられた。生成した概念とカテゴリー間の関係性については、総括的評価による形成的評価への阻害要因の結果図としてまとめた。このカテゴリー間の影響関係は、教師の語りにより明示的または暗示的に示されたものを基に研究者が解釈し整理した。なお、結果図は調査協力者の教師に確認してもらい、インタビュー内容を反映した妥当なものであると判断された。次に、総括的評価による形成的評価への阻害要因の解消方法について、教師の語りから抽出した具体的事例を、3つのアプローチ(総括的活用型、形成的活用型、評価の分離型)に分類し紹介す

表1 研究協力者の概要

名前	教科	教職歴	名前	教科	教職歴
A	数学	34年	G	社会	12年
B	社会	29年	H	英語	8年
C	英語	24年	I	社会	8年
D	数学	23年	J	国語	8年
E	国語	21年	K	理科	6年
F	理科	18年			

る。そして、これらの解消方法が、M-GTAにより生成した総括的評価による形成的評価への阻害要因を示す概念とどのように対応しているのかを解釈的な事例分析により検討し、3つのアプローチそれぞれの有用性を明示する。

3. 結果

A. 総括的評価による形成的評価への阻害要因

分析により生成された概念は20、カテゴリーは6である。生成された、カテゴリー、概念、定義、具体例を表2に、カテゴリー間の関係を示した結果図を図1に示す。以下、カテゴリーと概念の意味及びカテゴリー・概念間の関係性を説明する。なお、文中ではカテゴリーは〈 〉、概念は【 】で示す。

1 授業スタイルの硬直化

〈授業スタイルの硬直化〉は4つの概念から成る。【テスト演習型の授業】は、テスト対策のための過去問やその解説を中心とした授業実践を示す。評定や学力証明、選抜のためのツールとしてテストが重要なポジションに位置付けられることで、テスト演習の繰り返しの授業がなされ多様な評価課題を用いることや教師と生徒の対話が積極的に生まれるような形成的評価を活用する環境になりにくい。

このようなテスト対策に偏重した授業では、より多くの知識を効率的に伝えるための【知識伝達型の授業】が展開される。知識伝達を中心とした授業では、生徒の理解を確認したり思考を表出させたりするような活動がしにくい。

テストが大きな価値をもつ中では、教師は【テスト範囲を網羅するための授業】を行う。これは網羅主義的傾向（石井, 2020）²⁰として指摘されている課題と一致し、生徒の学びの定着や深まりを確認したりフィードバックをしたりする形成的評価の機会が生まれにくい。

【形骸化した授業内の活動】は、深い学びをもたらすための教師の見取りやフィードバックといった形成的評価を伴わない、活動自体が目的化した授業実態が語られている。

2 評価方法の限定化

〈評価方法の限定化〉は3つの概念から成る。【入学試験に合わせたテスト評価依存】は、高校入学試験の合否において大きな比重を占めるペーパーテストの存

在に影響され、テスト形式の評価方法に依存している語りである。また、【入学試験で活用するための評定を算出しやすいテスト評価依存】もペーパーテストへの依存を表しているが、こちらは評定が入学試験において重要な資料となるために、評定換算に適していると教師が認識しているペーパーテストが評価方法として偏重的に選択されていることを示す語りである。入学試験という選抜のための評価の影響により、それとの親和性の高い評価方法としてペーパーテストが最重要視された結果、授業における多様な評価方法を選択する可能性を低下させる。

実際、【評定に反映しにくい評価方法の排除】では、数値化が難しい文脈依存的で質的な評価や、高次の学力を対象とした評価、教師の主観性が強く反映される評価については、評定換算することへの抵抗感が強く教師が敬遠する傾向があることが示されている。

3 フィードバック効果の弱体化

〈フィードバック効果の弱体化〉は3つの概念から成る。【成果物の数値化】は、レポートやワークシートといった成果物を数値化してフィードバックすることを表す語りである。教師は、コメントフィードバックも行うが、そこに数値化した評価が伴うことが特徴的である。この背景には、明確なフィードバックを与えたいという教師の意図とともに、これらの成果物を評定算出のための資料として活用することを想定していることがある。

【やり直しの機会の喪失】は、総括的評価が優位な状況下では、フィードバックを活用したやり直しの機会を十分にもたないことを表す概念である。評定の良し悪しのみを意識が向けられていると、やり直しの結果が評定に反映されない限りは、学び直しを積極的にやる動機が生まれず、やり直しをすることの価値が見出されにくくなる。

【観点別評価の評定への従属化】は、観点別評価は、本来は生徒の学習状況を分析的に捉え、指導・学習改善のために機能することが期待されているが（文部科学省, 2018）²¹、実際は、総括的な評定を出すための一過程に成り下がり、その形成的な機能を失っていることを示す語りである。具体的には、評定「1」を回避させるために、関心・意欲・態度の観点を「C」から「B」操作するなどの語りが聞かれた。このような評定に従属した観点別評価は、観点ごと分析的に生徒の学習状況を捉え、指導や学習改善のためのフィードバックとして用いられることはなくなる。

表2 総括的評価による形成的評価への阻害要因を示す概念リスト

カテゴリー	概念名	事例数	定義（上段）／具体例（下段）
授業スタイルの硬直化	テスト演習型の授業	4	テスト対策のための演習の繰り返しを中心とした学習指導 特に入試前の3年生の授業や、1,2年生でもテスト前の授業だったりすると、過去問やったり、単元テストで演習したりっていう授業になってます。[D教諭]
	知識伝達型の授業	6	テストに出題されやすい内容を効率的に教えるための、講義型の授業 3年生の入試が近づくと、考える価値のある問いよりも、出題頻度が高そうな問題を中心に、一方的に説明するような、あのそういうところだけを扱ってしまうことはよくあります。[I教諭]
	テスト範囲を網羅するための授業	7	指定されたテスト範囲を網羅することに迫られた授業 今度、県の学力テストがあるんですけど、その範囲が決まっているので、そこまで終えてないと大変なことになるので、今急いで教科書を進めているんですよ。[G教諭]
	形骸化した授業内の活動	4	授業における質的な見取りが、評定算出に馴染まないことによって、評価機会として生かされない学習活動 公開授業とかで、アクティブラーニングを実施しても、テスト点のような、評定に反映するような評価は難しいので、あまり評価を意識することなく、活動してるだけという感じのときもあります。[K教諭]
評価方法の限定化	入学試験に合わせたテスト評価依存	6	高校入学試験形式に対応した、ペーパーテストに偏重した評価方法の採用 入試では、生徒たちはテストを受けて、一発勝負でいい点をとらないといけないので、日ごろの評価がそれと違ってたりするとよくないかなとも思うので、どうしてもテストを重視してしまいます。[I教諭]
	入学試験で活用するための評定を算出しやすいテスト評価依存	9	高校入学選抜資料となる評定を算出するための、ペーパーテストに偏重した評価方法の採用 評定は、可否を左右するので、当然公平に、なるべく主観なくつけたいといけなくて思ったときに、テスト点のように明確にできたくないが、はっきりわかる評価っていうのがやっぱりテストだから、うん、テストでの評価が多いな。[A教諭]
	評定に反映しにくい評価方法の排除	3	評定換算が難しい、またはそぐわない評価方法は用いない 調べ学習とか問題解決学習とかして、そこから思考力とか意欲とか見ることができればいいんですけど、自分の見れている部分だけで評価したり、生徒によってわかることもばらつきがあるので、それを一律に成績に使わずらいので、あまり評価としては使わないです。[G教諭]
フィードバック効果の弱体化	成果物の数値化	10	学習の成果物を数値に変換して示すこと 実験レポートを書かせたりして、後日チェックして返すんですけど、やはりコメントというよりは、よければA、あまりよくない場合はCって言った感じで返してます。(中略) そのレポート評価は成績の資料にするので、やっぱりABC付けといたほうがいいのか、便利かなと思っています。[F教諭]
	やり直しの機会の喪失	3	フィードバックを使って、学習をやり直したり、振り返ったりする機会をもたないこと テストを返したら、簡単に説明とか全体の傾向とかいいますね。まあ、ただその後にはやり直したり、またそれを提出させたりとかはないですね。(中略) まあ、とりあえず最初に取れた点で成績つけないです。[J教諭]
評価意識による評価の信頼性への制約	観点別評価の評定への従属化	4	評定の数値に合わせて、逆算的に観点別の評価を決定することで、観点別評価が学習改善のための機能を失うこと 本来は、観点別の評価をして、それで評定を決めるんですけど、たまに評定をイメージして、それに合うように観点別の評価を操作するようなこともあって、特に関心意欲とかは。そうすると、本当は形成的評価のために使える観点別の評価はあまり意味がなくなってしまっているかもですよ。[D教諭]
	評価の信頼性への意識	3	評価の信頼性を担保するために、教師の見取りによる質的な評価を活用できないという制約意識 やっぱり自分が評価したものが、また他の人が見たら違うかもしれないし。テストとかならそういうことなくちゃんとした評価ができるけど、そうですね、うん、その都度結果がちよっとずつ変わるような主観的な評価はまずいいのかなと思います。[E教諭]

評価意識による制約	エビデンスのある評価への意識	3	<p>アカウントビリティを果たすために、明確なエビデンスを示せる評価しかできないという制約意識</p> <p>成績とかは、保護者もみるし、まあ管理職もチェックするので、「なぜ」って聞かれたときに当然答えられないといけません。だから、知識理解は、テストで何点なので、AですとかBですとか、そういう数字で証拠を出せる評価が大事になるかなと思います。[F教諭]</p>
	評価は負担のある行為という意識	5	<p>評価は、説明責任を考慮して実践しなければならないという負担感</p> <p>評定つけるのが結構、大変というか負担の多い事務作業ですね。もちろん何となくで付けられないし、事前の情報も集めておかないとだし。ちゃんと評価してますよって生徒に保護者にも見せないとなので。[I教諭]</p>
	教師が唯一の評価者という意識	3	<p>評価主体を、教師に限定して捉える意識</p> <p>生徒に自分で採点させたり評価させたりすることもありますけど、成績レベルになったらそれは使えないですよ。不公平になってもいけないし、教師が責任もってつけないとですよ。[B教諭]</p>
評価リテラシーの停滞	指導と分離した評価理解	4	<p>学習の成果を示すために、指導を完全に終了後に、評価を実施する</p> <p>ちゃんと全部教えたあとに、それがわかっているかを見てあげないと、生徒にとっても不公平になるので、一通り授業やって、単元テストとかしてって感じですね。[G教諭]</p>
	成績としての評価理解	8	<p>評価を、成績評価として捉える限定的な理解</p> <p>確かに評価っていうと、話していて、成績だけでなく、日ごろから見取ったり、アドバイスしたり、なんかそういうの全て評価なんだなと思ったんですけど、割と今まで、どう成績をつけるかみたいなのが評価の中心になってたなって思いました。[K教諭]</p>
	形成的評価の技能不足	7	<p>総括的評価の影響を強く受けた実践に偏る結果、形成的評価技能が高まらない</p> <p>どうやって授業しようかなは考えるんですけど、評価は小テストとかが何か当たり前になっていて、授業内での生徒の学習を見取るような、何か方法はあまり考えたことなくて、うん、なのでそうですね、やり方みたいなものがあるのかなって思いました。[H教諭]</p>
学習者の総括的評価志向	学習者の数値志向性	6	<p>学習者が、成果物を数値化したフィードバックを求める</p> <p>多くの生徒は、成績がどうかとか、テスト順位がどうかとかそういうことを意識してますよね。だから、ただ「よかったよ」とか言ったりコメントつけたりするよりも、具体的に何点とか言った方が励みになるかなってのはあります。[F教諭]</p>
	失敗のリスク認識	3	<p>学習者が、学習の失敗をリスクと捉え、教師の学習介入が消極的になる</p> <p>生徒もできないとか、間違っているとかが知られたくないので、教師にも友達にも。成績も下がっちゃうかもとか考えると特に。だから、あまり積極的に発言してくれないので、正直授業での理解はよくわからないところもあるんだよね。[G教諭]</p>
	学習者の成果志向性	4	<p>学習プロセスよりも学習結果が重視され、フィードバックによる学習改善サイクルがおこりにくい</p> <p>とりあえず、テスト返して、点数みてそれで終わりみたいのはどうにかしたいですけどね。まあ、その結果が変わるわけではないので生徒としては、やり直しのモチベーションが上がらないのかもしれないですけど。[E教諭]</p>

以上のように、評価を形成的に機能させるために不可欠なフィードバック過程への総括的評価の影響は、総括的評価による形成的評価への阻害要因における中核となる。

4 評価意識による制約

【評価意識による制約】は4つの概念から成る。【評価の信頼性への意識】は、評価したデータが安定的なものかどうかという点における教師の過度な意識を示している。当然、評価において信頼性は妥当性同様に

重要なものであるが(田中, 2008b)²²⁾、その過度な意識は指導や学びと統合された形成的評価を活用することへの制約ともなりうる。

【エビデンスのある評価への意識】は、評価には明確なエビデンスが伴わなければならない、それが欠如した評価はフィードバックに値しないという教師の意識を示している。教師は子どもや保護者へのアカウントビリティを果たすために、数値化されたエビデンスを必要とする。そして、アカウントビリティを果たすためのエビデンスを得ることへの過度な意識は、流動的

でプロセス性のある形成的評価の実践を消極的にする。

上述したように、説明責任や証明、選抜のために使用される総括的評価は、学習者やその保護者の選択に与える影響力が大きいが故に、【評価は負担のある行為という意識】を教師が抱き評価への積極的関与を後退させる。形成的評価は計画的にフォーマルに実施することもできるが、日常的にインフォーマルな形式でも使用できる強みがある (William, 2018)²³⁾。教師の評価への負担意識が強まれば、手軽にできるインフォーマルな評価までも敬遠するようになり形成的評価の活用機会が失われる。

説明責任や証明と結びつきハイスティクス化した評価では、【教師が唯一の評価者という意識】を強める。構成主義の学習観に支えられた形成的評価では、学習者自身も自らの学習に介入し調整する役割を負っており、教師と学習者でその責任が共有されると考えられている (Harlen, 2005; William, 2018)^{24) 25)}。しかし、総括的評価が支配的になると、その責任は教師が一手に担うことが要求され、自己評価や相互評価といった形成的評価が使用されにくくなる。以上のように、総括的評価に影響された教師の評価意識が背景となり、教師の形成的評価の活用を消極的にしている。

5 評価リテラシーの停滞

〈評価リテラシーの停滞〉は3つの概念から成る。【指導と分離した評価理解】は、指導のあとに評価があるという、指導と評価の時系列的な分離を示しており、ここでは学習を総括するようなペーパーテストが評価方法として好まれる。

【成績としての評価理解】は、教師が総括的評価の優位な環境に身を置くなかで、「評価＝成績」という様に評価を限定的に理解することを示す (石井, 2008)²⁶⁾。このような評価理解が基盤となって、生徒の成績を数値化しやすいペーパーテストへの傾倒をもたらし、形成的に機能する多様な評価方法の活用機会を狭める。

評価理解の総括的評価への偏重は、【形成的評価の技能不足】にもつながる。形成的評価ツールが各種考案されているが (例えば, Harlen & Deakin, 2003; William, 2018)^{27) 28)}、教師がこれらを実践的スキルとして身につけるためには、形成的評価に価値を置いた評価実践の蓄積と省察が必要である。しかし、総括的評価と密接に結びついたペーパーテストに依存してしまうと、教師は多様な形成的評価実践を経験する機会を失

い、実践と省察を通した形成的評価の力量形成がなされにくい。

6 学習者の総括的評価志向

〈学習者の総括的評価志向〉は3つの概念から成る。【学習者の数値志向性】は、学習者が得点や順位といった数値化したフィードバックを求めているという教師の認識によって、形成的評価の活用が妨げられることを示す語りである。【成果物の数値化】で論じたように、教師自身も数値化したフィードバックを好む傾向があるが、学習者の要求がこの傾向に拍車をかける。

【失敗のリスク化】は、評価が成績や選抜と結びつきハイスティクス化した環境下では、学習者はできないことをリスクと捉え、評価が下がるのを防ぐために教師の介入に消極的になることを意味している。形成的評価は、ブラックボックス化した学習者の学びのプロセスを明らかにし、つまづきに対して必要な対処を行い学習改善することを目的とするため、学習者が失敗の表出をリスクと捉え隠そうとすることは教師の形成的評価の活用の障壁となる。

【学習者の成果志向性】は、学習者が他者と比較して何点高かったのか、先生にどれだけいい評価をもらったのかといった学習成果の証明に偏重した状態を意味する。このような学習者の志向性を受けて、教師も学習成果を学習者のモチベーションの向上や学習を強制するために利用し、学習の本質的な理解を促進するような形成的評価の活用からは遠のく。以上のように、総括的評価は学習者の学習観にも大きな影響を与え、それを認識した教師が生徒の志向性に合わせることで、形成的評価の活用が妨げられている。

7 カテゴリー・概念間の関係

次に、カテゴリー・概念間の関係を教師の語りを引用しながら説明する (図1)。まず、〈評価意識による制約〉と〈フィードバック効果の弱体化〉の関係について、E教諭が「ちゃんとした、なんかきちっと説明できる評価をしないといけないってのがあって、曖昧な感じにならないようにするには一応『4』とか『3』とかっていう評価を書いて返すことが多いです。」と述べているように、教師は説明責任に耐え得る信頼性が高い評価へと意識づけられた結果、曖昧性の残る質的評価を敬遠し成果物を数値化してフィードバックする傾向を強める。

また、〈学習者の総括的評価志向〉と〈フィードバック効果の弱体化〉の関係については、F教諭が「多く

の生徒は、成績がどうかとかテスト順位がどうかとかそういうことを意識してますよね。だから、ただ『よかったよ』とか言ったりコメントつけたりするよりも、具体的に何点とか言った方が励みになるかなってのはあります。」と述べているように、学習者が数値志向や成果志向を強く持っていることも教師の成果物の数値化志向に影響する。また、K教諭が「テストとかレポートとかもう1回やるのは時間的にも厳しいですけど、生徒としては『もう一回やって成績変わるの』みたいに言う子もいて。」と述べているように、学習者が結果のよし悪しのみを重視することで、結果をどのように活用するのかという学習改善への意識が弱まり学習のやり直しの価値が低下する。

〈学習者の総括的評価志向〉と〈評価意識による制約〉の関係について、G教諭が「レポート評価のときに、まあ良い悪いくらいはあるんですけど、必ずしも細かく点数をつけられるわけではないんですけど、生徒とか保護者がそれを見ることを考えると曖昧なことは書けないし、うん、結構そこは負担になってるかもですね。」と述べている。学習者の数値志向や成果志向によって、教師は評価のエビデンスを示しアカウントビリティを果たさなければならないという意識を強く持ち、それが評価に対する過度な負担意識や責任意識の増大に繋がる。

〈評価意識による制約〉や〈学習者の総括的評価志向〉と、〈評価方法の限定化〉の関係について、A教諭が「ちゃんと評価付けるのって結構大変なので毎回はできません。(中略)まあ、テストは絶対あるので、それで成績つけるまあそうですね。」と述べているように、評価は負担のある行為という意識は、教師を慣れ親しんだペーパーテスト依存へと向かわせる。同様に、C教諭が「生徒が、とくにテストできる子とか女の子ほどそうかもしれないけど、わからないのを知られたくないみたいな意識があって、聞かれたくない感じはありますね。そうするとテストくらいでしかちゃんとした目安は持ちにくいかな。」と述べているように、学習者の失敗のリスク意識が、学習者の積極的な関与を要せず明確な数値として算出できるペーパーテスト依存をもたらす。

〈フィードバック効果の弱体化〉と〈評価方法の限定化〉の関係については、J教諭が「授業中の様子とか記録して成績に反映しようとしても、実際言われたこともあるんですけど、それは教師の好みみたいなものだからって感じで、そういうのを数値化するのは難しいので、テストだったらとりあえずズバリその点で

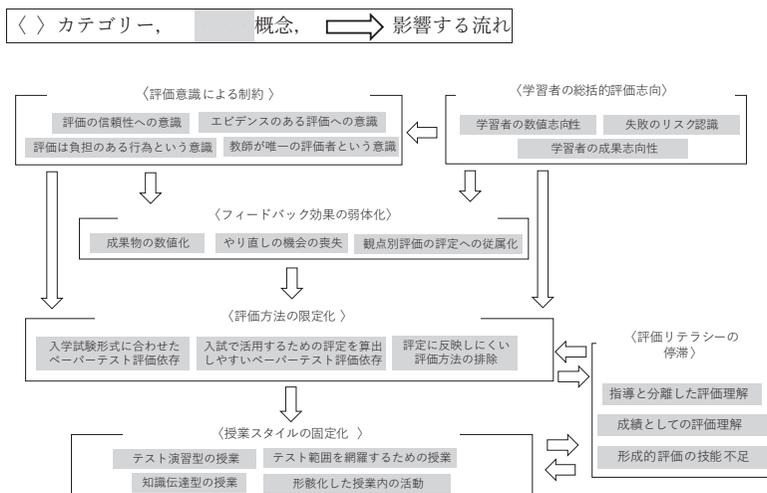
すって言えるので。」と述べているように、教師が成果物の数値化にこだわることにより、数値化することが難しい文脈依存的で質的な評価の活用を消極的にし教師が測定的なテスト評価依存に陥る。

〈評価方法の限定化〉と〈授業スタイルの固定化〉の関係については、I教諭が「ちょうど来週はテストがあるので、ちょっと範囲が終わらなそうなクラスがあったので、急いで範囲のページ・教科書までやります。」と述べているように、評価方法がテストに限定化されると、テスト演習を中心とした授業やテストに関連した知識伝達の授業を行い、テスト範囲を網羅するための授業へと繋がる。

また、〈評価方法の限定化〉と〈評価リテラシーの停滞〉の関係については、J教諭が「テスト作りはかなり慣れましたけど、確かにそれ以外の評価のスキルはついていないような。」と述べているように、テスト評価依存によって、形成的評価に必要な評価スキルを向上する機会を失い、形成的評価の技能不足へとつながる。これは、逆も然りであり、K教諭が「テスト点以外でどう成績をつけたらいいんだろうってのがずっとありますね。」というように、評価技能の不足が故に、教師が慣れ親しみ容易に数値化できるペーパーテスト依存に陥ることもある。また、A教諭が「評価はとりあえず、一通り内容をやって確認テストしてみたいなルーティンがあるのでまあ基本はそれです。」というように、指導と分離した評価理解が脱文脈的なペーパーテストへの偏重に繋がる。

最後に、〈授業スタイルの固定化〉と〈評価リテラシーの停滞〉にも相互の影響関係が見られる。I教諭が「授業中にどのように見取ったらいいのかわからないところもあって、あとそれが評価的のどの程度正しいのかとか、そうするととりあえず小集団で話し合ってみたいにしてもそれで終わりみたいな。」と述べているように、形成的評価の実践の難しさや自信のなさが、評価の伴わない活動主義的な形骸化した活動をもたらす。また、B教諭が「まあ、実際に授業でいろいろ試してみないと、うまくいかなくてもね、やっぱり試してみないとできないままだね。」と述べているように、授業実践での評価経験の蓄積がないと、評価リテラシーも高まっていけないことを示している。以上のように、総括的評価は教師の形成的評価の活用に影響を与えている。

図 1 総括的評価による形成的評価への阻害要因の結果図



B. 総括的評価による形成的評価への阻害要因の解消事例

次に、総括的評価による形成的評価への阻害要因を解消する具体的方略を教師の語りから抽出し、先行研究で示した阻害要因を解消する3つのアプローチ（総括的活用型、形成的活用型、評価の分離型）を枠組みに整理する。そして、上述のM-GTAにより生成した、

総括的評価による形成的評価への阻害要因を示す20概念と各事例との関係を検討する。表3に、各事例と20概念の関係を示し、総括的評価による形成的評価への阻害要因を解消する可能性がある項目には●印をつけた。なお、文中の○数字は表3中の概念を指している。

表 3 総括的評価による形成的評価への阻害要因を示す概念と各事例の関係

概念	I 教諭の事例	G 教諭の事例	B 教諭の事例
①テスト演習型の授業			●
②知識伝達型の授業			●
③テスト範囲を網羅するための授業			●
④形骸化した授業内の活動			●
⑤入学試験に合わせたテスト評価依存	●		●
⑥入学試験で活用するための評定を算出しやすいテスト評価依存	●		●
⑦評定に反映しにくい評価方法の排除			●
⑧成果物の数値化	●		
⑨やり直しの機会の喪失		●	
⑩観点別評価の評定への従属化	●		
⑪評価の信頼性への意識	●		●
⑫エビデンスのある評価への意識			●
⑬評価は負担のある行為という意識			
⑭教師が唯一の評価者という意識		●	
⑮指導と分離した評価理解	●		
⑯成績としての評価理解	●		
⑰形成的評価の技能不足	●		●
⑱学習者の数値志向性	●	●	
⑲失敗のリスク化	●	●	
⑳学習者の成果志向性	●	●	

●印は、各事例が、総括的評価による形成的評価への阻害要因を解消する可能性がある項目を示す

1 形成的評価の総括的活用（総括的活用例）

「総括的活用例」は、形成的評価を第1目的として収集した評価情報を、総括的評価としても利用する方略である。この方略に該当する事例は4事例あった。代表的にI教諭の事例を以下囲みで紹介する。

研究者 I教諭	授業内ではどのような評価をしていますか。そうですね、頻繁にやるのは毎授業の最後に書かせている「振り返りシート」ですかね。子どもたちには、その授業で学んだこととか気づきとか、あと意見・質問などを記入してもらってます。あと、たまに私から授業でやった内容に関する具体的な問いとか課題を与えて、それについて書かせることもあります。
研究者 I教諭	振り返りシートは、その後どうされますか。授業後に集めて、一人一人に簡単なコメントを書きます。で、次の授業の最初で返ししながら、少し気になったこととかおもしろい感想とかを紹介したりはします。
研究者 I教諭	振り返りシートに記述されていた、生徒の質問とかに答えたりもしますか。しますね。何か、まあその質問の意味を直接口頭で聞いたり、何がわからないのかをより具体的に聞いたりして、まあその説明をしたりしますね。あとまあ、その疑問を授業での問いにしたりもできるし、子どもたちはここがわかりにくいんだってわかるので、他のクラスで授業するとき参考にしてたりとかはあります。

I教諭は、振り返りシートに書かれた生徒の感想や疑問は理解度の把握の手がかりとなり、別のクラスでの同授業の実践や次の授業の進め方を決める参考資料となる。また、記述内容について直接生徒に詳しく聞いたり口頭で質問に回答したりしており、振り返りシートが教師と生徒間の対話を生み出す形成的評価ツールとして機能している。このように、I教諭は振り返りシートを学習者の学習理解の把握、フィードバックの提供、授業改善を実施するために利用しており、形成的評価を第1目的としているといえる。

以上の、I教諭の事例と総括的評価による形成的評価への阻害要因を表す概念との関係を見ると(表3)、I教諭は形成的評価を第1目的として振り返りシートを使用しているため、フィードバックを数値化するような事態には陥りにくく(⑧)、生徒の数値志向性も抑制する(⑱)。また、毎時間の振り返りシートは日々の授業での生徒の様子を文脈の中で見取りその変化を中長期的に把握することを可能にするため、評価の信頼性への過度な制約意識を緩め評価の妥当性にも目を向けることや(⑪)、妥当性のある分析的な観点別の評価を行うこと(⑩)、テスト評価依存から脱却すること(⑤⑥)ができる。加えて、授業内で実施される質的な評価は、指導と分離した評価理解(⑮)や成績

としての評価理解(⑯)から解放され、学習状況を把握しフィードバックする力量を高める(⑰)。また、教師が生徒の意見や質問に価値づけを行うことで、学習者は疑問点を表出することは学びのプロセスの重要な過程と考えることができ、失敗をリスクとして捉えることや成果志向性からの脱却を促す(⑲⑳)。

2 総括的評価の形成的活用（形成的活用例）

形成的活用例は、総括的活用例とは反対に総括的評価を第1目的として収集した評価情報を、形成的評価としても利用する方略である。この方略に該当する事例は5事例あった。代表的G教諭の事例を以下囲みで紹介する。

研究者 G教諭	テスト返却はどのようにしていますか。そうですね、あの毎回ということではないけど、テスト返しの時に、論述問題でよかった解答を、実際の生徒の解答として見せることはありますね。教師が見つけた模範解答よりも、子どもの言葉で書かれてるからいいかなと思って。
研究者 G教諭	その、解答を見せた後は何かしますか。ええ、生徒はそれを読んで、私が採点基準をいくつか挙げているので、内容とか論理性とか、そういうのと照らし合わせながら、その解答をみてなぜいいのか考えたり話し合わせたりっていうのをします。この活動を入れることでどんな効果がみられますか。
研究者 G教諭	そうですね、まあこれをして点数がどうなるとかではないんですけど、たぶんこれをした後に答案返却をすると、ただ点数みて終わりみたくならず、じゃあ自分の解答はどうかみたい、さっきの他の人の解答と比べながら振り返る機会になるんじゃないかなと思います。また、実際に書き直しもしたりするので、それを見てあげてってこともします。

G教諭は、テスト実施後に採点過程に学習者を関与させる方略を使っている。特に、論述形式の問題について、教師の採点結果を返却する前に実際の生徒の論述から見本となる事例を抽出し、サンプルとして提示している。生徒たちは、サンプルとして提示された論述を読み、採点基準と照らし合わせながらどこがどのように優れているのかを検討する。このような活動を経た後に答案返却をすることで、サンプル検討を通じて理解を深めた評価基準に照らして自らの解答を振り返ることができる。

以上のような、D教諭の事例と、総括的評価による形成的評価への阻害要因を表す概念との関係を見ると(表3)、テスト後にサンプルを基に評価基準への理解を深め、それを受けて論述のやり直しの機会を創造することで(⑨)、テスト結果を学びの終着点ではなく次の学びに向かうための学習経過の一部と位置付ける

ことを可能とし、学習者ができないことをリスクと捉えたり (19)、数値や成果のみに捉われたりする状態から脱却する (18)(20)。また、教師の評価をそのまま受け入れるのではなく、自らの評価基準の解釈と照らし合わせて内在化していくことで、学習に対する責任を教師と共有することができ、教師が唯一の評価者であるという意識からも脱却できる (14)。

3 評価機能の分離 (分離型)

「分離型」は、授業における形成的目的の教師の評価を説明責任や証明の手続きから完全に切り離すことで、総括的評価による形成的評価への阻害要因を解消する方略である。この方略に該当する事例は2事例あった。代表的にB教諭の事例を以下囲みに紹介する。

研究者 B教諭	教室で評価するとき特に意識していることはありますか。 私は、とにかく、まあ評価は大事なのはわかるんだけど、あまり評価を意識して勉強させたくないというか、そういうことをあまり気にせず自由に学んでほしいなどというのがあるんですよ。だから、授業の中では、評価とかいう言葉は使わないで、でも生徒が何しているかとかは教師がちゃんとみないとだから、机間巡視したりしてアドバイスしたり、何書いているかを見たりとかして、まあ評価というかチェックしたりしますよ。
研究者 B教諭	それが成績評価と関係したりはしますか。 その、授業での今言ったようなのは、あくまでその時その時の確認なので、直接的に成績どうこうってことはないですね。記録もしてません。(中略)あと、単元テストとか小テストもやったりしますが、それもあくまで確認テストなので、成績には使わないよってことは生徒にも言ってます。逆に、定期テストとか、あと調べ学習の最終レポートとかは、その評価が成績になるよってことは言ってますね。そこをあえて明確にしたいなあってのがあります。

B教諭は、テストや成績に縛られた授業や学習は好ましくないという考え方をもち、授業内では極力テストや成績を意識させないという。そのため、単元テストや小テストはあくまでも理解の確認や定着のために実施し、一切評定には利用しないことを生徒とも共有している。一方で、調べ学習のまとめレポートや定期テストを評定算出のデータとして活用しておりこのことも生徒と事前に共有している。

以上のB教諭の事例と、総括的評価による形成的評価への阻害要因を表す概念との関係を見ると(表3)、B教諭は評価のための授業をすることを意図的に避けテスト演習に終始するような授業はしない(1)。むしろ、授業ではなるべく評定やテストに委縮せず生徒が

自由に思考し表現できる機会を創ることを目指しており、知識伝達型の授業や形骸化した授業内の活動に陥ることを防いでいる(2)(4)。また、成績評価にはテスト点を利用しているが、それを教室での教師の評価の代替とはせず、教師の質的な評価を形成的に利用することにも価値を置いている。これは、テスト評価依存から脱却し(5)(6)、評定化しにくい評価方法の排除を抑制する(7)。また、評定やテストと切り離して教師が日常的に形成的評価を使おうとすることで、アカウンタビリティの要請に基づく、評価の信頼性やエビデンスへの過度な意識を軽減し(11)(12)、見取りやフィードバックといった形成的評価技能が実践的に磨かれていく(17)。このように、分離型は授業スタイルや多様な評価方法の選択において有用性が確認できる。

4. 考察

本研究を通して、ペーパーテストや評定と密接に結びついた総括的評価は、授業スタイル、評価方法、フィードバック、評価意識、評価リテラシーにおいて教師の形成的評価の活用に影響していることが明らかになった。また、総括的評価が優位な環境で学習を行うことで、生徒も総括的評価志向をもち、それが教師の形成的評価の活用に影響を与えるという、生徒を媒介とした総括的評価と形成的評価の関係も確認できた。

総括的評価と形成的評価の関係の問い直しが迫られている今日において(西岡, 2015)²⁹⁾、これまで十分に研究されてこなかった日常的に形成的評価を行う主体である教師の語りからボトムアップに総括的評価による形成的評価への阻害要因について明らかにしたことは意義がある。なぜなら、教師がどのような授業や評価方法、フィードバックを選択するのか、またその背景となる評価技能や評価意識が教師の形成的評価にどのように影響を与えるのかなど、教師の経験やその認識に基づいた実際の教授・学習のプロセスに即して、総括的評価と形成的評価の関係を示す知見となったためである。また、総括的評価による形成的評価の阻害要因を解消する方略について、その方略を示すだけでなく、実際の教授・学習プロセスにおける阻害要因と対応させてその有用性を検討し、教室の評価に根差した実践的な知見を提供できた。

また、教師の中で形成的評価は質的な評価と、総括的評価は定量的な評価と不可分に結びつき、教師の評価活動を方向付けていることもわかった。これは、総括的評価と結びついたペーパーテストについては、「信頼

性]や「公平性」の高さといったメリットがある一方で、テストに向けた指導に授業が矮小化されること(石井, 2020)³⁰やテストが引き起こす学習行動の危機(村山, 2006)³¹などのこれまで指摘されてきた課題を示すものである。こうした従来から指摘されているペーパーテスト問題について、総括的評価による形成的評価への阻害要因という視点から検討し、総括的評価による形成的評価への負の影響を超克する方略の有用性を示せたことも本研究の重要な知見である。一方で、本研究は先行研究で示した総括的評価による形成的評価への阻害要因の3つの解消方法の有用性を示すにとどまっているため、今後既存の方略に当てはまらない方略について実証的な研究を進めていくことが求められる。

最後に、今後の課題を2点指摘する。1点目は、本研究は11名の非技能教科の公立中学校教員を対象とし、その共通性から概念やカテゴリーを生成したが、形成的評価の活用には教科の特性や管理職の方針なども影響していると思われる。今後、教科の領域固有性や管理職をはじめとした学校文化の特徴に着目して、形成的評価と総括的評価の関係を質的及び量的に研究することが必要になる。

2点目は、学習者側の視点から研究を進めることである。本研究により、学習者の数値志向性や失敗をリスクと捉える傾向が教師の形成的評価の活用に影響を与えていることを示したが、学習者が教師の評価方法や評価意識によって受けている影響については検討できていない。例えば仮説として、教師のペーパーテスト依存や知識伝達型授業が子どもを総括的評価志向に向かわせ、さらにそれが教師の総括的評価志向に拍車をかけているといった相互作用の関係が考えられる。形成的評価の効果的な活用を実現するには、教師が学習者に与える影響についても一層研究を進める必要がある。

注釈

- i) 学習者の将来や進路の選択に大きな影響を及ぼすこと

参考文献

- 1) Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004) Working inside the black box: Assessment for learning in the classroom. *Phi delta kappan*, 86(1), 8-21.
- 2) Organisation for Economic Co-operation Development, & Organisation for European Economic Co-operation. (2005) *Formative Assessment Improving Learning in Secondary Classrooms / Organisation for Economic Co-operation and Development*. Paris: OECD Publishing.
- 3) Black, P., & Wiliam, D. (1998) Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74.
- 4) Bennett, R. E. (2011) Formative assessment: A critical review. *Assessment in education: principles, policy & practice*, 18(1), 5-25.
- 5) 佐貫浩 (2012) 評価論をめぐる論争点の検討 (No. 1) 田中耕治氏・中内敏夫氏の評価論の検討. *生涯学習とキャリアデザイン = 生涯学習とキャリアデザイン*, (9), 87-110.
- 6) 西岡加名恵, 石井英真, & 田中耕治. (2015). 新しい教育評価入門一人を育てる評価のために. 東京: 有斐閣コンパクト.
- 7) Harlen, W. (2004) A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes (EPPI-Centre Review), *Research Evidence in Education Library*, issue 3 (London, EPPI-Centre, Social Science Research Unit, Institute of Education). Available on the website at: http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=reel/review_groups/assessment/review_three.htm.
- 8) Harlen, W. (2005) Teachers' summative practices and assessment for learning – tensions and synergies, *Curriculum Journal*, 16(2), 207-223.
- 9) Stiggins, R. J. (1999) Assessment, student confidence and school success, *Phi Delta Kappan*, 81(3), 191-8.
- 10) Linn, R. (2000) Assessments and accountability, *Educational Researcher*, 29, 4-16.
- 11) 石井英真 (2020) 現代アメリカにおける学力形成論の展開——スタンダードに基づくカリキュラムの設計, 東信堂.
- 12) 前掲5)
- 13) 前掲8)
- 14) 前掲1)
- 15) 前掲8)
- 16) やまだようこ (2007) 質的心理学の方法, 新曜社.
- 17) 木下康仁 (2003) *グラウンデッド・セオリー・アプローチの実践——質的研究への誘い*, 弘文堂.
- 18) 木下康仁 (2007) *ライブ講義M-GTA 実践的質的研究法——修正版グラウンデッド・セオリー・アプローチのすべて*, 弘文堂.
- 19) 前掲17)
- 20) 前掲11)
- 21) 文部科学省 (2018) 学習評価に関する資料.
- 22) 田中耕治 (2008) 学力調査と教育評価研究 (〈特集〉学力政策と学校づくり). *教育学研究*, 75(2), 146-156.
- 23) Wiliam, D. (2018) *Embedded formative assessment*. Solution Tree Press (Second ed.)
- 24) 前掲8)
- 25) 前掲23)
- 26) 石井英真 (2008) アメリカの思考教授研究における情意目標論の展開: 「性向」概念に焦点を当てて. *教育方法学研究*, 34, 25-36.
- 27) Harlen, W. & Deakin Crick, R. (2003) Testing and motivation for learning. *Assessment in Education: principles, policy & practice*, 10(2), 169-207.
- 28) 前掲23)
- 29) 前掲6)
- 30) 前掲11)
- 31) 村山航. (2006). テストへの適応 教育実践上の問題点と解決のための視点. *教育心理学研究*, 54(2), 265-279.