

ジョン・ハッティの研究成果と教育実践との関係を問う

——「わかりやすいエビデンス」の陥穽——

石田 智敬* 森本 和寿*

* Co-first authorship

1. はじめに

「どのような教育が効果的か」という問いは、古くて新しい問いである。実践家である教師にとって、どのようなカリキュラムや教育方法、指導方略を選択するかは、日々直面する切実な問いであり続けている。また、近年では、学校教育に対するアカウンタビリティ（説明責任）の要請が世界的に高まっており、教育を取り巻く判断とその説明において、根拠の明示を求める傾向も日増しに強まっている。客観的根拠に基づいた合理的な判断を促進することは、教育における質の向上に寄与すると考えられている。このような背景のもと、学校や教師、教育行政関係者が選択する／選択したカリキュラムや教育方法が効果的であるか否かを判断し、その判断について説明するために「エビデンス」を求める声が高まっている。

「どのような教育が効果的か」、「その根拠は何か」という問いに対して、私たちはどのように向き合うべきなのだろうか。2018年、邦訳『教育の効果』が出版されて以来、日本の教育界においても広く認知され始めているジョン・ハッティ（John Hattie）は、この問いに対して一つの明確な立場を提示している。ハッティによる「教育効果の可視化（visible learning）」研究（以下、可視化研究と呼ぶ）は、学力（educational achievement）に影響を与えるさまざまな要因（教育方法やカリキュラム、その他の教育的介入はもちろん、教育を取り巻く環境要因や学習者の発達要因をも含む）に対する系統的レビュー（メタ分析: meta-analysis）を大量に収集し、各メタ分析で示された効果量（effect size: ES）を、Cohen's *d* を指標として統合（super synthesis）するというものである。これは、要因や手法を定量的に比較可能にすることで教育の効果を可視化し、「最も効果的な教育は何か」を見出そうとする野

心的な試みである¹。

ハッティによる可視化研究が教育界に与えた影響は大きい。ハッティが研究成果を書籍として出版した Hattie (2009) と Hattie (2012) は、それぞれ 17967 件、4789 件の被引用数を誇っており、世界的に大きな注目を集めている²。このようなハッティの研究成果について、ドイツの教育学者エヴァルト・テルハルト（Ewald Terhart）は、「[教育に関わる] 研究における画期的な出来事であり、学校での学習が成功するための条件についての議論」であると述べ、そのインパクトの大きさを強調した（Terhart, 2011, p. 425）。

ハッティの研究業績に対するインパクトの大きさは、ハッティがそれによって得た地位からも垣間見ることができる。たとえばハッティは、その研究業績から、現在、豪州メルボルン大学の桂冠教授（laureate professor）の地位が与えられている。桂冠教授は、ノーベル賞といった世界的な学術賞の受賞者に贈られる教授職である。マスメディア等においても、ハッティの研究成果は、「聖杯」や「真の宝」、あるいは『指輪物語』に登場する「万物を支配する指輪」等のメタファーを用いて形容されている（Terhart, 2011）。たとえば、英国のタイムズ紙教育版は、「指導の聖杯、研究によってついに明らかに（“Research reveals teaching's Holy Grail”）」と報じたという。なかでも、各要因の順位表である「ハッティ・ランキング（Hattie ranking）」は、そのわかりやすさから、特に注目を集めている。

教育実践の場面や政策決定の場面においても、教育効果の可視化を謳うハッティの研究成果は、「わかりやすいエビデンス」として小さくない影響力をもち始めている。ただし、このようなハッティの研究成果が、筋違いな解釈や見当違いな判断に基づいて理解されたり、引用・利用されていたりする側面も見落とす

ことはできない³。

ハッティによる可視化研究については、これまでもいくつかの先行研究において、その意義と課題が検討されてきた (Snook, 2009; Snook, 2010; Myburgh, 2016; 熊井, 2019)。ただし、(1) ハッティが行った統合メタ分析の内実 (産出過程) を詳らかにし、(2) ハッティの研究と教育実践との関係性を問う試みが十分なされていないとは言えない。そこで本稿は、ジョン・ハッティによる可視化研究を一次研究にまで遡って批判的に検討することで、可視化研究の可能性と限界を浮き彫りにし、教育研究者や教育実践者がハッティの可視化研究とどのように向き合うべきかについて考察することを目的とする。この試みを通じて、教育におけるエビデンスのあり方をも検討したい。

2. ハッティは何をしたか：Super Synthesis の衝撃

2.1 ハッティの研究手法と研究成果の表現方法

ハッティによる可視化研究における研究手法はどのようなものか。簡潔に言えば、彼の研究手法は、教育効果の可視化を目指して、(1) 教育研究における学力 (学習成果) に関する効果を扱ったメタ分析論文 (研究) を徹底的に収集し、(2) それらを、効果量を指標として、ハッティ自身が設定した「要因項目」ごとに区分し、統合 (super synthesis) するというものである。こうしたハッティの試みは、134 のメタ分析研究を収集した Hattie (1992) の取り組みに端を発するものであり、1992 年以降も引き続いて、彼はメタ分析研究を収集し続けている。

ここでいう「要因項目」 (たとえば、「フィードバック」「宿題」「学級規模」等) は、ハッティによって、独自に設定されているものである。統合素材としての各メタ分析の指標は、各要因の項目名に従って割り振られて区分される。また、これらの要因項目は、「学習者」「家庭」「学校」「教師」「カリキュラム」「指導方法」といった大領域ごとで整理されるようになっている。このようなハッティによる可視化研究の取り組みは、教育研究におけるメタ分析を含む研究論文を徹底的に収集することで、量的エビデンスに関する巨大データベースを構築しようとするものとも言える。

メタ分析とは、「過去の様々な研究成果を検索・選択し、組み合わせるために正式な統計手順を使用する

研究の統合」のことであり (Wachter, 1988, 1407)、系統的レビュー (systematic review) とも呼ばれる。メタ分析は、同一のテーマについて行われた複数の研究結果を、統計的方法を用いて統合するために、信頼性が高いものとされる。ただし、メタ分析では、研究で報告された統計情報 (分析結果) を利用して研究の統合を行うため、その対象は実証的研究に限定され、データを用いない理論研究や質的な事例研究はメタ分析の対象からは除外されるという課題もある。

このようなメタ分析を収集し、さらに統合しようとするハッティの研究手法 (統合メタ分析) は、「メタ・メタ分析」ないし「メガ分析 (mega-analysis)」とも呼ばれる場合がある。統合メタ分析に代表される量的研究の世界的な流行は、それらが、エビデンス・ピラミッドの頂点にあるメタ分析をさらに統合するものであるということによって勢いづけられており、日本においても注目を集めている (亘理, 2014)。

2009 年時点において、ハッティによる可視化研究では、52637 件の個別研究を参入した 800 以上のメタ分析結果が収集され、それらが 138 の教育に関わる要因に区分され、分析されている。ハッティ・ランキングは数年ごとに更新されており、最新版のハッティ・ランキング (2018 年) では 80000 件の個別研究 (対象児童・生徒 3 億人) を参入した 1400 のメタ分析が、252 の要因項目で整理・統合され、ランキング化されている。ここで留意しておくべきことは、彼が量的研究やメタ分析を実際に行なっているわけではなく、(1) 教育研究のメタ分析論文を収集すること、(2) さまざまなカテゴリーを設定して、各メタ分析を整理しているということにある。この点が、彼の可視化研究における一番のオリジナリティであると言える。

可視化研究における成果の表現方法はどのようなものか。ハッティは、教育効果を可視化することを主張しており、「効果がある」と謳われてきた教育的介入 (教育方法、カリキュラム、教育政策等) は本当に効果があるのかを、効果量を用いることで徹底的に可視化しようと試みた。可視化するという言葉が如実に示すように、ハッティは研究成果として得られた統合効果量を、バロメータやランキングといった形式で表現することで、教育効果の大きさをわかりやすく示している。

また、ハッティは、収集した論文の効果量のほとんどが正であること、そして、収集した全てのメタ分析研究の平均効果量が 0.4 であることを論拠に、目標値をゼロに設定するのは不条理であると主張し、効果量 0.4 を一定の基準値に据え、0.4 以上を望ましい教育効果をもつものとすることを提案している。なお、ハッティは、諸関連データを踏まえつつ、 $d=0\sim0.15$ を「発達の効果」（学校に通わずとも学習者が到達しうる範囲）、 $d=0.15\sim0.4$ を「教師の効果」（学校に通って教師の指導を受けることによる一般的な効果範囲）と位置づけている。そのうえで、以下のようなバロメータをデザインし、各要因項目の効果量をわかりやすく表現している（図 1）。

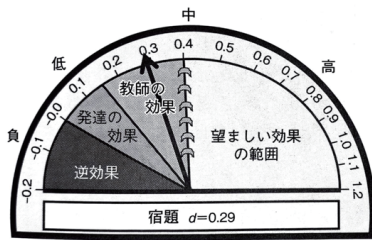


図 1. 統合効果量の大きさを示すバロメータ

出典：ハッティ・J. 著、山森光陽監訳（2018）『教育の効果：メタ分析による学力に影響を与える要因の効果の可視化』図書文化、p.53。

2.2 ハッティの研究に対する反応

「はじめに」で述べたとおり、このようなハッティの可視化研究は、世界的に大きな称賛と注目を得てきた。ただし、可視化研究を公開して以来、ハッティの研究成果は多くの賛同を得るとともに、厳しい批判にもさらされてきた。特に、彼の統合メタ分析という手法、そしてその素材となるメタ分析に対する批判は少なくない⁴。ハッティの研究に対して、研究手法上の批判をいち早く加えたのが、ニュージーランドの研究者であるスヌーク（Ivan Snook）らである。スヌークらは、ハッティが *Visible Learning* を刊行した翌年（2009 年）に、統合メタ分析において用いた 800 の研究の質に対してハッティが関心を払わなかった点を批判する論文を公開した。この論文でスヌークらは、Evidence-informed Policy Network⁵における「研究の統合」の定義を援用したうえで、ハッティが用いた「研究の統合」という手法を適切に用いるためには、「プロト

コルの開発、体系的で明示的な方法の使用、データ収集、分析、結果の解釈と報告」が必要であり、「不十分または不適切な研究を除外しないメタ分析は誤解を招く可能性があり、それが不適切な方針の策定につながる場合は潜在的に有害である」ことを指摘した（Snook et al., 2009, p. 94）。

翌年、ハッティは自らの可視化研究に寄せられた批判に応答する論文を出し、この中でスヌークらの論考を引用して反論している。ハッティは、研究の統合に際して用いられている「研究成果は査読を通じて最大限に質の上でのコントロールを受けているため、公表された研究成果には大きな信頼性がある」と主張した⁶（Hattie, 2010, p. 88）。このハッティの反論に対してスヌークらは、ハッティの主張はメタ分析一般に対する批判として存在する公表バイアスへの反論として成立していない点を指摘し、ハッティの研究の意義を認めたいうえで、参照に際しては留意が必要であるとしている⁷（Snook et al., 2010）。

このようなスヌークらによる厳しい指摘の背景には、メタ分析一般に対する批判である「ゴミを入れてもゴミしか出てこない（garbage in, garbage out）」問題がある。先にも示したとおり、ハッティの研究は膨大な量的研究に基づいている（2009 年の研究段階で、個別研究 52637 件を参入した 800 以上のメタ分析結果に基づく 138 の教育に関わる要因を分析している）。しかしこれは、5 万件以上ある個別研究ないし 800 以上あるメタ分析結果をハッティの研究内で一つ一つ精査したことを意味するものではなく、ハッティの研究において個別研究は直接的に参照されない。また、800 以上のメタ分析結果についても効果量が参照されるものの、個々のメタ分析研究すべてに対して個別に言及されることはない。したがって、大量のメタ研究を機械的に関数に入れて統合し、そこから導き出された平準化された効果量にのみ関心を向けるのが、ハッティの統合メタ分析研究において行われていることである。

このような処理手順は大量の情報を処理することに長けているという面がある一方で、大量に投入したメタ分析結果のなかに精度の低いものが入り込んでしまい、結果を歪めてしまうという危うさがある。「メタ分析は、質の悪い研究が存在するのと同じ理由で、質の低い研究を含めることに対して免疫があるわけでは

ない」ため、質の低い研究が結果に深刻な影響を与える可能性には十分に留意する必要がある（Myburgh, 2016, p. 7）。メタ分析が抱えるこの問題に対抗するひとつの手段として、スヌークらは、「大規模」で「よく設計されていて」、「数年にわたって実施された」研究を参照することを推奨するというように、メタ分析の入力値に対してバリデーションを設けることを提起している（Snook et al., 2010, p. 96）。

以上はハッティが用いた研究手法に対してメタ分析一般がもつ問題という観点から批判したものであるが、ハッティの研究の内実を質したものではない。では、ハッティの統合メタ分析において参照されているメタ分析、あるいはそのメタ分析で参照されている一次研究は、どのようなものであり、どのような手続きで参照されているのか。次章では、彼の研究における「入力値」、すなわち元になったメタ分析や一次研究について精査する。

3. ハッティによる教育の可視化研究の内実を探る：浮き彫りになる新たな疑念

ハッティは、教育研究におけるメタ分析を含む膨大な論文を探し出し、教育の要因項目として任意のラベルを設定し、その要因項目のラベルに基づいてメタ分析を分類し、各要因に対する統合効果量を算出している。なお、ある要因項目の統合効果量は、ある要因項目として集められたメタ分析研究における効果量を平均化したものである⁸。そして、次のように、それをランキング化している（表 1）。ただし、見てわかるとおり、要因項目のラベルには抽象的なものが多く、それぞれの要因項目が一体どのような教育実践や手法・概念を具体的に指すものなのかは、不明瞭である。

そもそも教育における用語や概念は、ある用語や概念が意味することが明確に一意に定まるように定義される科学的用語とは異なり、一意に定まる明確な定義をもたない多義的なものであることが、ほとんどである。ただし、ここでの要因項目のラベルの背後には、統合効果量を算出する源となった一次研究が存在する。つまり、ハッティがつけた要因項目におけるラベルが意味することは、そのラベルの下に収集された一次研究の内実によって帰納的に説明されることになる。

そこで、本章では、統合効果量算出の源流である一

次研究にまで遡ることで、各ラベルの下に収集された一次研究の内実を切り開き、その要因項目が何を意味するものなのか、そして、どのような算出過程を経て統合効果量が弾き出されているのかについて詳細に見ていこう。言うなれば、これはハッティのブラックボックスを切り開く試みである。この試みを通して、ハッティが提示する統合メタ分析という研究手法・方法がもつ可能性と限界についても考察していきたい。

表 1. 『教育の効果』におけるランキング表

順位	領域	要因	d
1	学習者	能力レベルの自己評価*	1.42
2	学習者	ピアジェによる発達段階*	1.28
3	指導方法	指導改善のための形成的評価	0.90
4	教師	マイクロティーチング*	0.88
5	学校	飛び級*	0.88
6	学校	学級での児童生徒の行動	0.80
7	指導方法	学習障害児に対する理解力向上のための介入	0.77
8	教師	教師の明瞭さ	0.75
9	指導方法	相互教授法	0.74
10	指導方法	フィードバック	0.73
11	教師	教師と学習者の関係	0.72
12	指導方法	分散学習と集中学習	0.71
13	指導方法	メタ認知的方略	0.69
14	学習者	過去の学力	0.67
15	カリキュラム	語彙獲得プログラム*	0.55
16	カリキュラム	繰り返し読み指導*	0.67
17	カリキュラム	創造性育成プログラム*	0.65
18	指導方法	自己言語化と自己質問	0.64
19	教師	現職教育	0.62
20	指導方法	問題解決的指導	0.61
21	教師	学習者を分類すること	0.61
22	カリキュラム	音声指導*	0.60
23	指導方法	教授方略	0.60
24	指導方法	協同的な学習と個別的な学習	0.59
25	指導方法	スタディスキル	0.59
26	指導方法	直接教授法	0.59
27	カリキュラム	幼児に対して触覚刺激を与えること*	0.58
28	カリキュラム	読解プログラム*	0.58
29	指導方法	完全習得学習	0.58

出典：ハッティ・J.著、山森光陽監訳（2018）『教育の効果：メタ分析による学力に影響を与える要因の効果の可視化』図書文化、p.330。

もちろん、ここですべての要因項目の一次研究を参照することは困難を極めるので、ここでは、「カリキュラムや指導方法・内容改善のための形成的評価（教師へのフィードバック）」を一例として取り上げることとする。この「形成的評価」の要因項目における効果量は、2009 年（訳本は、2018 年）の段階で、「指導方法」の領域で第 1 位、全領域でも第 3 位にある。これはすなわち、彼の研究知見において、「形成的評価」に取り組むことが、指導方法としては最も効果があると示されていることを意味しており、ハッティの研究成果と教育実践の関係を問うのに適した項目であると言える。

このような彼の研究成果を「量的」「実証的」根拠として引用する形成的評価研究は少なくない。では、ここでいう「形成的評価」が意味する、指導手法（教育実践）の内実はどうなのなのか。どのような学習者を対象にするもので、具体的にどのような方法によるものか。そして、どのような学力を対象にするものか。以下、本稿では、このような形成的評価を取り巻く言説を基礎づけるハッティの研究の、そのさらに基礎となる部分を検証する。

3.1 可視化研究における統合効果量の源流を辿る

ハッティは、「形成的評価」というラベルが与えられた要因項目に対して効果量 $d = 0.9$ という驚くべき「エビデンス」を提供している。この数値は、効果量 $d = 0.7$ と効果量 $d = 1.1$ を弾き出す Fuchs & Fuchs (1986) と Burns & Symington (2002) の2つのメタ分析を統合した（平均化した）ことによって算出されたものである。これら2つのメタ分析は、それぞれ21本と9本、合計して30本の一次研究に基づくものである（図2）。それでは、これら2本の研究論文と、それらの一次研究は、どのようなものなのか。



図2. 要因項目「形成的評価」の構造（著者作成）

3.1.1 Fuchs & Fuchs (1986) の概要

まず、Fuchs & Fuchs (1986) の内容について簡潔に見ていこう。フックスらは、特別支援教育において、個別化された指導が児童・生徒の学力向上に寄与するという原則を確認した上で、その一手法である「システムティックな形成的評価（systematic formative evaluation）」に焦点を合わせる。ここでいう形成的評価とは、提案されたプログラムの継続的な評価と修正に焦点を合わせるものであり、具体的には、さまざまな指導方法の下における障害のある生徒の学習成果を定期的にモニタリングするものであった。モニタリングの目的は、個に合わせたプログラムを作成する為のデータベースを提供することであった。

Fuchs & Fuchs (1986) は、このような形成的評価に関連する定量的研究のメタ分析を試みることによって、形成的評価の手続きが学習成果に及ぼす影響を調査した。メタ分析の結果、21件の一次研究に対する加重平均効果量（ES）は $d=0.7$ であった。この結果を踏まえて、フックスらは、「本研究の結果から、特別支援教育においてシステムティックな形成的評価手順を用いることで、学力向上が確実に期待できること、また、教師が行動修正やデータ評価ルール、グラフ化されたデータ表示を用いることで、その効果がより高まる可能性があることが明らかになった」と結論づけた。

Fuchs & Fuchs (1986) における対象研究（一次研究）は、博士論文4本、未発表研究11本、雑誌論文8本である（表2）。合計3,835人の被験者がこれらの研究に参加しており、83%は障害児を対象としていた。また、対象学年は就学前から高校までの範囲であった。

表2. Fuchs & Fuchs (1986) の一次研究

1. Beck, R (1976). <i>Report for the office of education dissemination review panel</i> . (Unpublished manuscript)
2. Beck, R (1979). <i>Report for the office of education dissemination review panel</i> . (Unpublished manuscript)
3. Beck, R (1981). <i>Curriculum management through a data base</i> . (Unpublished manuscript)
4. Beck, R (1981). <i>High school basic skills improvement project</i> . (Unpublished manuscript)
5. Bohannon, R. M. (1975). <i>Direct and daily measurement procedures in the identification and treatment of reading behaviors in children in special education</i> . (Unpublished doctoral dissertation, University of Washington.)
6. Bradfield, R. H., Brown, J., Kaplan, P., Rickert, E., & Stannard, R (1973). The special children in the regular classroom. <i>Exceptional Children</i> , 39, 384-391.
7. Brandtetter, G., & Merz, C. (1978). Charting scores in precision teaching for skill acquisition. <i>Exceptional Children</i> , 45, 42-48.
8. Bruening, S. E. (1978). Precision teaching in the high school classroom: A necessary step towards maximizing teaching effectiveness and student performance. <i>American Educational Research Journal</i> , 15, 125-140.
9. Crutcher, C. E., & Hofmeister, A. M. (1975). Effective use of objectives and monitoring. <i>Teaching Exceptional Children</i> , 7(3), 78-79.
10. Dubrule, M. N. (1984). <i>The study of precision teaching as a remedial method</i> . (Unpublished doctoral dissertation, Clark University.)
11. Frumess, S. C. (1983). <i>A comparison of management groups involving the use of the standard behavior chart and setting performance aims</i> . (Unpublished doctoral dissertation, University of Houston.)
12. Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. <i>American Educational Research Journal</i> , 21, 449-460.
13. Haring, N. G. (1971). Investigation of systematic instructional procedures to facilitate academic achievement in mentally retarded disadvantaged children. Final report. (ERIC Document Reproduction Service No. ED 071 246)
14. Haring, N. G., & Krug, D. A. (1975a). Evaluation of a program of systematic instructional procedures 208 for extremely poor retarded children. <i>American Journal of Mental Deficiency</i> , 79, 627-631.
15. Haring, N. G., & Krug, D. A. (1975b). Placement in regular programs: Procedures and results. <i>Exceptional Children</i> , 41, 413-417.
16. King, R. Deno, S. L., Mirkin, P. K., & Wesson, C. (1983). <i>The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison (Research Report No. 111)</i> . Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.
17. Lovitt, T. C., & Fantasia, K. (1983). A precision teaching project with learning disabled children. <i>Journal of Precision Teaching</i> , 3, 85-91.
18. Mirkin, P. K. (1978). <i>A comparison of the effects of three formative evaluation strategies and contingent consequences on reading performance</i> . (Unpublished doctoral dissertation, University of MN.)
19. Mirkin, P. K., Deno, S. L., Tindal, G., & Kuehne, K. (1980). <i>Formative evaluation: Continued development of data utilization systems (Research Report No. 23)</i> . University of

Minnesota, Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 197 510)

20. Peniston, E. G. (1975). *An evaluation of the Portage Project: A comparison of a home-visit program for multiple handicapped preschoolers and Headstart program*. (ERIC Document Reproduction Service No. ED 112 570)

21. Sevcik, B., Skiba, R., Tindal, G., King, R., Wesson, C., Mirkin, P. K., & Deno, S. L. (1983). *Curriculum based measurement: Effects on instruction, teaching estimates of student progress, and student performance* (Research Report No. 124). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

22. Skiba, R., Wesson, C., & Deno, S. L. (1982). *The effects of training teachers in the use of formative evaluation in reading: An experimental-control comparison* (Research Report No. 88). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.

23. Tindal, G., Fuchs, L. S., Christenson, S., Mirkin, P. K., & Deno, S. L. (1981). *The relationship between achievement and teacher assessment of short- or long-term goals* (Research Report No. 61). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 218 846)

* なお、13, 14, 15, は同一研究のため、21 の研究に対して 23 の文献が示されている。

3.1.2 Burns & Symington (2002) の概要

次に、Burns & Symington (2002) の内容について簡潔に見ていこう。バーンズらは、児童・生徒の問題行動や学習困難に対応することを目的とした「照会前介入チーム (Prereferral Intervention Team: PIT)」に焦点を合わせる。PIT モデルは、不必要な特別支援教育への紹介や配置を減らし、レッテル貼りや一般教育からの分離を減らすのに有効であるとされる。PIT モデルは、一般的に(1) 相談の依頼、(2) 相談、(3) 観察、(4) カンファレンス、必要に応じて、(5) 特別教育の適格性のための正式な紹介を含む。バーンズらは、PIT モデルは公立学校で一般的であるものの、その有効性に関する研究は十分行われていなかった点を指摘する。そこで、試験的なメタ分析を行うことで、PIT モデルの有効性を検証した。ただ、バーンズらは対象研究数が少ないため、メタ分析から得られる結論の信頼性には限界があるとし、PIT モデルの有効性を実証的に検証するというよりも、今後の研究が必要な領域を特定することに主眼を置いた。

Burns & Symington (2002) では、72 本の論文のレビューが行われ、その内 9 本がメタ分析の採用基準に合致した。メタ分析の結果は、平均効果量 (ES) 1.1 であった。バーンズらは、最も重要な発見は、大学基盤による PIT の効果量 (1.32) とフィールド基盤の PIT の効果量 (0.54) の差だと指摘した。大学基盤のチームでは PIT に対する肯定的な結果が多く見られ、フィールド基盤からは一貫性のない結果が報告された。以上から、肯定的な結果の欠如は、PIT の一貫した実施の欠如と関連している可能性がある結論づけた。

Burns & Symington (2002) における対象研究 (一次

研究) は、次の 9 つの論文であった (表 3)。すべての研究が、小学校または中学校 (1 学年から 7 学年) の学習者を対象としており、通常学級に参加している問題行動や学習困難を有する学習者 (ただし、障害を持つとは認定されていない) を対象とするものだった。

表 3. Burns & Symington (2002) の一次研究

1. Bahr, M., Fuchs, D., Fuchs, L. S., Fernstrom, P., & Stecker, P. (1993). Effectiveness of student versus teacher monitoring during prereferral intervention. *Exceptionality*, 4, 17 – 30.
2. Burns, M. K. (1999). Effectiveness of special education personnel in the intervention assistance team model. *The Journal of Educational Research*, 92, 354 – 356.
3. Fuchs, D., & Fuchs, L. S. (1989). Exploring effective and efficient prereferral interventions: a component analysis of behavioral consultation. *School Psychology Review*, 18, 260 – 283.
4. Fuchs, D., Fuchs, L. S., & Bahr, M. W. (1990). Mainstream assistance teams: a scientific basis for the art of consultation. *Exceptional Children*, 57, 128 – 139.
5. Graden, J. L., Casey, A., & Bonstrom, O. (1983). *Pre-referral interventions: effects on referral rates and teacher attitudes*. Minneapolis, MN: Institute for Research on Learning Disabilities, University of Minnesota (ERIC Document Reproduction Services No. ED 244438).
6. Graden, J. L., Casey, A., & Bonstrom, O. (1985). Implementing a prereferral intervention system: II. The data. *Exceptional Children*, 51, 487 – 496.
7. Kovaleski, J. F., Gickling, E. E., Morrow, H., & Swank, P. (1999). High versus low implementation of instructional support teams: a case for maintaining program fidelity. *Remedial Special Education*, 20, 170 – 183.
8. McKay, B., & Sullivan, J. (1990). *Effective collaboration: The Student Assistance Team model*. Paper presented at the Annual Convention of the Council for Exceptional Children, Toronto, Canada (ERIC Document Reproduction Service No. ED 322695).
9. Ponti, C. R., Zins, J., & Graden, J. L. (1988). Implementing a consultation-based service delivery system to decrease referrals for special education: a case study of organizational considerations. *School Psychology Review*, 17, 89 – 100.

3.2 浮き彫りになった可視化研究がもつ問題

以上のように、ハッティによる可視化研究のブラックボックスを切り開いていくと、それが抱えるいくつかの問題点が浮き彫りになってくる⁹⁾。ハッティによる可視化研究が抱える諸問題点は、大まかに次の二つのレベル (段階) で捉えることができる。第一に、ハッティによる可視化研究そのものがもつ問題点であり、第二に、可視化研究で用いられている「統合メタ分析」という研究方法に内在する問題点である。以下に、これら問題点について詳しく見ていこう。

3.2.1 ハッティによる可視化研究が抱える問題

第一に、上記で見てきたメタ分析研究は、そもそも「形成的評価」の要因項目として区分して扱うことが妥当なのかという「カテゴライズの妥当性問題」が指摘できる。ハッティは、上述した二つのメタ分析研究を「形成的評価」という要因項目ラベルのもとに統合しているが、これは、次の理由から妥当ではない。

まず、Burns & Symington (2002) の論文に関しては、第一に、これは、「照会前介入チーム (Prereferral

Intervention Team: PIT)」の効果に焦点を合わせるものであり、形成的評価の一手法として位置づけることは難しい。実際、Burns & Symington (2002) の論文では、形成的 (formative) という単語は一度も登場していなかった。第二に、Burns & Symington (2002) で弾き出された 1.1 という効果量の数値は、先述のとおり、サンプルの少なさから有効性を示すものとは位置づけられていなかったということがあった。以上より、Burns & Symington (2002) で示された効果量 1.1 という数値を「形成的評価」の項目に組み入れるのは適切ではないと言えよう。

次に、Fuchs & Fuchs (1986) の論文についてである。これは、Burns & Symington (2002) とは異なり、確かに、形成的評価という手法に焦点を合わせるものではあった。ただし、Fuchs & Fuchs (1986) の論文は、特別支援教育における指導方略の有効性を探ることを目的としてなされたものであった。そして、その一次研究にまで遡ると、Precision Teaching という手法の研究が含まれているというように、特別支援教育における独特な形成的評価実践の内容が含まれていた。よって、Fuchs & Fuchs (1986) で示された効果量 0.7 という数値を「形成的評価」の項目に組み入れるのは、適切でないとは言えないものの、この研究のみでは、教育全般における形成的評価の有効性を語る上で、十分な代表性を持ち得ていないとは言えよう。

以上より、形成的評価と言われて一般に想定されるものと、ハッティが「形成的評価」とラベルを貼ったものとの間には乖離があるのではないかということが推察される。要は、このような教育実践の中身を見て、ハッティ以外の教育研究者は、果たしてこの要因項目に「形成的評価」というラベルをつけるだろうかということである。そのため、これらを「形成的評価」という項目名で一括りにするのは、無理があると言わざるを得ない。

効果量 0.9 という数値は、一体どのような教育実践や要因に対して与えられるものであり、それらは一体何を意味するものであるのか極めて不明瞭である。少なくとも、効果量 0.9 という数値を、教育全般において形成的評価が有効であることを示すエビデンスとして用いることは、不適切であると言えよう。

第二に、対象としている学習者、学力、教育環境に

極めて偏りがあるという「サンプリングの偏りの問題」が指摘できる。「形成的評価」という項目において統合メタ分析の素材になっていた二つのメタ分析研究では、その対象の大部分が特別な支援を必要とする学習者であった。もちろん、一次研究まで遡って確認してみても、その対象となっている学習者群や、そこでの教育環境のほとんどが特別支援教育をコンテキストとするものである。少なくとも、このような文脈に限定されて得られた効果量の数値を、通常の学校におけるエビデンスとして用いることの妥当性は低いと言えるだろう。もっとも「形成的評価」にカテゴライズされた二つのメタ分析研究の著者らは、各論で算出された効果量の数値が、通常学校における形成的評価のエビデンスとして用いられることは、想定していなかったことだろう。

第三に、効果量の対象とする学力 (アウトカム) がいったい何を指しているのかが極めて不明瞭であるという曖昧性の問題が指摘できる。一次研究を参照していくと、そこにおいてアウトカムとして定義されている学力概念は、特別支援教育に固有なものであったり、ある特定の行動の確認であったり、スペルや計算といった認知的に単純で限定的な能力であったりするものを、その対象としていた。こうしたことを踏まえると、限定的で測定しやすい学力に、測定対象が偏っていると指摘せざるを得ず、学力全般に対して効果があることを示すエビデンスとみなすことは難しい。

以上を踏まえると、ハッティの研究における統合メタ分析から得られた「形成的評価」に対する統合効果量の妥当性や信頼性は、低いものであると言わざるを得ない。「何に対して」、「どのような手法」が有効なのかということが、そもそも統一されておらず、その概念規定もまったく行われていないのである。そのため、0.9 として算出された統合効果量は、そのラベルによって表されているものとは無関係な、意味をもたない「ただの数値」に過ぎない可能性がある。それにもかかわらず、ハッティは、統合効果量を有効数字 2 桁、つまり 0.01 の差によってランキング化してしまうのである。

加えて、このようなハッティの研究成果が提供する「数値」が、形成的評価が学力向上に有効であることの「わかりやすいエビデンス」として、しばしば引用されてしまっているという現状がある。たとえば、

Google 等の web 検索エンジンで「formative evaluation (assessment), effect size, 0.9」というキーワードで検索を行うと、ハッティの示す 0.9 という「エビデンス」を形成的評価が有効であることの論拠として用いている研究論文が数多くヒットする。また、国内の文献においても同様の状況が確認できる¹⁰。ここで言いたいことは、形成的評価が有効でないということではない。そうではなく、形成的評価の有効性を語るエビデンスとしてこのハッティの数値を用いることは、不適当であり、形成的評価研究への信頼を失いかねないということである。

以上のハッティによる可視化研究に対する批判に対しては、たとえば、「カテゴライズを精査しよう」「アウトカム概念規定を精査しよう」と応答していくことが想定される。つまり、こうした批判の一部は、ハッティによる可視化研究における瑕疵や不手際を批判するものであっても「統合メタ分析」という手法（方法論）自体を明確に否定するものとはならない。そこで、次に、ハッティによる可視化研究の手法が構造的に抱える問題について検討する。

3.2.2 可視化研究の方法が構造的に抱える問題

第一に、メタ分析研究が、さらに統合化されること（二階層化）によって、コンテキストが極度に抽象化されてしまう問題である。メタ分析を行う研究では、一般的に、その対象とする一次研究を厳密にレビューし、その取捨選択が行われる。一方、ハッティの手法は、項目要因としてのラベルを設定して、それに関連するとみなされるメタ分析研究の効果量を集めて平均化するというものである。結果として、得られるデータや数値は、極度に文脈が捨象されたものとなり、適切な推論を行うことが難しい状態となっている。なお、ハッティ自身は、著書『教育の効果』で、各項目の統合効果量に対する因果関係について推論しているが、それは必ずしもメタ分析研究等を参照して厳密に論じているものとはなっていない。

ここで言いたいのは、メタ分析研究を通して得られる量的データが、教育実践や教育研究にまったく無意味なものだということではない。メタ分析研究においても、もちろん制限や課題もあるものの、その限界性を自覚したうえで、メタ分析から得られる量的データ

を読み解いていくことは、教育研究においても教育実践においても意義がある。なぜならば、メタ分析を試みる研究論文においては、多かれ少なかれ、その著者は、一次研究の内実を目を通した上で、メタ分析の対象とする研究を取捨選択し、結果として得られる数値に対して慎重な解釈や考察を行っているからである。しかし、ハッティの場合、一次研究までは目を通してない旨を公言しており、加えて、その読者も、彼の論文や著書を読むことで、統合効果量が算出された教育文脈や、その教育実践の具体を参照することは難しい（後述するが、近年は追跡不可能な状態になっている）。結果として、コンテキストが極度に捨象されてしまっており、統合効果量を読み解いていくこと、データを適切に推論・検証していくことがほとんど難しい状態になってしまうのである。

教育実践は、主に教師の専門的判断によって舵が取られていく。教育に限らず、専門職の仕事において、臨床的・実践的判断が重要であることは、エビデンスに基づく政策・実践の推進者も批判者も合意する点である。たとえば、「エビデンスに基づく医療（EBM）」¹¹の提唱者であるサケット（David Sackett）は、「外的な臨床のエビデンスは、個人の臨床的専門知（clinical expertise）に情報を提供することはできるが、それに取って代わることはできない。外的なエビデンスが個々の患者に適用されるかどうか、そして、適用される場合、それをどのように臨床的意思決定に組み込むべきか判断するのは、この専門知なのである」と述べ、専門家の専門知を中心に据えつつ、それに深い情報と思慮を与えるものとしてのエビデンスという理解を示している（Sackett et al., 1996, p.72）。

では、統合メタ分析がもたらす「知」は、教師という専門職がよりよい判断と実践を行うための「有益な知」を提供するものとなり得ているのだろうか。これが、重要な論点となる。簡潔に言ってしまうと、ハッティが示す研究知見（統合効果量の数値）から、何かの効果的な教育実践を編み出すことは、原理的に難しい。たとえば、「形成的評価」のラベルから、応答責任（教育改善）に向けた形成的評価の具体的方法論が明らかになるだろうか。ハッティが示す情報では、どのような形成的評価を、どのように実践すればいいかが、明らかにはならない。そして、どのような条件下

やコンテキストにおいて、どのような手法や要因が、どのような学力（目標）に対して、どの程度有効に機能するのを思慮深く推論するだけの十分な情報が得られないのである。

第二に、以上の論点とも関連して、ランキング化に代表される、研究成果の公表や表現における問題がある。ハッティは算出された統合効果量に基づいて、さまざまな教育方法等を序列化(ランキング化)するが、これには重大な問題が伴う。こうした表現方法は、結果として教師に対して上位の方法を安直に採用することを促し、エビデンスを「使う」側に、不適切なメッセージを送っている可能性がある。なお、ハッティは、この問題を他の研究者から指摘された際に、「ランキング化はやめる」と公言したものの、一向にそれを止める気配は示していない（図3）。

Hattie's 2018 updated list of factors related to student achievement: 252 influences and effect sizes (Cohen's d)

Source: J. Hattie (December 2017) visiblelearningplus.com
Diagram: S. Waack (2018) visible-learning.org

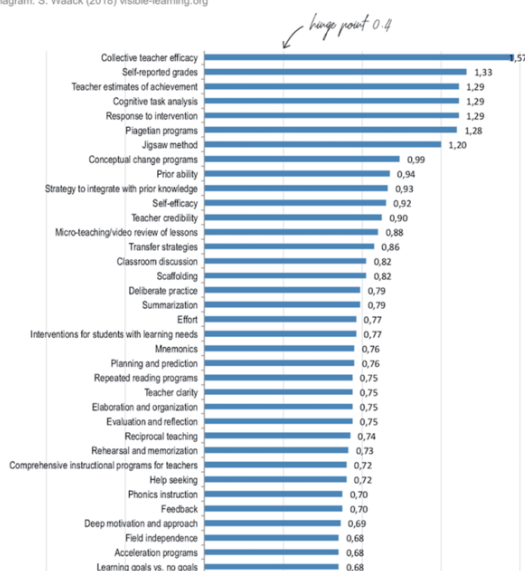


図3. ハッティによる2018年版（最新版）ランキング
出典：Hattie Ranking: 252 Influences And Effect Sizes Related To Student Achievement (URL: <https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/>)

一般的に、エビデンスには、「つくる（産出）」「つたえる（伝達）」「つかう（活用）」という三つのプロセスがあるとされる。この三つの視点を踏まえると、ハ

ッティの研究は、エビデンスを「つくる」というよりも、むしろ「つたえる」役割に重きがある。統合効果量という数値はハッティが生み出している（つくる）ものではあるものの、彼の研究は、既存のメタ分析研究を収集して分類し、統合効果量を算出するものであり、その結果を実践者に働きかける役割が大きいことに鑑みると、「つたえる」側面が強い。

エビデンスを「つかう」のは、主に教育の実践者や教育政策の立案者であるが、つくる主体とつかう主体の間には、しばしば職務内容・文化等の面でギャップが存在し、よって、このギャップを埋める仲介機関や利害関係の調整役が求められるという（石井, 2015）。たとえば、効果が実証的に確認されたエビデンスをレビューし要約し、ウェブサイトで情報提供をしている米国のWWC（What Works Clearinghouse）や、定量的データのみならず定性的データも含めて、政策や実践に活用可能なエビデンスを定量的に集めてレビューしている、英国のEPPI（Evidence for Policy and Practice Information and Co-ordination Center）が知られている。こうした機関は、実践者や政策立案者が参照し、彼らの活用にあつたような情報を提供し、エビデンスを作る側と使う側の橋渡しをうまく行おうとしている。ただ、ハッティによる可視化研究は、このような役割を果たすものとしては不十分である。この役割を適切に果たすためには、少なくとも、エビデンスを使う側、すなわち教育実践者らが、さまざまなデータを思慮深く読み解き、自身の実践的知識と合わせて、教育実践に対する意思決定に用いることができるようなものでなければならないのである。

第三に、ハッティによる近年の可視化研究では、効果量算出の元となっているメタ分析研究が明示されておらず、結果として、統合の元となったメタ分析やその一次研究への追跡可能性が失われてしまっているという深刻な問題がある。そして、この問題は、教師の専門性養成のための教材に見られる問題とも、結びついている。

少なくとも Hattie (2009) の段階では、統合メタ分析の元となったメタ分析の論文リストが巻末で提示されており、それらメタ分析研究を参照したり、一次研究にまで遡ったりすることができた。本稿において、「形成的評価」の項目の中身を検証することができた

のも、彼の研究の元となった情報源が公開されていたからである。しかし、近年のハッティによる可視化研究の web サイトでは、統合メタ分析の元となったメタ分析の論文が何かが、明らかにされない状態となってしまうている。そこで示されているのは、教育効果に関する要因項目としてのラベルと、それに対する統合効果量の数値のみである。そして、各要因は、変わらず 0.01 刻みで序列化されている。また、教師の専門性発達に関するプログラム・コンテンツ等を提供している「Visible Learning Plus（<https://www.visiblelearning.com/>を参照）」では、赤・黄・青などとシグナル表示を用いて、さまざまな指導方法における効果（「エビデンス」）が一眼でわかるような、ポスターやリーフレットが作成されている（図 4）。ハッティ自身、教育実践へのランキングの安直な適応は、好ましくないとはいっているものの、「Visible Learning Plus」において提供されている専門性発達プログラムにおいては、次に示すような教材が用意されているのである。

このような情報の提供方法は、教師の実践における専門的判断を助けるようなエビデンスとはなり得ていない。これらは、実践者の専門的な意思決定に資するようなデータの提供とはならず、そこにあるのは、何を意味するのか不明瞭な「項目名」とそれに付加された単なる「数値」だけなのである。

以上、本章では、ハッティによる可視化研究の内実

を探り、それが抱える諸問題を浮き彫りにしてきた。次章では、ハッティの研究を手がかりとしながら、教育実践と実証的・量的エビデンスの関係性のあり方について検討する。

4. ハッティの研究成果と教育実践の関係

4.1 ハッティの研究が教育実践に対して与える影響

ハッティの研究成果が教育実践にもたらす影響を素朴で好意的に考えるなら、これまで自明とされてきた教育実践の常識を批判的に見る視点を与えてくれるものだと言える。すなわち、実際には効果がないのに効果があるように語られてきた教育実践を白日のもとにさらし、より効果のある実践へと改善する契機を与えてくれるわけである。ハッティ自身も「教育のカーゴ・カルト死すべし」と題する論考において、迷信や民間伝承のように信じられてきた誤った教育実践を正す必要があると主張している (Hattie & Hamilton, 2018)。彼は教師をはじめとした教育関係者に対して、「能力が伸びるとはどういうことかを論理的・省察的に議論すること、児童生徒の成果物、評価結果、そして肉声といった児童生徒の能力の伸びに対する介入の効果のエビデンスを収集すること、そして、児童生徒の学力、学習態度、学び続ける意欲に対して一定の効果をもたらす介入を確実に実施すること」を求める一方で、彼の研究成果は「効果量の順位表のトップ 10 を引き合い

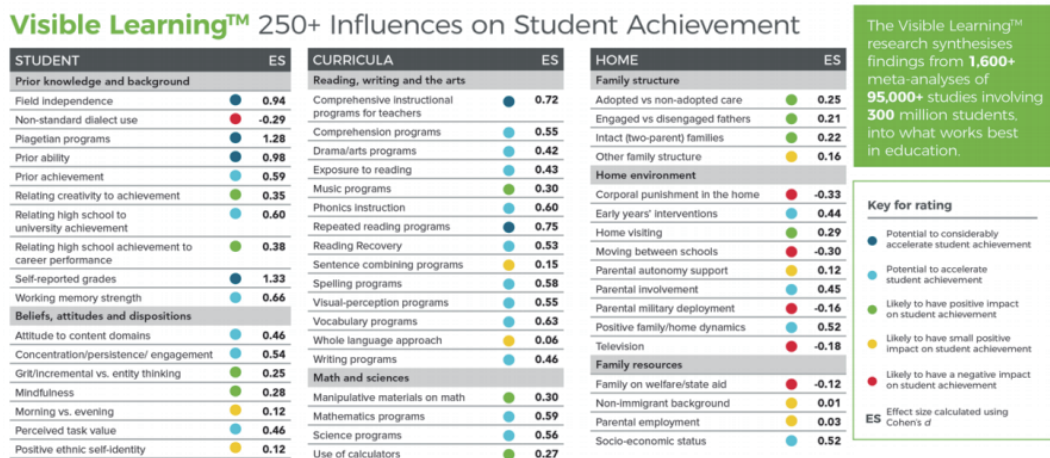


図 4. 教師向けコンテンツの一例（一部抜粋）

出典：Visible Learning Influences (URL: <https://osiriseducational.co.uk/blog/latest-research/visible-learning-infographics/>)

に出したり、ワースト 10 はやらないようにする、といった使い方のためにあるのではない」と述べ、自らの研究が実践において参照される際に留意するように促している（ハッティ、2018, p. 11）。

このような留意点を見落として彼の研究成果が受容された場合、ハッティのランキングだけを見て、より「教育効果が高い」¹²実践を優先させればよいという素朴で乱暴な教育実践改善観に陥ってしまう。すなわち、個別の実証研究がもっていた文脈や目の前の子どもが置かれている文脈が捨象され、ハッティの研究成果が生んだランキングを「科学的に正しいもの」として絶対視し、順位が上のものが個別具体的な文脈にかかわらず優れていると機械的に判断することで教育実践が改善されるという考え方である。このような素朴な「エビデンス・ベース」の捉え方に対して、ハッティ自身は批判的なスタンスをとっており、教師が自らの教育実践を「エビデンスの光の下での批判的省察」（Hattie, 2013, 281）にさらすことを求めている。

一方で、少なくとも現状において、彼の研究成果は、批判的省察を求める彼の意図とは関係なく、教育実践がもつ複雑性や文脈依存性を乱暴なまでに単純化し、教育実践の形式化、空洞化を招く危険性をもつものとなっている。ハッティが紹介された国において、彼の研究成果が「聖杯」「すべてを支配する指輪」と形容され、受容されたのも、教育政策論における決定打という意味よりも、広く教育実践を含む教育全般に対する決定打であるという言説であった。日本の教育実践におけるハッティの受容のされ方も、おおむねこのような素朴な見方に基づいていた¹³。結果として、彼の研究成果は、批判的で省察的な実践家としての教師を生み出すよりもむしろ、「エビデンス」の内実を理解せず、ハッティ・ランキングを絶対視して礼賛し、教育実践を空洞化させる教育関係者を生み出すことに寄与してしまっている。

4.2 教師はハッティの研究とどう向き合えるか：これからの熟達者教師

ハッティの研究は教育実践に対して利する面もあれば、そうではない面もある。先述のように教師がハッティの研究成果を「ファストフード」のように理解してしまえば、教育実践に対してはむしろ害になると

さえ言える。では、教師はハッティの研究に対してどのように向き合うことができるのだろうか。ハッティの研究を実践改善のために活用するための条件や留意点は何だろうか。ここでは、教師のリテラシーとエビデンスを「つくる一つたえる一つかう」関係性の二点から考えてみたい。

第一に、教師のリテラシーである。教育実践家としての教師にとって、実証された技術知は実践上のレパートリーを提供するものとして活用できる。たとえば、心理学等が提供する人間の学習に関する新たな知見は、教師が日々の実践を省察する視点を提供してくれる。一方、具体的・総合的で複雑な判断が求められるリアルな教育実践においては、「実践の中の理論（直観的なものだが、なぜそれが妥当かの事後的説明は可能なもの）」が必要であり、「子どもと向き合う教室レベルに近づくほど、より質的で解釈的な知」が求められる（石井、2019, pp. 7-8）。このような臨床知・実践知もまた、実践における教師の省察には不可欠なものであり、理論知を参照し、臨床知・実践知に学びつつ、自らの教育実践を日々修正するという営為が、教師に求められる専門性／専門職性を基礎づける。

このように、技術知を参照しつつ、臨床知・実践知で教育実践を組み立てる場合、たとえば、教師は「今、Aさん／B学級が直面しているXという課題に対しては、一般的にYという方法が有効だとされています。一方、この子／この学級の場合は、Zという方法の方がむしろしっくりくる（効果的）でしょうから、今回はZを試みましょう」というように思考・判断することが想定される。「なぜしっくりくると考えた／感じたのか」というところが、教師の臨床知・実践知が働く部分である。もちろん、ここで行われている教師の判断は、単なる素人の勘ではなく、熟達者のそれである必要がある。このように思考・判断ができることが、エビデンスに基づく教育の時代において教師に求められるリテラシーである。

このとき、熟達者の臨床知・実践知が生活世界的エビデンスにのみ押し込められるものではない点に留意する必要がある。既存の教育言説においては、日々子どもと向き合う生活世界的エビデンスの中で、下町の老舗さながらの職人気質による神秘的・魔術的な「見とり」（教育的タクトと言い換えてもよい）を発揮する

熟達者教師像が強調される傾向にあった¹⁴。教育政策が透明性の論理に基づいて説明責任を果たすことが求められるとされるのに対して、教育実践が納得可能性の論理に基づいて応答責任を果たすことが求められるという言説の構造にもその傾向は看取される（石井, 2015; 今井, 2015）。仮にクライアントを子どもと保護者に限定してみたとしても、「現場の先生がおっしゃっているから」というのが納得につながり、応答責任を果たしたことになる場合もあれば、そうではない場合もある。むしろ、教師がなぜそのように判断したのかに対して、より普遍的で一般性のある「科学的な」回答がなければ納得に至らないという場合も大いに想定される。応答責任をクライアントの納得可能性と解するとき、クライアントの納得の水準が意図せず低く（あるいは極めて好意的で善意に基づく「納得」として）見積もられてしまう可能性には留意しなければならない。

このように考えてみると、熟達者の実践知は、近代科学的エビデンスと生活世界的エビデンスとを往還するなかで磨かれるものとして捉えなおされる必要がある。教育に関する基礎研究から提起された理論知に学びながら、自らの臨床知・実践知を不断に省察する教師像こそが、現代における熟達者としての教師のあり方として、より望ましい。先のクライアントに対する応答責任に係る問題についても、理論知と臨床知・実践知を相互補完させながら日々の教育実践に取り組んでいる教師のあり方にこそ、応答責任が果たされたものとして納得可能性に開かれていると言えよう。

4. 3 エビデンスと向き合う教師を支える仕組み

第二に、エビデンスを「つくる一つたえる一つかう」関係性である。先述のように、教師が理論知を参照しながら、臨床知・実践知を磨き上げることで、自らの教育実践における判断を見極める力が、エビデンスに基づく教育の時代における教師のリテラシーの重要な一角をなす。しかしながら、教師が教育研究を読み解いて参照し、さらに臨床知・実践知も体得せよというのは、教育実践家であるはずの教師に対して多くを求めすぎているという点で現実的ではない。教師のリテラシーの射程は、少なくともその平均値においては、より限定的に想定されるべきだろう。では、教師が先

述のようなリテラシーを発揮するためには、どのような環境整備が必要になるだろうか。

奇しくもこの問いはハッティの問題意識と重なる。彼は自らの研究が、教育政策における政策立案者の意思決定に貢献することのみならず、教育実践における教師の意思決定に貢献することをも企図している。彼はエビデンス・ピラミッドの頂点に位置するメタ分析をさらにメタに用いて知見を統合することで、教育に影響を与える要因を効果量によって一直線上に置き、どの要因が効果的かをわかりやすく可視化することを試みた。

しかし一方で、教育実践は文脈依存性の高い営為である。ハッティのような統合メタ分析から導出された研究結果は、平均的傾向を示すに留まるものであり、教育の対象である子ども（学習者）が多様性に富む存在である以上、平均的傾向を示す数値に従うことだけでは効果的な教育にはならない。たとえば「日本の子ども」というような学習者の総体に対してどのような教育政策を採用するかというマクロな文脈においては、「これが有効であると数字で出ています」と主張することが説明責任として機能するとしても、目の前の子どもに合致しない教育方法を「これが有効であると数字で出ています」と主張することは、説明責任を果たすという名の下に、それぞれの子どものに適した教育方法の選択を怠った教師の責任回避としてしか機能せず、単なる強弁に陥る¹⁵。ドイツにおいてハッティ受容に対する批判としてすでに指摘されているとおり、ハッティの研究は教育にとって「何が有効か（What works）」を示すものではあるが、「いつ有効か（When it works）」については捨象されている（熊井, 2016）。

以上を踏まえると、教師とエビデンス、教育実践とエビデンスの関係性を構想するときには、エビデンスを強弱の垂直方向（縦軸）のみで考えるのではなく、この文脈において妥当かどうかという水平方向（横軸）でも考える必要がある。ハッティの研究は教育のアウトカムが効果量として定量化されたもののみを「強いエビデンス」として採用し、そうではないものは「弱いエビデンス」として検討対象外としている。しかしながら、たとえば、子どもへの個別の対応における判断力を養う場合には、ケース・スタディのような手法に基づく方がより有効な策であることが少なくない。

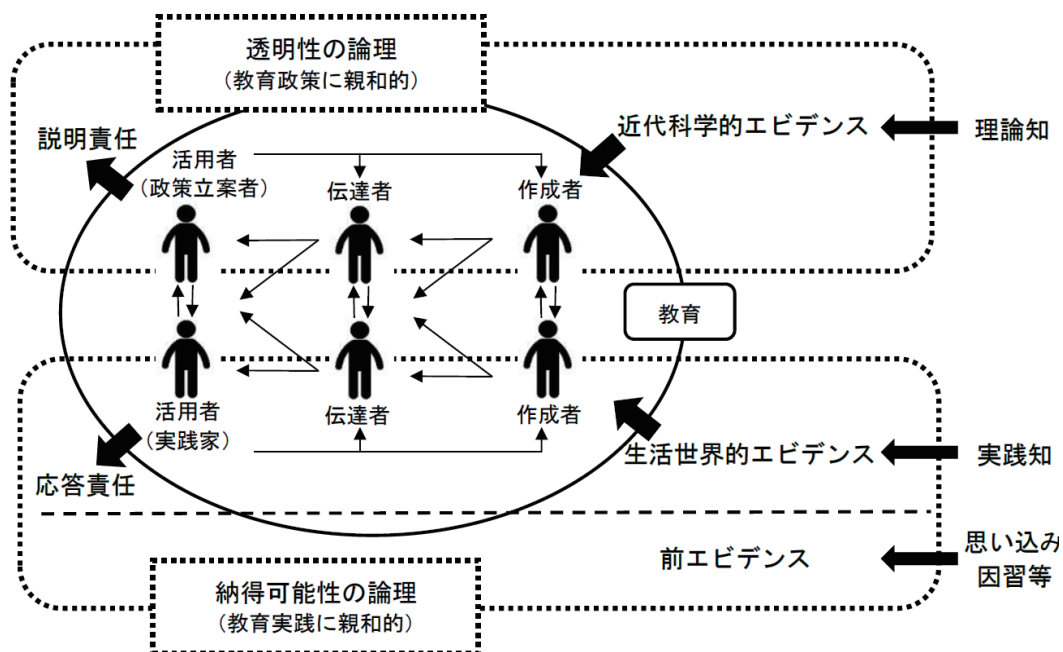


図 5. 教育とエビデンスを取り巻く「つくる一つたえる一つかう」関係
出典：今井 2015 の図を参考にしつつ筆者作成

ケース・スタディは、授業の事例や子どもの事例、あるいは学校内で生じた事件や事故のようなトラブル等から、一般性の高い原理や法則を帰納的に導き出すものであり、たとえば、個別の授業事例を分析することで教授／学習の改善方法を探る授業研究等がこれに当たる。ケース・スタディは、その文脈依存性の高さゆえ、エビデンス・ピラミッドでは「弱いエビデンス」とされているが、個別の事案に対応する実践的な方策を探る際にはむしろ、このようなケース・スタディを参照の方がより示唆的である。このように、少なくとも教育実践においては、エビデンス・ピラミッドで「弱い」とされているエビデンスは、「強い」とされているエビデンスに対して絶対的に劣後しているわけではなく、目的に応じて使い分ける必要がある。

以上を踏まえると、教育実践におけるエビデンスの活用を目指す場合、エビデンス・ピラミッドだけでなく、エビデンス・マトリックスとして構想することが有効であろう (Hansen & Rieper, 2009, 157)。エビデンス・マトリックスでは、エビデンスがもつ強弱 (縦軸)

のみならず、どのような場合に、どのようなエビデンスを用いるのが適切なのか (横軸) が提示される。これによって、教師は今自分が依拠している文脈に沿って、参照すべきエビデンスを判断・選択することが可能になる。

もちろん、参照すべきエビデンスの選択には、教師が熟達した臨床家である必要がある。エビデンスがマトリックスとして配列されても、参照に際して教師が判断・選択を誤っては意味がない。そのため、教師には提示されたエビデンス・マトリックスを適切に見る目、すなわちリテラシーが求められる。ただし、教師がエビデンスの背後にある個別の研究すべてを見極めよというのは現実性に欠ける。そこで、先述した「つくる一つたえる一つかう」という教育研究、教育実践における関係が重要になる (図 5)。現代では、「エビデンス」とされるものが近代科学的エビデンスと生活世界的エビデンスに二分され、透明性の論理と納得可能性の論理という二つの軸においてそれぞれの世界に分断されている (今井, 2015)。このような状況を打開

するために、(1) 個別の研究によってエビデンスをつくる教育研究者（エビデンスの作成者）、(2) エビデンス・マトリックスを構想・伝達するものとしての教育学研究者（エビデンスの伝達者）、(3) 提示されたエビデンス・マトリックスを活用・省察する実践家（エビデンスの活用者）という三者によって、近代科学的エビデンスと生活世界的エビデンスを架橋し、教育実践の改善につなぐことが求められている。

エビデンスの作成者、伝達者、活用者という三者のなかで、実践家である教師は作成・伝達された研究成果を批判的に読み解き、活用・省察する役割を主に担う。一方、伝達者としての教育学研究者には、教師の実践を下支えするために、教育研究の論理と教育実践の論理を架橋する役割が期待される。このような関係性のなかであれば、教師はエビデンスに基づく研究成果を見極めるリテラシーを養うことができる。

ただし、ここで留意したいのは、教師がエビデンスを活用するだけでなく省察する役割を担っている点と、教師は活用者という役割に固定されるわけではない点である。これはすなわち、教師が活用の役割のみを担い、ただ提示されるエビデンスに関する情報を理解し、受容せよというのではないことを意味する。活用する過程で自らの教育実践を通して研究成果を批判的に省察することもあれば、伝達されたものの源流に遡って個別研究を参照することもあるだろう。あるいは、自ら授業研究によってエビデンスをつくり、また、提示されたエビデンス情報を伝える、使う、すなわち作成者でもあり伝達者でもあり活用者でもあるという教師像も十分にあり得る¹⁶。教師がエビデンスを読み解いて活用することを「理論の実践化」とするならば、実践知・臨床知に基づいてエビデンスを批判的に省察することは「実践の理論化」である。

日本の場合、校内研修という形での授業研究文化があり、そこには教師の専門性（実践・現場を重視する日本の教師文化に依拠するならば「専門家」「専門性」という言葉は必ずしも好んで使用されるとは言えないが、より現場に近い言葉で言えば「教師ならば〇〇はできなければならない」という職人的な気質に基づくプロ意識）が、小さくない範囲で共有されている。また、このような教師文化（教員文化）を基盤としながら、教師が自らの実践を実践記録としてナラティブに

綴ることで、臨床知・実践知の共有が試みられてきた歴史がある。このような日本の授業研究の伝統の中で連綿と受け継がれてきた臨床知や実践知が、認知心理学や現象学等の観点から検討対象とされ、教育の理論と実践を往還しながら研究されてきたことも忘れてはならない（金井他、楠見編、2012；姫野、生田編、2019等参照）。自生的な授業研究文化がある日本の学校には、教師共同体のなかで相互に実践を磨き上げる教育の文化を再発見することが可能である。

以上のとおり、教師が基礎研究の成果物を参照しつつ、しかし基礎研究の成果物に盲従することなく実践を築き上げるためには、(1) 近代科学的エビデンスと生活世界的エビデンスとを往還する関係性のなかで、教師が研究を見極めるリテラシーを養うこと、(2) それを下支えするものとして、「つくる一つたえる一つかう」関係性のなかで、教師が省察の実践家としての力を最大化するための環境を整備することが必要である。ハッティはさまざまな新しい考え方を、無批判かつ一貫性もなく受け入れる学校・教師を問題視しているが、教師がハッティの研究に向き合うときに無批判かつ一貫性もなく受け入れることもまた同様に問題視されて然るべきことである。

5. おわりに

本稿ではハッティによる可視化研究に焦点を合わせて、統合メタ分析という手法が産出された「エビデンス」に、教育実践に携わる者がどのように向き合うべきかを検討した。本稿の問いは次の2つである。(1) 現在までのハッティによる可視化研究がもつ問題点は何か、(2) 仮にハッティによる可視化研究が限りなく完成に漸近し、統合メタ分析による完全な可視化研究が実現したとして、それは教育実践において直接参照可能なエビデンスとなるのか。

まず、第一の問いに対して、ハッティによる可視化研究における「形成的評価」の要因項目を一次研究にまで遡ってみた結果、カテゴリーの妥当性の問題、サンプリングの偏りの問題、効果量の対象となる教育アウトカムの曖昧性の問題があることが明らかになった。これらはハッティによる可視化研究そのものにおける瑕疵である。次に、第二の問いに対して、実践的文脈が捨象されるために教育実践の具体像を提起し得

ない問題、ランキング化のような研究成果の公表や表現に関する問題、元となる研究を公開しないことによる追跡・検証不可能性の問題を指摘した。これらはハッティによる可視化研究において用いられている研究方法が構造的にもつ弱みである。

以上を踏まえ、本稿では、ハッティが採用した統合メタ分析という研究手法一般が、教育政策立案等、教育をマクロに検討する際のひとつの指標として有効性をもつことを認めつつも、しかしながら、少なくともハッティによる可視化研究には現在のところ瑕疵があり、統合メタ分析が具備しておくべき要件を十分に満たしているとはいいたいと結論づけた。統合メタ分析が導き出した効果量という脱文脈化された情報を受動的に受け入れて教師が実践を行うことが「科学的に正しく」、教育改善に資すると考える教師観・教育観は、「エビデンス」に唯々諾々と従って機械的に実践を行うだけの教師、あるいはファストフードのように断片化された「エビデンス」を消費するだけの教師を増やすことにつながる。結果として教師の自律性は損なわれ、エビデンスによって脱専門化された教師が量産されれば、教育実践は空洞化する。ハッティ自身がデューイを引用して述べているとおり、「エビデンスは行動規範を提供するものではなく、知的な問題解決のための仮説、あるいは教育の目的についての探究をするための仮説を提供してくれるだけである」ことには、実践家こそ自覚的になる必要がある(Hattie, 2009, p. 247)。

そこで本稿では、教育とエビデンスを取り巻く論理と責任のダイナミズムを整理し、エビデンスを「つくる一つたえる一つかう」関係のなかで、教師が必要に応じて理論知、臨床知・実践知を自在に活用できるようなあり方がどのように構想されうるのかを示した。エビデンス・ピラミッドに基づくエビデンス観は、教育実践に対して平均値として「何が有効か」を提起するものの、多様な学習者／学級／学校を対象とするために個別の文脈を重視する教育実践という営為には必ずしも親和的ではない。このような教育実践がもつ特性を踏まえ、本稿では、より教育実践に資するエビデンスの捉え方として、エビデンス・マトリックスという考え方が有効であることを提示した。

「つくる一つたえる一つかう」関係のなかでエビデンス・マトリックスを通して理論知を参照しつつ、臨

床知・実践知に基づきながら教育実践を日々つくっていくには、教師のリテラシーが不可欠である。これからの熟達者教師は、基礎研究の成果物を参照しつつ、しかし基礎研究の成果物に盲従することなく、日々の実践を築き上げる省察的实践家としての役割が期待される。そのような教師ならば、ハッティの研究成果を批判的に検討しつつ、日々の実践に活かしてしていくだろう。このような批判的な鑑識眼をもつ教師は、ハッティ自身が望んでいた力量のある教師像でもある。

註

¹ こうした試みは、松下良平が指摘するところの「直進の文明」の典型である。松下は、「逍遙の文化」と対比させながら「直進の文明」という思想が、教育研究や教育実践の根底でどのように広がっているのか、その根源的な理念について考察している(松下, 2020)。

² 被引用数は Google Scholar より。https://scholar.google.co.jp/scholar?hl=ja&as_sdt=0%2C5&q=visible+learning&btnG= [2021/02/28確認]

³ ハッティの *Visible Learning* の邦訳を監修した山森は、監訳者解説において「本書のセンセーショナルな特徴や視覚的訴求力の高さは、ややもすれば筋違いな解釈や見当違いな判断を招きかねない」と指摘している(山森, 2018, p. 13)。

⁴ メタ分析が一般的に抱える問題点としては、「リンゴとオレンジ」問題、「ゴミを入れてもゴミしか出てこない」問題、公表バイアスの問題等が挙げられる。本稿ではハッティの研究に焦点化するため、メタ分析一般がもつ問題点について屋上屋を架すことは避ける。

⁵ Evidence-informed Policy Network は、「保健システムを強化し、必要とする人々に適切なプログラム、サービス、薬剤を提供するために、保健政策立案における研究エビデンスの体系的な利用を促進するために、世界保健機関によって設立されたネットワーク」である(World Health Organization, n. p.)。

⁶ ここで引用した Hattie (2010) は、原田他編著(2015)に邦訳されて掲載されている。本稿では、この邦訳を参考にしつつ、原典の語義に照らして一部筆者訳出した。

⁷ スヌークらは次の4点に留意するように述べている。

(1)「彼[ハッティ]自身の頻繁な警告にもかかわらず、政治家は彼が支持しない方針を正当化するために彼の研究を使用するかもしれない、彼の研究はそれを認めていない」、(2)「教師や教師教育者は、単純なやり方で調査結果を使用しようとするかもしれない」が、エビデンスは行動規範を提供するものではなく、あくまで仮説を提供するだけのものなので、「彼[ハッティ]の発見は教育政策や実践に直接適用することはできない」、(3)「学校の効果に関する定量研究は、その「歴史的、文化的、社会的背景、ならびに家庭や地域社会の背景との相互作用から切り離して」提示されている」、(4)「教育の目的や学校教育の目的についての議論が不十分であり、それがなければ研究はほとんど意味をもたない」(Snook et al., 2010, pp. 4-5)。

⁸ 管見の限り、ハッティは統合効果量の算出方法を明示していないが、邦訳版『教育の効果』の巻末付録AとBにある情報に基づいて、各メタ分析における効果量の平均値を計算し、統合効果量の数値と比較した結果、これらは一致した。

⁹ なお、これらのメタ分析研究2本の元となっている一次研究に関しても、公表されていて入手可能なものは基本的に入手し、内容の確認を行った。ただし、その詳細については紙幅の都合上、本稿では割愛する。

¹⁰ たとえば、近年では次のような言及がなされている。「ハッティ (J. Hattie) が2009年に刊行した『Visible Learning』では、5万件を超える研究論文を含む800以上のメタ分析調査にもとづき、様々な教育的働きかけの効果が算出されている。その中でも教師による「形成的アセスメント」は、効果量が0.9と算出されており、非常に高い効果をもたらすことが示唆されている」(二宮, 2019, p. 109)。

¹¹ 「エビデンスに基づく医療 (EBM)」は、一定有効に機能しているのに対して、「エビデンスに基づく教育 (EBE)」は、なぜ難しさを抱えるのか。石井 (2015) や松下 (2015) を参考にしながら考えてみたい。

第一に、教育と医療が目指すところ、すなわち、その目的や目標の性質に違いがあることによる。(1) 医療は、治療や快方という比較的に明確で一義的な目標を追求するのに対して、教育は、「何のためか」「何が

良い教育か」といった、教育の目的や目標それ自体が常に論争的であり、多様な能力・態度・人格といった多義的でファジーな目標を追求する (目標の複雑性と非明確性)。(2)目的と手段の関係が、医療の場合、比較的リニアで因果的に捉えることができるのに対して、教育の場合、目的と手段の関係は、相互規定的であり、非因果的な関係にもある。つまり、目的と手段を完全に分離し、教育における目的や目標を固定的で所与なものとし、その効果的・効率的な達成のみを問うというモデルは、教育実践に馴染まない側面がある。

第二に、医療と教育における概念規定の明確さに違いがあることによる。医療における概念は、病名、症状、医薬品などに対しては、それが一意に定まるように明確に定義されている (シャープな用語)。一方、教育における概念は、教育研究・教育実践どちらの世界においても、ほとんどの用語が一意に定まるように明確に定義されているわけではない。用語の定義や解釈それ自体が論争的・多義的であり、それ自体が研究対象ともなっている。ある特定の集団内での共通概念は形成されていても、それが広く一般に定義される概念とはなっていないのである (教育概念のファジーさ)。したがって、どのような状況で、どのような目標に対して、どのような手法・介入・操作が、どれほど効果があるのかということを明確に規定された用語に従って普遍的に語ることは困難である。

第三に、物質的な相互作用か、それとも、意思や感情を持った人間同士の相互作用かという違いがあることによる。医療において、特に、EMEの対象となっているのは、主にどのような薬や治療法が、どのような症状や病に対して、どれほど有効かということである。この場合、基本的には、物質的な相互作用がその中心となっているが、これに対して教育の場合、意思や感情を持った人間同士の相互作用がその中心にある。学習が生じたとすれば、それは学習者が教えられたことを解釈し意味を生成したということであり、相互の解釈によって教育という営為は成立する。また、教えるものと学ぶものの関係性も重要であり、別の場所や別の相手であまうといったことをそのまま適応してもらうまいかないということが多い。

¹² 前節までの検討を通して見えてきたハッティの研究成果に対する一抹の疑念を踏まえ、ここでは実際に教育効果が高いかどうかではなく、彼の研究成果において効果量が高いということが、すなわち教育効果が高いとされたという点を含意させるため括弧書きで「教育効果が高い」と記載している。

¹³ たとえば、オンラインショッピングサイト Amazon においてハッティの研究の邦訳『教育の効果』のレビューを見るだけでも、その一端は十分に垣間見える。このような言説の形成に対して、先述のとおり、ハッティは効果量の順位表（ハッティ・ランキング）を絶対視しないように促したり、教師の批判的省察を説いたり、合理的な対応を求めているが、一方で、「教育効果が低い」とされたものを捨てて「教育効果が高い」とされたものへと移行することを素朴に呼びかけ、素朴な実践改善観を垣間見せたり（Hattie et al., 2016）、最新（2018年）のハッティの研究においてもランキングはいまだ健在であったりと、彼自身の語りには揺れがある。

¹⁴ このような教師のタクトを重視する立場に対しては、「教育の目的はそもそも教師の自由を獲得することではない」にもかかわらず、「教師の裁量と自由をより認める方向に帰結する」主張に終始しているのではないかとの批判的な見方もある（杉谷, 2020, p. 98）。しかし、教師の裁量と自由を重視する主張の基盤には教師の専門性を高めるという目的がある点には留意しておきたい。教師が熟達者となる過程で、専門性を高めるためには、あるいは高めた専門性を発揮するためには、教師の自律性（裁量と自由）が必要であるということである。教師が実証的エビデンスに隷属することを自明視することで教師の自律性が掘り崩され、結果として教師が臨床知・実践知を体得する機会を失い、教育実践が痩せ細ることは避けるべきである。

¹⁵ もちろん、教育の結果に対する教師の責任が無制限に拡大されるべきではない。医療過誤の問題においても予見可能性と回避義務が問われるが、医療以上に曖昧で複雑に要因が絡み合う教育という営為において、教育における成果の予見可能性と回避義務が軽々に拡大されることは、ますます教師の実践を委縮させ、結

果として子どもに不利益をもたらす。

¹⁶ ただし、「エビデンスづくり」が行われる際に、予算の獲得や獲得した予算の正当化と抱き合わせになれば、「実証」と呼ぶものとしては常に疑問がつきまとうことになる。どこからともなく出てきた右肩上がりのグラフによって、いつの間にか成績が伸びたことになっているというようなことが行われれば、メタ分析批判の一角である公表バイアスと同類の問題をもつことになる。

参考文献

- Burns, M. K. and Symington, T. (2002). A meta-analysis of prereferral intervention teams: Student and systemic outcomes. *Journal of School Psychology*, 40 (5), 437-447.
- Fuchs, L.S. and Fuchs, D., (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional children*, 53 (3), 199-208.
- Hansen, H. F., and Rieper, O. (2009). The Evidence Movement: The Development and Consequences of Methodologies in Review Practices. *Evaluation*, 13 (2), 141-163.
- Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Hattie, J. (2010). On Being a “Critic and Conscience of Society”: The Role of the Education Academic in Public Debates. *New Zealand Journal of Educational Studies*, Heidelberg, 45 (1), 85-96.
- Hattie, J. (2012). *Visible Learning for teachers: Maximizing impact on learning*. Routledge.
- Hattie, J., Masters, D., and Birch, K. (2016). *Visible Learning into Action: International Case Studies of Impact*. Routledge.
- “Hattie Ranking: 252 Influences And Effect Sizes Related To Student Achievement” Visible Learning. <https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/> [Access: 2019/05/28]
- Hattie, J. & Arran Hamilton. (2018). Education Cargo Cults Must Die, Corwin. <https://www.visiblelearningplus.com/groups/cargo-cults-must-die-white-paper> [Access: 2019/06/20]

- Snook, I., O'Neill, J., Clark, J., O'Neill, A., and Openshaw, R. (2009). Invisible Learnings?: A Commentary on John Hattie's Book: "Visible Learning: A Synthesis of over 800 Meta-analyses Relating to Achievement." *New Zealand Journal of Educational Studies*, 44 (1), 93-106.
- Snook, I., Clark, J., Harker, R., O'Neill, A., and O'Neill, John. (2010). Critic and conscience of society: A reply to John Hattie. *New Zealand Journal of Educational Studies*, 45 (2), 93-98.
- Terhart, E. (2011). Has John Hattie really found the holy grail of research on teaching? An extended review of Visible Learning. *Journal of Curriculum Studies*, 43 (3), 425-438.
- Myburgh, S. (2016). Critique of Peer-reviewed Articles on John Hattie's Use of Meta-Analysis in Education, University of Bath Department of Education Working Papers Series.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., and Richardson, W. S. (1996). Evidence Based Medicine: What It Is and It Isn't. *British Medical Journal*, (312), 71-72.
- Wachter, K. W. (1988). "Disturbed by Meta-Analysis?" *Science*, 241 (4872), 1407-1408.
- World Health Organization. EVIPNet: Evidence-informed Policy Network. <https://www.who.int/initiatives/evidence-informed-policy-network>
- 石井英真 (2015) 「教育実践の論理から『エビデンスに基づく教育』を問い直す：教育の標準化・市場化の中で」『教育学研究』日本教育学会、第 82 巻、第 2 号、216-228。
- 石井英真 (2019) 「学力調査時代を読み解く評価リテラシー」『教育方法の探究』京都大学大学院教育学研究科・教育方法学研究室、第 22 号、1-10。
- 今井康雄 (2015) 「教育にとってエビデンスとは何か：エビデンス批判をこえて」『教育学研究』第 82 巻、第 2 号、2015 年、188-201。
- 金井壽宏、楠見孝編 (2012) 『実践知』有斐閣。
- 熊井将太 (2019) 『『エビデンス・ベース』時代の教育実践研究：ジョン・ハッティの Visible Learning をめぐる議論から』杉田浩崇、熊井将太編『『エビデンスに基づく教育』の関を探る』春風社、250-280。
- 杉谷和哉 (2020) 「エビデンスに基づく政策における責任論の再考：医療・教育学のエビデンス論を参考に」『医療福祉政策研究』日本医療福祉政策学会、第 3 巻第 1 号、85-100。
- 二宮衆一 (2019) 「安藤輝次著『みんなで「深い学び」を達成する授業：形成的アセスメントで子どもが自ら学びを把握し改善する』」『教育方法学研究』第 44 巻、109-110。
- ハッティ・J. 著、山森光陽監訳 (2018) 『教育の効果：メタ分析による学力に影響を与える要因の効果の可視化』図書文化。
- 原田信之、ヒルベルト・マイヤー編著、宇都宮明子他訳 (2015) 『ドイツ教授学へのメタ分析研究の授業：ジョン・ハッティ「可視化された学習」のインパクト』デザインエッグ社。
- 姫野完治、生田孝至編 (2019) 『教師のわざを科学する』一草書房。
- 松下良平 (2015) 「エビデンスに基づく教育の逆説：教育の失調から教育学の廃棄へ」『教育学研究』第 82 巻、第 2 号、2015 年、202-215。
- 松下良平 (2020) 「二一世紀における教育観の転換：直進の文明と逍遙の文化の対立を調停する」『教育哲学研究』121 号、112-137。
- 亘理陽一 (2014) 「メタ分析の手続きについて：大切なことは全て効果量が教えてくれるのか？」『外国語教育メディア学会 関西支部メソドロジー研究部会報告論集第 5 号』外国語教育メディア学会 関西支部メソドロジー研究部会、64-74。

謝辞

本稿は、第 30 回カリキュラム学会（京都大学大会）における自主企画セッションⅡ「教育における『エビデンス』を取り巻く論争点—ジョン・ハッティの研究をどう読むか—」（発表者：原田信之、山森光陽、森本和寿、石田智敬、指定討論者：松下佳代、司会者：徳島祐彌）における議論を踏まえて、執筆したものである。本自主企画セッションに発表・参加いただいた方々に、感謝申し上げたい。

（日本学術振興会特別研究員・博士後期課程、
日本学術振興会特別研究員・博士後期課程）
受理 2021 年 3 月 1 日