

# 現代米国の「教師の効果」(teacher effectiveness) 測定 をめぐる論争状況

—Value-Addedモデルに関する誌上討論を手がかりに—

西野倫世<sup>†</sup>

Recent Debates over Measuring “Teacher Effectiveness” in the U.S. :  
Focusing on Journal Discussions on Value-Added Models

NISHINO Michiyo

## 1. はじめに

2000年前後の「学力低下」論争等を契機として学力向上政策が展開する中、わが国では学力テスト結果に対する「教える側の貢献度」への注目・期待が高まってきた。例えば、2007年以降実施されている全国学力・学習状況調査をめぐっては、テスト点数の高低やランク付けなど結果の公表が争点とされがちであり、「教える側」の責任を厳しく追及する動きもみられる。とはいえ、学力結果には家庭や地域の実状等が少なからず反映されており、学校や教師の努力・成果を正当に評価するものではない。こうした点を踏まえないまま上記の動向が進展すれば、学校や教師に際限ない努力が求められ、不条理かつ窮屈な教育を生む危険がある。

こうした問題を検討する上で示唆的なのが、現代米国の「教師の効果」測定をめぐる議論である。「教師の効果」(teacher effectiveness)とは、教師がいかに上手く仕事をしたかを高低で語ろうとする発想であり、米国では1920年代からその研究が緒に就き、1950年には米国で最も権威ある教育学会で研究委員会が立ち上げられる等、一世紀にわたり議論が蓄積されてきた(Lavigne & Good 2013)。本稿の立場から特筆すべきは、米国で厳しい学力テスト政策が進展した2000年代以降、「教師の効果」を問う際に学力テスト結果を重

---

<sup>†</sup> 大阪産業大学 全学教育機構 教職教育センター 准教授

草稿提出日 7月28日

最終原稿提出日 8月21日

視する趨勢がみられる点である。例えば、2002年の「どの子も置き去りにしない法 (No Child Left Behind Act : 以下NCLB法)」では、学力向上に対する「学校=教師集団」の責任が追及され、続く2009年の「頂点への競争 (Race to the Top : 以下RTTT)」プログラムでは、「教師個人」に対してまでも責任が問われるようになった。とりわけRTTTプログラムで教員評価結果を教師の報償・解雇等と結びつける「人事直結型教員評価 (high-stakes teacher evaluation)」(高橋 2021) が推奨され、そこで学力テスト結果を偏重する向きがある点については論争を呼んでいる。

このように、米国ではわが国よりも厳しいテスト政策が進展する中で深刻な問題も散見される一方 (cf : 北野・吉良・大桃 2012), 学力結果に対する「教師の効果」測定をめぐって慎重に議論が重ねられてきた点は注目に値する。その中心的論題となっているのが、Value-Addedモデルと呼ばれる比較的新しい測定原理である。その特徴については後述するが、これは統計技術やビッグデータに伴う測定技術の高度化を背景に、テスト点数のうち「教師が貢献した部分」を分離・析出する試みが制度化されたものと説明できる<sup>1)</sup>。米国では1990年代初頭にテネシー州で運用が開始され、とりわけ上述のRTTTプログラムで「人事直結型教員評価」の指標に同モデルが推奨された経緯等から、政策的な広がりを見せてきた。

では、連邦政策の影響下で学力テスト結果を通じた「教師の効果」測定が広がる中、具体的にどのような議論が展開し、何が論争点となってきたのか。この問題関心から本稿で着目するのが、Value-Addedモデルに関する複数の誌上討論である。米国では上述した連邦政策の展開に伴い、とりわけ2000年代以降同モデルに関する議論が盛んになってきた。こうした議論がまとまった形で刊行されることもあり、例えば2015年に刊行された米国教育に関する機関誌*Educational Researcher*の特集では、Value-Addedモデルをテーマにした共同討議が組まれている。この共同討議についてはわが国でも注目され、全国学会誌における海外の教育経営事情の一例として、米国の教員評価制度に詳しい藤村祐子が、いち早くその内容を整理・紹介した (照屋・藤村 2016)。なお、他の議論にも目を向けると、同特集よりも早い時期に、少なくとも3つの誌上討論が実施されていることを確認できる。時系列に即してそれらを整理すれば、第一の討論は、2008年にワシントンD.C.で開催されたワークショップの議論をまとめた報告書であり、別稿で既に紹介・検討した (小島 2014)。第二の討論は、2012年にSAGE Publicationsが刊行した米国教育における論争集で展開されたマホニー (Mahoney, J.) とキヌキャン=ウェルシュ (Kinnucan-Welsch, K.) らの論争である。そして第三の討論にあたるのは、季刊誌*Teachers College Record*で特集された2012年アリゾナで開催されたシンポジウムである。

後に詳しくみるように、第四の討論にあたる上述の*Educational Researcher*の特集も含めてこれらの誌上討論を比較・検討すると、第一の討論とそれ以降の討議ではその性質がやや異なっていることがわかる。つまり、Value-Addedモデルをめぐる議論は時期によって論点が異なっている可能性があるため、その展開を追ってこそ、基本的論点の洗い出しが可能になると思われる。第一の誌上討論については既述した通り検討済みのため、本稿では第二および第三の誌上討論を中心に取り上げる。以下、これらの内容を考察・分析することを通して、特に2009年の連邦政策の影響を受けながら米国では「教師の効果」測定をめぐるどのような議論の展開があり、何が基本的論争点とされているのか、その論争状況を整理・分析することを目的とする。

## 2. Value-Addedモデルをめぐる討議—支持派と慎重派の主張

まず取り上げるのは、上述した第二の誌上討論にあたるもので、米国の学術出版社SAGE Publicationsが2012年に刊行した米国教育に関する論争集(全10巻)のうち、第9巻『学校におけるスタンダードとアカウンタビリティ』に収録されている。その中の一つが、「Value-Addedモデルは優れて効果的な教師を識別するのに使用されるべきか」という主題であり、同モデルの「支持派」と「慎重派」の主張が対立的に描かれる構成となっている。以下では、「支持派」と「慎重派」それぞれの主張をみることで、Value-Addedモデルの基本的論争点を導出する手がかりとしたい。

### (1) 支持派の主張

Value-Addedモデルの「支持派」の立場から論を展開するのは、バテル子ども財団(Battelle for Kids)の代表を務めるマホニー(Mahony, J.)である。マホニーが主張する同モデルの意義は、次の3点にまとめられる(Mahony 2012)。

第一の意義は、Value-Addedモデルが“教育の景観”(landscape)を大きく変えた点である。例えば、従来の議論では“優れた教師”の「構成要素が何か」を規定することが論点になっていたのに対し、今日の議論では“優れた教師”が児童生徒の学習に与える影響に関して客観的かつ信頼たりうるよう「どう測るか」ということに焦点が移行した。とりわけ本稿の立場から示唆的なのは、マホニーが述べる次の見解である。すなわち、彼によると、米国の歴史を振り返れば、そもそも初期の頃には優れた教師(great teachers)の判断材料が、児童生徒の「成績」ではなく、その「成長度」であったという。もっとも、ここでは出典が明記されていないため、この見解がいかなる論拠に基づいているかは定かではない。けれども、仮にこの前提からValue-Addedモデルを捉え直せば、学力テストとい

う限定的範囲ではあるものの、まさに「成長度」を規準に据えて「教師の効果」測定を試みるという点で、米国教育史における教員評価の初発の原理が復権されたものとも理解できる。

第二の意義は、教育者に強力な士気を喚起しうる点である。詳しくは第一の討論で検討したが、同モデルの評価規準は“発達可能性を踏まえた伸び”であり、児童生徒の入学時の学力状況を考慮する唯一の測定原理である。この特徴から、たとえ教育困難校に勤める教師であっても、子どもの“学力伸長”に貢献できていれば、その努力や成果が識別されうる仕組みとなっている。したがって、「素点」などその他の測定原理では十分捉えられなかった「教師の効果」が可視化され、それゆえ教育者の士気向上につながると考えられているのである。この特徴は、米国の教員採用をめぐる問題状況に照らした際、いっそう重要な意義を有するよう思われる。というのも、わが国とは異なり、米国では各学校の校長が教職員をリクルートすることが通例であるため、比較的裕福な家庭が多い郊外の学校には“優秀な”教師が集まる傾向にある一方、貧困家庭の割合が高い都市部の学校では教師の質保証や離職率の高さが深刻な問題となってきたからである。だからこそ、たとえ教育困難地域で働いていたとしても、自らの教育成果が承認されうる同モデル固有の特徴は、教師の効力感形成に寄与し、ひいては離職防止や勤続につながる可能性があるという意味で、大いに注目に値する。

第三の意義は、Value-Addedモデルの提供する“レンズ”によって、学校教育におけるカリキュラムや指導、プログラム、実践の長短を発見でき、教師の専門性開発に活用しうる点である。この論考では詳しく述べられていないが、この意義は次のように解釈できる。まず、Value-Addedモデルの評価規準は“学力伸長”であるため、これが“レンズ”として機能することによって、子どもの学力を伸ばすことが得意な教師を発見できる。次に、その教師が実践するカリキュラムや指導法等を手がかりとして、同僚間で効果的な教授について話し合う機会が創出する可能性がある。したがって、広く教師の専門性開発につながりうるというわけである。

以上が、マホニーが挙げたValue-Addedモデルの意義である。とはいえ、次に挙げる諸課題から、これらの意義は潜在的なものに止まっていると付言されている点には注意が必要だろう。その課題とは、①教授＝学習 (teaching=learning) は複雑であるため、教師の力量を判断する上で同モデルを唯一の指標としてはならない、②教師はデータに基づく意思決定に関する研修をほとんど受けていない、③プログラムの改善や児童生徒の学力向上に対して同モデルの結果が利用可能な情報であったとしても、教師が事後分析としてデータをみることはほとんどない、というものである。つまり、同モデルは上記の潜在力を秘

めている一方、前提条件や使用状況上の課題から、現時点ではその適切な使用方法を理解するための機会が十分整えられているとは言い難いと理解できる。

以上からマホニーの主張を総括すれば、次のように結論できる。すなわち、Value-Addedモデルは児童生徒の学力伸長に焦点化するという特性から、上記の潜在的意義を有するものであり、協同的な対話の触媒、あるいは学力テスト結果をより意義深いものにする「手段」となりうる。だからこそ、同モデルに基づいて“効果的な教師”を識別し、その教師が教室で何をしているかについて知ることが、全ての教師をより成功に導く支援の出発点になる、と。なお、マホニーがこの結論に至ったのは、彼が同モデルの“成功例”を知っていたからと考えられる。彼はこの誌上討論に先立って執筆した別稿において、テネシー州チャタヌーガ市の学校改善事業を紹介し、その中で同モデルが果たした役割に注目している(Mahoney 2006)。この事例については別稿で検討したが(西野 2016)、同事例を知っていたからこそValue-Addedモデルの潜在力を主張し、同時にその潜在力発現の障害となる現状を問題視していたと推察できる。

## (2) 慎重派の主張

Value-Addedモデルに対して慎重な姿勢を示すのは、デイトン大学のキヌキャン=ウェルシュら(Kinnucan-Welsh, et al. 2012)である。ここでは議論の前提として、まず次の2点を確認された。一つは、児童生徒の学習に関する研究およびアカウンタビリティの中心に「教師」が位置づけられたのは、ごく最近であること。そしてもう一つは、アカウンタビリティの近年の方向性が「教師の効果」の定義において児童生徒の「学力」を最も重要な位置に置いたことである。すなわち、近年のアカウンタビリティ政策において「教師の効果」への注目が集まり、中でもテストで測れる「学力」が重視されていると理解できる。

キヌキャン=ウェルシュらはこの前提を出発点としながら、①Value-Addedモデル自体の問題点、②教員評価の中心に学力テスト結果を据える今日の趨勢の2点について、次のように批判した。以下、その内容をみていこう。

まず、①については、3つの次元から重層的な批判が加えられた。第一の次元は、Value-Addedモデルの制度設計自体がはらむ問題であり、次の3つの課題が指摘された。すなわち、同モデルは、(a)推定値の妥当性・信頼性を支える根拠が欠如している、(b)望ましい教育目標を十分に代理していないテストに対する不信がある、(c)児童生徒は教師・学校に無作為に選出・割り当てられるという原則に反して、教師・学校・学区の効果を児童生徒の学力に帰するという軽率な推論を含む、というものである。これらの課

題を鑑みれば、同モデルには教師の昇進やテニユア、雇用・解雇等に関わるハイ・ステイクスな決定には十分対応しえないという根本的問題があると言えよう<sup>2)</sup>。にもかかわらず、政策的には同モデルの結果に基づく「人事直結型教員評価」が進展しつつあると批判された。

第二の次元は、政策レベルの問題である。現在米国には数種類のValue-Addedモデルが存在しており、例えば社会経済的状况や過去のテスト結果、人種・性別・母語・転校等の要因を考慮するモデルもあれば、過去の教師の影響も含むモデルも存在するなど、それぞれ目的や前提が異なるという。そのため、諸州で異なるモデルが用いられている場合、教師に対するValue-Addedモデルの結果も州によって大きく意味合いが変わるはずである。しかし、実際のところ、為政者が当該モデルを使用する方法や目的を結合させているかは定かでないという。すなわち、キヌキャン＝ウェルシュらが問題視しているのは、例えばRTTTプログラムの競争的資金獲得に向けて導入しやすいモデルを採用するといったように、政策意図と選択されるモデルの不整合についてであり、教育目標や課題に応じて適切なモデルを選択するという本来の趣旨に即さない利用実態が広がっていることに対して批判が加えられた。なお、この点について経済学の観点から教育問題を研究する前述のハヌシュエックに尋ねたところ、Value-Addedモデルが複数存在する背景として、統計解析を専門とする民間企業の利益追求の問題があると指摘していた<sup>3)</sup>。

第三の次元は、学校現場における運用レベルの問題である。キヌキャン＝ウェルシュらによれば、学校や諸州の実状では児童生徒と教師を結びつけるデータの質に深刻な問題があるため、測定結果の信頼性が損なわれているという。例えば、Value-Addedモデルで「教師の効果」を測るには膨大な量のデータが必要不可欠だが、都市部の場合は児童生徒の転入出が多く、特定の教師が同じ児童を1年間担当するとは限らない等、現行の州データシステムでは技術的な限界も多い。その意味で、現時点で質の高いデータを収集することは困難であると指摘された。

次に、②については、教員評価の中心に学力テスト結果を据える今日の趨勢に対して批判が加えられた。上述した三次元にわたる諸問題を鑑みれば、教師の処遇等ハイ・ステイクスな決定の根拠としてValue-Addedモデルを用いるのは不十分と言わざるを得ず、同モデルを用いて「効果的な教師」を識別することについても慎重な姿勢が示されたのである。その上で、「教師の質」をより良く問うには、その人柄および効果的な教師 (teacher) と効果的な教え (teaching) の関係を含む「複合的な見方」が必要であると彼女たちは主張した。すなわち、教授に関する研究では、以前は個々の教授行為と児童生徒が教授される環境を関連づけていたのに対し、近年では教授と学習をめぐる経験的知識に基づく見方

を土台として、より全体論的な観点が包含されてきた。例えば、効果的な教授を特徴づける試みを通して、教室におけるパフォーマンスを視察したり記録に残す等、教育実践の産物に基づいて教授の質を評価することを企図したツールも開発されている(cf:石井2015)。こうした蓄積があるにもかかわらず、「教師の質」に関する昨今の定義は、児童生徒の学習の一助となる学校の文脈や教育実践、そして教師の特質等を看過するものであるとの批判が加えられた。ここに、「教師の効果=学力テスト結果」とみなす今日的趨勢への批判を見て取れる。

以上から、キヌキャン=ウェルシュらの結論を総括すれば、次のように理解できる。すなわち、Value-Addedモデルに基づく教員評価をめぐる近年の論調は、「効果的な教師」の意味内容が、無前提に「学力テスト」と直結して捉えられており、本来「教師の質」に内在している多義性や異質の類型化の可能性が十分配慮されていない。昨今の風潮のように、学力テスト結果に基づき「効果的な教師」を識別することは、教師の性格や特質等を看過し、「教師の質」を矮小化することになるのではないか。こうした問題意識から、キヌキャン=ウェルシュたちは、児童生徒の学習に寄与するもの・ことに関する理解を発展させる上でValue-Addedモデルが重要な役割を担うという議論を支持できる一方、こうした論調は本来「教授の質」を測る他のアプローチと結び付けて活用する方向に向かうべきと述べ、教員評価や給与、アカウントビリティへの使用を求める昨今の趨勢については支持できないと主張した。

### (3) 連邦政策の展開に伴う基本的論点の変化

マホニーとキヌキャン=ウェルシュらの討論から、米国における「教師の効果」測定の展開を読み取れば、①Value-Addedモデルの評価対象が「学校=教師集団」から「教師個人」へと移行し、②さらには教師の処遇を左右するハイ・ステイクスな決定において同モデルの結果が重視されている点が明らかになった。

この背景には、言うまでもなく連邦政策の存在がある。2002年のNCLB法では学区や学校単位、つまりは「教師集団」の教育責任が問われていたのに対して、2009年のRTTTプログラムによってその対象が「教師個人」へと細分化された。とりわけRTTTプログラムでは、教師個人の「効果」を問う上で、①Value-Addedモデルをはじめとした学力テスト結果における「児童生徒の学力成長度」(student growth)を教員評価の指標に採用し、②さらに教員評価結果を教員処遇と結びつけることを競争的資金獲得に向けた要件の一部としたため、「教師の効果」測定が「人事直結型教員評価」という議論の俎上に載った。ゆえに、この共同討議が展開された2012年時点においては、Value-Addedモデルをハイ・

ステイクスな教員評価の文脈で論じる素地が必然的に生じており、だからこそ同モデル固有の問題を越えて、広くその前提となる学力テストの使用へと論点が拡張してきたと把握できる。

この政策上の展開に伴い、教師の身分保障の問題ないし“よい教師”の定義を学力テストに一元化／矮小化してしまうことの是非という論点から、Value-Addedモデルに対して批判が高まってきた点を確認できた。キヌキャン＝ウェルシュらの議論から捉えれば、ここで批判されているのは同モデル自体がはらむ問題はもとより、より広く学力テスト結果を通じて「教師の効果」を捉える昨今の風潮と理解すべきだろう。つまり、この誌上討論で争点となったのは、学力テスト結果を教員評価の中心に据えることの是非であると判明した。

### 3. 「人事直結型教員評価」への利用に伴う問題

続いて、季刊誌 *Teachers College Record* の特集に焦点を当て、「教師の効果」測定に関する近年の論争状況を確認する。同特集は、2012年アリゾナで開催された「人事直結型教員評価：高い費用－大きな損失」と題した大会の議論をまとめた7本の論文から成り、①歴史的な分析、②信頼性・妥当性に関する問題、③より広域な政策関連等の観点から整理されている (Lavigne, Good & Marx 2014)。この中でValue-Addedモデルに焦点化した論文は4本と半数以上を占める点から、教員評価の文脈において同モデルへの注目が高まっている点を確認できる。以下では、この4本のうち最も包括的に議論を展開する2つの論考を中心的に取り上げ、その他の論考も補足的に用いながら、近年の論点について考察・分析を行う。

#### (1) 「教師の効果」測定モデルの全米使用状況

季刊誌 *Teachers College Record* の特集の中で第一に取り上げるのは、教育政策に詳しいアリゾナ州立大学コリンズ (Clarín Collins) とコロラド州立大学アムライン・ビーズリー (Audrey Amrein-Beardsley) の論考である。同論考は、Value-Addedモデルをはじめとした「学力成長に基づく教師の効果測定」が、全米でどの程度広まっているかを調査したものである (Collins & Amrein-Beardsley 2014)。

彼女たちの調査によると、2011年時点で学力成長を規準とする「教師の効果」測定モデルに関して使用・試行・開発のいずれかを行っているとは回答したのは、40州とワシントンD.C.であった。これは、全米の80%以上に相当する。そのうち、最も使用率の高いものが「成長度パーセンタイル」 (Student Growth Percentiles) で12州 (24%)、次いで



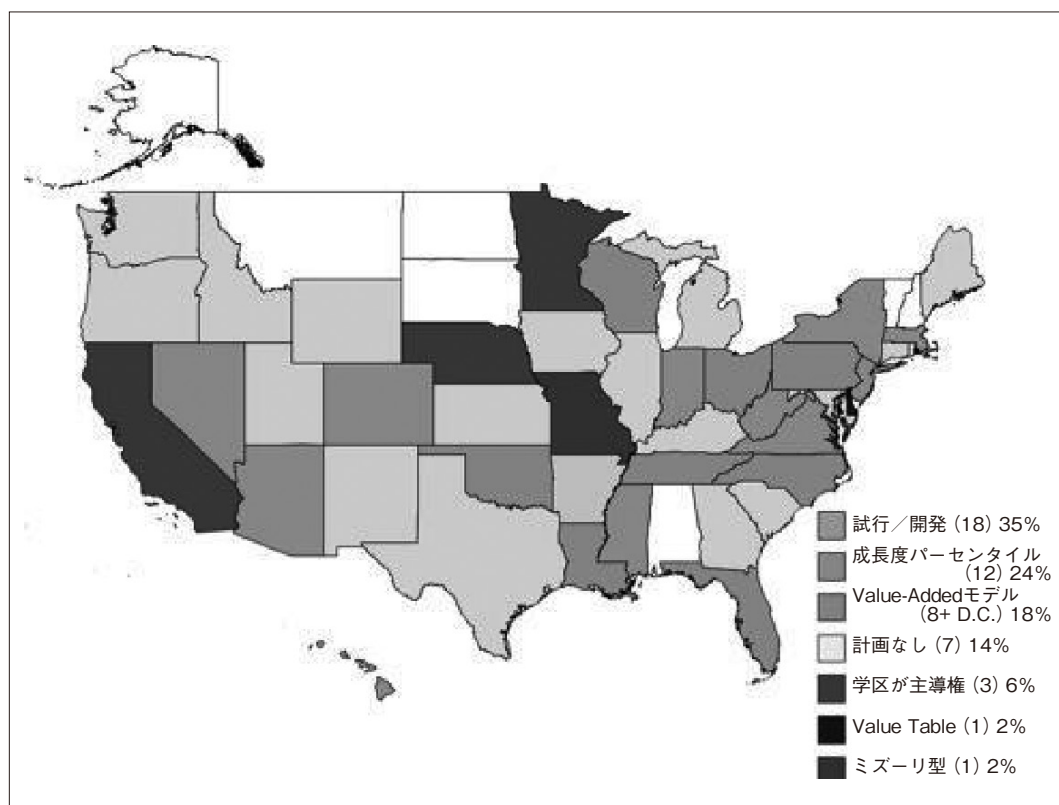


図1 学力成長に基づく測定モデルの全米使用状況 (2011年時点)  
(出所: Collins & Amrein-Beardsley 2014: 7)

Value-Addedモデルで8州とワシントンD.C. (18%) であった。その他、成長度パーセンタイルとValue-Addedモデルの両方を試験的に用いているのがミズーリ州 (2%)、Value Tableモデルを用いているのがデラウェア州 (2%) であった。なお、18州 (35%) については特定のモデル名を回答しなかったものの、何らかの測定モデルを使用中ないし開発中と述べた。これに対し、上記諸測定モデルの使用について、州規模ではなく学区レベルに委ねていると回答したのが3州 (6%)、学校単位では使用したとしても教師個人の評価を企図して使用する計画はないと回答したのが7州 (14%) だった (図1)。

## (2) 成長度パーセンタイル (Student Growth Percentiles)

では、最も多くの州で使用・試行されている成長度パーセンタイルとは、一体どのような測定原理なのだろうか。これは、教育関連企業であるCenter for Assessmentのベトベナー (Betebenner, D.) が2000年代後半に提唱・開発したとされる比較的新しい測定モデルである。成長度パーセンタイルを用いて「教師の効果」を算出する手順について概要のみ

述べれば、次のように説明できる (Betebenner 2008)。

まず、ある児童の今年度の成長度パーセンタイルの成績 (SGPスコア) は、州内で同じ学力レベルをもつ児童らと比較する中で算定される。具体的には、州統一学力テストで過去に同様の成績をとった児童が同じ母集団となり、その母集団の中で今年度獲得した点数が1位から最下位までランク付けされる。つまり、同様の学力レベルと思われる子どもたちの今年度の進捗状況と相対的に比べることによって、当該児童生徒の今年度のSGPスコアが算出されるという仕組みである。例えばある4年生のSGPスコアが67の場合、昨年度(第3学年時)に同水準の成績をとった児童の2/3よりも、当該児童は良いパフォーマンスを発揮したことがわかる (Goldhaber, Walch & Gabele 2014)。

次に、成長度パーセンタイルに基づく「教師の効果」については、次のように説明できる。すなわち、当該教師が担当する児童生徒をSGPスコアの順に並べた際、ちょうど中央に位置する児童生徒のSGPスコアが、直接その「教師の効果」を示す成績となる。例えば31人学級の場合、15番目に位置する児童のSGPスコアが67ならば、その「教師の効果」を示す成績も67になるという具合である。

以上から、成長度パーセンタイルはValue-Addedモデルと同様、「学力成長」を規準とする測定原理と言え、それゆえ同モデルの一種に分類する論者もみられる (Henry, Rose & Lauen 2014)。では、これらの相違点は一体何か。ここで重要なのは、Value-Addedモデルが「社会経済的要素を考慮する」という特徴をもつものに対して、成長度パーセンタイルはこの特徴をもたないとされる点である。そのため、理論上はValue-Addedモデルと異なる測定原理として分類することが通例であるという (National Research Council and National Academy of Education 2010)。加えて、Value-Addedモデルが「教師の効果」析出に際して多次元・多変量のデータを要するのに対して、成長度パーセンタイルで必要とされるデータは少なくとも過去2年分のテストスコアのみと、データの必要数においても相違を認められる。

もっとも、Value-Addedモデルは1990年代初頭から学校現場で運用が始まったため、実践面・理論面双方において蓄積が進んでいるのに対し、成長度パーセンタイルは2000年代後半から使用され始めた比較的新しいモデルであり、その検討は近年緒に就いたばかりである。成長度パーセンタイルの普及に伴い、両モデルを比較・検討する動きもあるものの、どちらが優位かについては未だ十分解明されていないという (Goldhaber, Walch & Gabele, op. cit.)。にもかかわらず、現時点で成長度パーセンタイルの採用数が多いのはなぜか。その理由として、ベーカー (Baker, B.) らは次の3点を指摘する (Baker, Oluwole & Green 2013)。第一に、連邦政策RTTTプログラムで教員評価への使用が推奨

されているモデルの一つであること。第二に、Value-Addedモデルよりも単純・明瞭なモデルであること。そして第三に、Value-Addedモデルの使用をめぐる課題についての研究蓄積が進んだことである(cf: Darling-Hammond, et al. 2012)。

第一の点に関しては、前節で述べたように、RTTTプログラム以降アカウンタビリティの対象が「教師個人」に細分化されたことに起因する。第二および第三の点については、ベーカーらの指摘によると、成長度パーセントイルが広がっているのはValue-Addedモデルよりも“優れた”測定原理と判断されたからではない。上述のキヌキャン＝ウェルシュらの主張のように、測定原理としての優劣というよりは、“わかりやすさ”や“無難さ”が採用の根拠となっている可能性が指摘されているのである。むしろ、こうした選定基準は教師たちの“納得感”につながる可能性があり、それゆえ既に多くの問題が指摘されるValue-Addedモデルと比べれば、使用にあたっての抵抗感は少ないかもしれない。他方、成長度パーセントイルはValue-Addedモデルよりも「教師の効果」を測定する上で“公平さ”が損なわれるという指摘もある点を踏まえれば(Walsh & Isenberg 2013)、成長度パーセントイルの妥当性や優位性についても、Value-Addedモデルと同様に、慎重な議論を要するだろう。

### (3) 「人事直結型教員評価」への使用の広がり

同論考における第二の論点は、Value-Addedモデル等を「人事直結型教員評価」に使用することの是非である。例えば、成長度パーセントイルやValue-Addedモデルの結果をテニユアと結びつけている(またはその予定の)州は、全体の22%(10州とワシントンD.C.)であり、このうち1州を除く9州とワシントンD.C.においては、解雇事由や給与の差異化にも用いているという。これに対して、「教師の効果」を数量化するという発想を完全に拒否し、連邦政策RTTTプログラムの基金獲得に向けてこうしたモデルを採用することに反対する州はわずか6%(3州)に止まった。

もっとも、成長度パーセントイルやValue-Addedモデルを適用できる教師は、学力テストのデータ制約等の問題から全体の1/3程度とされるため、その注目度の高さとは裏腹に、学校現場に与える影響はさほど大きくないと指摘する声もある(大桃・吉良・堀・宮口・金子 2016)。だが、教師の3人に1人が該当するという数字は決して少ないとは言えず、看過すべきでないという考えもありうるだろう。さらに言えば、当時これらの測定原理で用いるテストは、第3～第8学年の3教科(算数・国語・理科)に限られていたが、学力テスト結果のデータ蓄積に伴い、今後適用可能な教科・学年が広がる可能性が高いとも言われている。これらの点を鑑みれば、Value-Addedモデルをはじめとした「教師の効

果」測定結果を教員評価にどう位置づけるかについては、依然として重要な検討課題と考えられる。

なお、「人事直結型教員評価」については、同特集における別の論考においても言及がみられる (Lavigne 2014)。そこでは、新任教師の初年度の離職率は14%、5年目までの離職率は50%に上ることから、教師はストレスの高い職業であると確認され、さらに若手の離職率の高さに関しても指摘が加えられた。同論考によると、1988年度の新任教師の割合が13%だったのに対して、2008年度はその割合が34%に増加しており、近年では若手教師の離職防止がより差し迫った問題になっている。それゆえ、「人事直結型教員評価」は教師の専門性開発や学力向上といった本来の目的をむしろ弱める恐れがあると主張された。他方、こうした主張とは対照的に、冒頭で触れた第四の誌上討論においては「人事直結型教員評価」でなければValue-Addedモデルの活用が教師の資質向上につながらないと主張する論者もいる (Winters & Cowen 2013)。これらの主張をみただけでも、同モデルを「人事直結型教員評価」に使用することの是非に関しては、米国内でも賛否両論あることがわかる。

#### (4) 「教師の効果」測定をめぐる議論の変容

季刊誌 *Teachers College Record* の特集の中で第二に取り上げるのは、コンスタントプロス (Konstantopoulos, S.) の論考である。ここでは議論の前提として、Value-Addedモデルが注目されるに至った2つの背景が指摘されている。ひとつは、同モデルによって教師を評価・ランク付けする「枠組み」が提供された点、もうひとつは、「人事直結型教員評価」の文脈における使用が争点となってきた点である (Konstantopoulos 2014)。

まず、後者の内容から先に述べれば、「人事直結型教員評価」に学力テストを用いる動きが教師の資質向上や児童生徒の学力保障につながるかは論争的という趣旨であり、これについては上述の内容と重なるため割愛したい。他方、本稿の立場から特筆すべきは、Value-Addedモデルが教師を評価・ランク付けする「枠組み」を提供したという、前者の指摘である。これは、2つの意味で、米国教員評価制度の問題を示唆するものと思われる。第一に、測定結果の妥当性については後述の通り課題が残るものの、近年の統計技術の発展に伴い同モデルを用いることで“発達可能性を踏まえた伸び”が可視化されうるとの想定が現実味を帯び、だからこそ、その数値を教員評価に活用しようという発想が必然的に生じたこと。そして第二に、少なくとも近年では、教員評価結果を処遇と連動させることに対して、強いニーズが存在するということである。

コンスタントプロスによれば、1960年代のコールマン報告以降、学校も含めた「教師の

効果」に関して多くの研究が蓄積されてきた。その一例として、「教師の効果」は学校内／学校間で劇的に変化しうる点や、「教師の効果」は教育現場での経験や学歴等、従来の教員評価において重視されてきた典型的特性からは十分説明できないという指摘を挙げられる(Hanushek & Rivkin 2012)。こうした研究蓄積があったからこそ、ここ十数年で「教師の効果」に関する研究関心が「児童生徒の学習をいかに促進できるか」へと移行し、「教授(=teaching)は児童生徒のパフォーマンス向上に帰着すべき」という命題が生まれたというのである。

こうした理解のもと、コンスタントプロスはValue-Addedモデルに対して、次のような見解を述べた。すなわち、教授過程と授業実践の把握に向けて、教員評価はより包括的なものでなくてはならず、それゆえ他の指標と同様に、量的な変数も含む必要がある。その意味で、「教師の効果」を測定する上で多変量・多次元の要素を用いるValue-Addedモデルは、一つの手引きになると評価できる。けれども同時に、その裏腹の問題として、教室内のダイナミズムや教師のふるまいを十分把握する上で、同モデルには限界があると言わざるを得ない。したがって、「教師の成果」(teacher performance)が本来多面的な性質をもつ以上、とりわけ処遇に関わるハイ・ステイクスな決定においては、例えば同モデルの結果といった単一指標のみに依拠することは難しいと結論された。

#### 4. 結びにかえて

これまでValue-Addedモデルをめぐる2つの誌上討論を中心的に取り上げる中で、各論者の主張を確認してきた。最後に、ここで提示された諸論点をマイクロ・メゾ・マクロという視点から重層的に整理し直すことによって、本稿の目的である現代米国における「教師の効果」測定の基本的論争点の洗い出しを試みたい。

第一に、「教師の効果」測定モデルというマイクロレベルの範疇で米国の動向を再確認すれば、①Value-Addedモデルなど学力成長を評価規準とするモデルが複数存在している点、②その測定結果は妥当性・信頼性が十分でないという指摘されている点が判明した。前者については、Value-Addedモデル・成長度パーセンタイル・Value Tableモデルなどの「教師の効果」測定モデルが存在し、さらにValue-Addedモデルに限ってもその中で数種類のモデルが存在していることがわかった。後者については、「教師の効果」測定に必要なデータ収集の制約やその前提となるテスト問題への不信等から、測定結果の妥当性・信頼性が十分に認められていないという深刻な問題が指摘されていた。

しかしながら第二に、学校現場というメゾレベルから捉えれば、教育実践上の意義も示唆される。それは、Value-Addedモデルは教師の専門性開発および士気向上につながりう

るという主張である。「素点」など従来の測定原理とは異なり、同モデルは「子どもの学力をいかに伸ばすか／伸ばしているか」という発想を制度化したものである。もとより「学力テスト」というごく狭い範囲であるものの、この発想自体は“子どもの成長・発達”という教育の原理に即するものであるため、教育実践の成否を問う参照軸として機能する可能性があり、教師の成長等につながると期待されていた。とはいえ同時に、これらの意義の発現に向けては、主に2つの問題が足かせになっていると指摘されていた点も重要である。ひとつは、研修機会の不足等から、そもそも教師たちが同モデルのデータ活用方法に精通していないという問題。もうひとつは、例えば「人事直結型教員評価」などハイ・ステイクスな文脈で同モデルを使用すると、こうした意義が縮減されうるという問題である。すなわち、Value-Addedモデルの活用にあたっては、まずその目的・目標を明確にし、教職員の理解を得ながら、その使用方法や条件整備を慎重に進めることが不可欠であると言えよう。

第三に、マクロレベルから米国の政策動向を概括すれば、①近年のアカウンタビリティ政策の文脈で「教師」が中心に位置づけられた点、②その「効果」を判断する上で、学力テストが主眼とされている点、③学力成長を規準とする「教師の効果」測定モデルが全米的に普及しつつある点が、本稿の検討を通して明らかになった。この一連の流れから、近年の学力向上政策の影響を強く受けながら、学力テスト結果を通じた「教師の効果」測定が興隆してきたと把握できる。

以上、現代米国の「教師の効果」測定をめぐる基本的論争点を洗い出す中で、Value-Addedモデルは測定結果の妥当性・信頼性という根本的な不安要素を抱えているにもかかわらず、いわゆる“納得解”として社会に受容され、普及するという矛盾が浮かび上がった。ここからさらに踏み込んで総合的に考察すれば、次の理論的・実践的含意を読み取れる。

まず、理論的含意として重要なのは、米国ではValue-Addedモデルの登場によって、教師への“まなざし”が再構築されたと考えられる点である。米国では「教師の効果」指標の典型例が教職経験や学歴等とされてきたように、かつては“優れた教師”の構成要素を実存的に問い続ける“構成主義”が「教師の効果」をめぐる議論の基盤にあった。これに対し、2010年代初頭における「教師の効果」測定モデルの普及状況を鑑みれば、近年では“優れた教師”の絶対条件として「学力を伸ばすことに成功している」ことが暗黙裡に認識されていると言え、この前提を基に“優れた教師”が有する／置かれた諸条件を緻密に検討・析出していく“効果産出の条件精製”へと、その議論の基盤が変容している様相を見て取れる。特筆すべきは、この変容に伴い、学力結果における教師の位置について次

の捉え直しが生じた点である。すなわち、従来では「学力」が児童生徒の学習成果として把握されがちで、学ぶ側にのみ過剰な責任・負担を課す向きがあったのに対し、現代米国の「教師の効果」測定の動向には、「学力」を教授＝学習過程の結果として再認識し、教える側も一定の責任を引き受けようとする新たなスキームの萌芽を感じられる。もっとも、学力テストという限定的範囲でこの動きが活発化している点には十分注意しなければならないが、「子どもの能力や親の意識評定だとされて疑われなかった教育評価のしごとを、逆に、教育条件と教師の教育力の測定と評価のしごとへと180度むきをかえる」(中内1985:169-170)志向性がうかがえる点は、理論的含意として大いに注目できる。

他方、実践的含意としては現代的課題が浮かび上がった。それは、とりわけアカウントビリティへの対応が急務とされる現代社会では、「教師の効果」を可視化・数値化した途端に、その目的の正当性にかかわらず人事管理に利用されてしまうという問題である。Value-Addedモデルの開発者とされるサンダース(Sanders, W.)によれば、同モデルは「アカウントビリティ」に用いるための指標ではなく、あくまで「全ての子どもの学力保障」を目指すものであった(Sanders 2003)。にもかかわらず、学力向上が至上命題、より踏み込んで述べれば、その成果を“分かりやすい形”で提示することが求められる昨今においては、必然的に学力テスト結果への注目が集まり、それを通じて算定された「教師の効果」がアカウントビリティや「人事直結型教員評価」と結びつきやすい。例えば、Value-Addedモデルをめぐる議論も、「測り方」といった測定原理固有の特性・問題を検証するものから、「教員評価における学力テスト結果の位置付け」、さらには解雇を含む「人事直結型教員評価」に関する議論へと、その中心的議題が拡張／分散したことは既述の通りである。

本稿の検討から、本来Value-Addedモデルは子どもの学力保障に向けた教師の成長や条件整備の目安として活用することを目指して制度化されたものであるにもかかわらず、学力向上の成否が厳しく問われる現代社会では、アカウントビリティや人事管理に利用する力学が働き、教職の矮小化や教師の身分保障といった深刻な問題にまで発展する恐れが示唆された。こうした問題はわが国も他人事ではない。全国学力学習状況調査のみならず、GIGAスクール構想によってビッグデータの集積が加速化する今後を見通せば、そこで可視化されるデータの限界および危険性を念頭におきつつ、子どもの学習権保障に向けてどう活用できるか／すべきでないかを慎重に議論する必要があるだろう。

## 【付記】

本稿はJSPS科研費JP15J01975, JP19K14078の研究成果の一部である。

## 【注】

- 1) 1980年代以降の第三次産業革命に伴う統計技術の発展, そして2010年代以降の第四次産業革命におけるビッグデータの集積・活用に伴い(佐藤 2021), 近年ではテスト点数のうち教師が貢献した部分のみを析出することが技術的に可能になったという前提で教育政策が進みつつある。英国・中国といった先進工業諸国においても Value-Addedモデルの運用が始まっている(中国については高妻ほか 2018, 英国については志水 2014を参照)。
- 2) 米国の教員身分保障制度については, 入江 2005や高橋 2011が詳しい。
- 3) ハヌシュエック氏へのヒアリング調査より。なお, 調査概要は次の通り。

ヒアリング対象者	調査日程	調査場所(所要時間)
スタンフォード大学・教授 ハヌシュエック(Hanushek, E.)氏	2012年11月5日	同氏の研究室(1時間)

## 【文献一覧】

- 石井英真(2015)『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(増補版)』東信堂。
- 入江彰(2005)『教員の処分と手続制度－アメリカ合衆国の教員解雇における手続保障の法制度に関する研究』多賀出版。
- 大桃敏行・吉良直・堀ひかり・宮口誠矢・金子友紀(2016)「公立学校の多様化とアカウントビリティ政策の展開－ワシントンD.C.を事例として－」『東京大学大学院教育学研究科紀要』第55巻, 425-443頁。
- 北野秋男・吉良直・大桃敏行編著(2012)『アメリカ教育改革の最前線－頂点への競争－』学術出版会。
- 高妻紳二郎・榊原禎宏・末松裕基・張揚・辻野けんま・照屋翔大・藤村祐子(2018)『中・独・英・米4カ国における教育経営学研究の国際的潮流』日本教育経営学会国際交流委員会。
- 小島倫世(2014)「現代米国における学校の「成果」測定をめぐる議論の動向－Value-Addedモデルの理論的位置」関西教育行政学会『教育行財政研究』第41号, 27-32頁。
- 佐藤学(2021)『第四次産業革命と教育の未来－ポストコロナ時代のICT教育－』岩波書店。
- 志水宏吉(2014)「『学力テスト結果』の社会的背景」『教育』10月号。
- 高橋哲(2011)『現代米国の教員団体と教育労働法制改革－公立学校教員の労働基本法



と専門職性をめぐる相剋』風間書房。

- 高橋哲(2021)「米国のハイスティクスな学力テスト活用モデル」佐藤仁・北野秋男編著『世界のテスト・ガバナンスー日本の学力テストの行く末を探るー』東信堂, 59-79頁。
- 照屋翔大・藤村祐子(2016)「アメリカの教員評価をめぐる付加価値評価モデル(Value-Added Model)の動向」『日本教育経営学会紀要』第58号, 118-130頁。
- 中内敏夫(1985)『指導過程と学習形態の理論』明治図書。
- 西野倫世(2016)「現代米国の学校改善事業にみる学力測定結果の活用状況と課題ーテネシー州チャタヌーガ市のValue-Added Assessmentをめぐる動向ー」『日本教育行政学会年報』No.42, 130-146頁。
- Baker, B., Oluwole, J. & Green, P. (2013) “The Legal Consequences of Mandating High Stakes Decisions Based on Low Quality Information: Teacher Evaluation in the Race-to-the-Top Era” *Education Policy Analysis Archives*, 21 (5) .
- Betebenner, D. W. (2008) “Toward a normative understanding of student growth” Ryan, K. E., & Shepard, L. A. (Eds.) *The Future of Test Based Accountability*, New York: Routledge, pp.155-170.
- Collins, C. & Amrein-Beardsley, A. (2014) “Putting Growth and Value-Added Models on the Map: A National Overview” *Teachers College Record*, 116 (1) .
- Darling-Hammond, L. et al. (2012) “Evaluating teacher evaluation” *Phi Delta Kappan*, 94 (3) , pp.8-15.
- Goldhaber, D., Walch, J. & Gabele, B. (2014) “Does the Model Matter? Exploring the Relationship Between Different Student Achievement-Based Teacher Assessments” *Statistics and Public Policy*, 1: 1, pp.28-39.
- Hanushek, E. & Rivkin, S. (2012) “The Distribution of Teacher Quality and Implications for Policy” *Annual Review of Economics*, 4, pp.131-157.
- Henry, G., Rose, R., & Lauen, D. (2014) “Are value-added models good enough for teacher evaluations? Assessing commonly used models with simulated and actual data” *Investigaciones de Economía de la Educación*, 9, pp.383-405.
- Kinnucan-Welsh, K. et al. (2012) “Should value-added modeling be used to identify highly effective teachers?” Lasley, T. J. et al., *Standards and Accountability in Schools*, SAGE Publications.
- Konstantopoulos, S. (2014) “Teacher Effects, Value-Added Models, and

- Accountability” *Teachers College Record*, 116 (1) .
- Lavigne, A. & Good, T. (2013) *Teacher and Student Evaluation: Moving Beyond the Failure of School Reform*, Routledge.
- Lavigne, A., Good, T. & Marx, R. (2014) “Introduction to High-Stakes Teacher Evaluation” *Teachers College Record*, 116 (1) .
- Lavigne, A. (2014) “Exploring the Intended and Unintended Consequences of High-Stakes Teacher Evaluation on Schools, Teachers, and Students” *Teachers College Record*, 116 (1) .
- Mahoney, J. (2006) “How Value-Added Assessments Helps Improve Schools” *Phi Delta Kappa Edge*.
- Mahoney, J. (2012) “Should value-added modeling be used to identify highly effective teachers?” Lasley, T. J. et al., *Standards and Accountability in Schools*, SAGE Publications.
- National Research Council and National Academy of Education (2010) *Getting Value Out of Value-Added: Report of a Workshop*, The National Academies Press.
- Sanders, W. L. (2003) “Beyond *No Child Left Behind*” paper presented at the 2003 Annual Meeting American Educational Research Association.
- Walsh, E. & Isenberg, E. (2013) *How does a value-added model compare to the Colorado Growth Model*, Mathematica Policy Research, Working Paper.
- Winters, M. A. & Cowen, J. M. (2013) “Who Would Stay, Who Would Be Dismissed? An Empirical Consideration of Value-Added Teacher Retention Policies” *Educational Researcher*, 42 (6) , pp.330-337.