

教育評価の意義と課題

— Alfie Kohnの「評価批判」をめぐって—

友野 清文 (現代教育研究所所員 総合教育センター)

はじめに

2019年度から教職課程の科目編成の変更が行われ、その中心に「教職課程コアカリキュラム」が置かれる。この教職課程コアカリキュラムでは、領域毎に「学生が修得する資質能力を『全体目標』、全体目標を内容のまとまり毎に分化させた『一般目標』、学生が一般目標に到達するために達成すべき個々の規準を『到達目標』¹として規定している。

評価については、例えば以下のような到達目標が示されている。([]内は科目の領域)

- ・主体的学習を支える動機付け、集団づくり、学習評価のあり方について、発達の特徴と関連付けて解している。[幼児、児童及び生徒の心身の発達及び学習の過程]
- ・学習評価の基礎的な考え方を理解している。[教育の方法及び技術 (情報機器及び教材の活用を含む。)]
- ・カリキュラム評価の基礎的な考え方を理解している。[教育課程の意義及び編成の方法 (カリキュラム・マネジメントを含む。)]

この他、各教科の指導法、道徳科、総合的な学習の時間、特別活動、進路指導などで評価について触れられている。

このように評価は教育活動のすべての側面で行われるものであり、教職課程科目でもそれを扱うことが明確化されている。

同時に2017年3月に告示された小学校学習指導要領でも総則において「学習評価の充実」について触れられている。すなわち「第3 教育課程の実施と学習評価 2 学習評価の充実」で以下のように述べられている。やや長くなるが解説編の該当部分を引用する。²

(1) 指導の評価と改善 (第1章第3の2の(1))

(1) 児童のよい点や進歩の状況などを積極的に評価し、学習したことの意義や価値を実感できるようにすること。また、各教科等の目標の実現に向けた学習の状況を把握する観点から、単元や題材など内容や時間のまとまりを見通しながら評価の場面や方法を工夫して、学習の過程や成果を評価し、指導の改善や学習意欲の向上を図り、資質・能力の育成に生かすようにすること。

本項と次項は、学習評価の実施に当たっての配慮事項を示している。

学習評価は、学校における教育活動に関し、児童の学習状況を評価するものである。「児童にどういった力が身に付いたか」という学習の成果を的確に捉え、教師が指導の改善を図るとともに、児童自身が自らの学習を振り返って次の学習に向かうことができるようにするためにも、学習評価の在り方は重要であり、教育課程

や学習・指導方法の改善と一貫性のある取組を進めることが求められる。

評価に当たっては、いわゆる評価のための評価に終わることなく、教師が児童のよい点や進歩の状況などを積極的に評価し、児童が学習したことの意義や価値を実感できるようにすることで、自分自身の目標や課題をもって学習を進めていけるように、評価を行うことが大切である。

実際の評価においては、各教科等の目標の実現に向けた学習の状況を把握するために、指導内容や児童の特性に応じて、単元や題材など内容や時間のまとまりを見通しながら評価の場面や方法を工夫し、学習の過程の適切な場面で評価を行う必要がある。その際には、学習の成果だけでなく、学習の過程を一層重視することが大切である。特に、他者との比較ではなく児童一人一人のもつよい点や可能性などの多様な側面、進歩の様子などを把握し、学年や学期にわたって児童がどれだけ成長したかという視点を大切にすることも重要である。

また、教師による評価とともに、児童による学習活動としての相互評価や自己評価などを工夫することも大切である。相互評価や自己評価は、児童自身の学習意欲の向上にもつながることから重視する必要がある。

今回の改訂では、各教科等の目標を資質・能力の三つの柱で再整理しており、平成28年12月の中央教育審議会答申において、目標に準拠した評価を推進するため、観点別学習状況の評価について、「知識・技能」、「思考・判断・表現」、「主体的に学習に取り組む態度」の3観点に整理することが提言されている。その際、ここでいう「知識」には、個別的事実的な知識のみではなく、それらが相互に関連付けられ、さらに社会の中で生きて働く知識となるものが含まれている点に留意が必要である。また、資質・能力の三つの柱の一つである「学びに向かう力、人間性等」には①「主体的に学習に取り組む態度」として観点別学習状況の評価（学習状況を分析的に捉える）を通じて見取ることができる部分と、②観点別学習状況の評価や評定にはなじまず、こうした評価では示しきれないことから個人内評価（個人のよい点や可能性、進歩の状況について評価する）を通じて見取の部分があることにも留意する必要がある。

このような資質・能力のバランスのとれた学習評価を行っていくためには、指導と評価の一体化を図る中で、論述やレポートの作成、発表、グループでの話し合い、作品の制作等といった多様な活動を評価の対象とし、ペーパーテストの結果にとどまらない、多面的・多角的な評価を行っていくことが必要である。

(2) 学習評価に関する工夫（第1章第3の2の(2)）

(2) 創意工夫の中で学習評価の妥当性や信頼性が高められるよう、組織的かつ計画的な取組を推進するとともに、学年や学校段階を越えて児童の学習の成果が円滑に接続されるように工夫すること。

学習評価の実施に当たっては、評価結果が評価の対象である児童の資質・能力を適切に反映しているものであるという学習評価の妥当性や信頼性が確保されていることが重要である。また、学習評価は児童の学習状況の把握を通して、指導の改善に生かしていくことが重要であり、学習評価を授業改善や組織運営の改善に向けた学校教育全体の取組に位置付けて組織的かつ計画的に取り組むことが必要である。

このため、学習評価の妥当性や信頼性が高められるよう、例えば、評価規準や評価方法を明確にすること、評価結果について教師同士で検討すること、実践事例を蓄積し共有していくこと、授業研究等を通じ評価に係る教師の力量の向上を図ることなどに、学校として組織的かつ計画的に取り組むことが大切である。さらに、学校が保護者に、評価に関する仕組みについて事前に説明したり、評価結果についてより丁寧に説明したりするなどして、評価に関する情報をより積極的に提供し保護者の理解を図ることも信頼性の向上の観点から

重要である。

また、学年や学校段階を越えて児童の学習の成果が円滑に接続されるようにすることは、学習評価の結果をその後の指導に生かすことに加えて、児童自身が成長や今後の課題を実感できるようにする観点からも重要なことである。このため、学年間で児童の学習の成果が共有され円滑な接続につながるよう、指導要録への適切な記載や学校全体で一貫した方針の下で学習評価に取り組むことが大切である。

さらに、今回の改訂は学校間の接続も重視しており、進学時に児童の学習評価がより適切に引き継がれるよう努めていくことが重要である。例えば、法令の定めに基づく指導要録の写し等の適切な送付に加えて、今回の改訂では、特別活動の指導に当たり、学校、家庭及び地域における学習や生活の見通しを立て、学んだことを振り返りながら、新たな学習や生活への意欲につなげたり、将来の生き方を考えたりする活動を行うこととし、その際、児童が活動を記録し蓄積する教材等を活用することとしており（第6章特別活動第2〔学級活動〕の3（2））、そうした教材を学校段階を越えて活用することで児童の学習の成果を円滑に接続させることが考えられる。

以上のように、「指導と評価の一体化」が掲げられ、教育目標により適合した評価観点と評価規準が国レベルで作成される、というシステムが一層進められようとしている。もちろん評価のあり方や方法についてはこれまでも多くの議論が交わされており、評価の実際にも変化が見られる（例えば「相対評価」と「絶対評価」、指導要録、内申書、通知表など）。しかし評価が教育活動の不可欠で重要な部分であることは議論の余地がないものであるかのように思われる。

しかし本当にそうなのであろうか。教育に評価が必要であるというのは、実は「根拠のない常識」なのではないか。このような疑問を持った人物がいるのである。本稿は、評価それ自体の問題点を指摘してきた、アメリカの教育研究者である Alfie Kohn（1957年～）の「評価批判」を足がかりとして、教育評価の意義と課題を考察しようとするものである。

1 評価をめぐる Kohn の著作・論文について

Kohnとその著作については別稿で紹介を行った³ので、ここでは評価についての著作・論文について整理する。Kohnの評価批判の論点はいくつかあるが、一つは標準テスト（Standardized Test [s/ing]）への批判である。

この問題について単行本では *The Case Against Standardized Testing: Raising the Scores, Ruining the Schools*（Heinemann, 2000）が主題的に扱っている。この本は *The Schools Our Children Deserve: Moving Beyond Traditional Classrooms and “Tougher Standards”*（Houghton Mifflin, 1999年）を基にして再構成されたものである。本文60ページあまりのコンパクトな本で、問いと答えの形で議論が進められる。

また主な個別論文としては以下のものがある。⁴

- 1) “Confusing Harder with Better” *Education Week* September 15, 1999年
- 2) “Why Students Lose When ‘Tougher Standards’ Win” (interview) *Educational Leadership* September, 1999年
- 3) “Tests That Cheat Students” *New York Times* December 9, 1999年
- 4) “Sell Schools Not Test Scores” *Realtor Magazine* January, 2000年

- 5) “Standardized Testing and Its Victims” *Education Week* September 27, 2000年
- 6) “Fighting the Tests: A Practical Guide to Rescuing Our Schools” *Phi Delta Kappan* January, 2001年
- 7) “Two Cheers for an End to the SAT” *Chronicle of Higher Education* March 9, 2000年
- 8) “One-Size-Fits-All Education Doesn’t Work” *Boston Globe* June 10, 2001年
- 9) “Emphasis on Testing Leads to Sacrifices in Other Areas” *USA Today*, August 22, 2001年
- 10) “Beware of the Standards, Not Just the Tests” *Education Week* September 26, 2001年
- 11) “Accelerated Direct Success” (parody ad for a test-prep company) *English Journal* September, 2001年
- 12) “Standardized Testing: Separating Wheat Children from Chaff Children”
Foreword to *What Happened to Recess...?* by Susan Ohanian, 2002年
- 13) “Requesting Testing”, *Rethinking Schools*, Summer, 2002年
- 14) “The Worst Kind of Cheating”, Streamlined Seminar [NAESP], Winter, 2002年-03年
- 15) “Test Today Privatize Tomorrow: Using Accountability to “Reform” Public Schools to Death”
Phi Delta Kappan April, 2004年
- 16) “Debunking the Case for National Standards: One-Size-Fits-All Mandates & Their Dangers”
Education Week January 14, 2010年
- 17) Introduction to *More than a Score: The New Uprising Against High-Stakes Testing*
edited by J. Hagopian, 2014年
- 18) “What ‘No Child Left Behind’ Left Behind”, blog post, December 18, 2015年

発表された時期としては2000年～2004年頃が最も多いが、この期間はKohnが学校教育の様々な問題に対して積極的に発言を行っており、標準テストはその一つの柱であった。

同時にルーブリックによる評価や標準準拠評価 (standards-based grading) についても批判を行っている。前者については、The Trouble with Rubrics (*English Journal* March 2006年 vol. 95, no. 4)、後者についてはThe Case Against Grades (*Educational Leadership* November 2011年⁵⁾) が各々代表的なものである。

ただ後者のタイトルからも分かるように、Kohnの批判は標準テストやその他の個別の評価方法を超えて、評価 (grading)⁶⁾ 自体へと向かっているのである。個別論文としては例えばGrading: The Issue Is Not How but Why (*Educational Leadership* October 1994年)、From Degrading to De-Grading (*High School Magazine* March 1999年) などがあるが、この評価批判は、Kohnの二冊目の単行本であり、彼の主著の一つである*Punished by Rewards :The Trouble with Gold Stars, Incentive Plans, A’s, Praise, and Other Bribes* (Boston: Houghton Mifflin, 1993年, 1999年) で議論されているものである。副題にあるように「A」という成績が「賄賂」の一つであるのだ。以下、順次Kohnの議論を検討する。

2 Kohnの評価批判

2-1 標準テスト批判

先ずKohnの標準テスト批判を検討するが、その前に簡単にアメリカの標準テストについて見ておく。

標準テストは（standardized test）は（1）すべての受験者に同一の設問、あるいは共通の群から選択された設問に、同じ方法で解答することを求め、（2）「標準化された」つまり一貫した方法で採点され、個々の生徒や生徒集団の相対的な成績が比較できるようにされている、試験の総称である。⁷

つまり国や地域全体の子どもの能力を測定するために行われる共通試験であって、入学や卒業の際に用いられる場合や、学校の「教育成果」を確認する場合など、その目的は様々である。

アメリカで代表的なものはNAEP（National Assessment of Educational Progress 全米学力調査）である。1969年から実施されており、この結果は「The Nation's Report Card（国家の成績通知表）」と呼ばれ、教育政策立案の重要な情報源となっている。⁸ あり、また大学進学の際の標準テストとしてはSAT⁹とACT¹⁰が代表的である。また2006年から各州共通基礎スタンダード（Common Core State Standards）が導入され、小学校から高校までの教育水準の向上・統一が図られた。これによる共通テストも実施されている。1980年代以降、アメリカの公教育の「危機」が叫ばれ、NCLB（No Child Left Behind Act 2001年）¹¹とその後継のESSA（Every Student Succeeds Act 2015年）により「学力向上」が目指されている中で、標準テストについて多くの議論が交わされてきているのである。

例えばNCLBでは、Adequate Yearly Progress（AYP 「年間の学力向上の目標」「教育改善指標」などと訳される）が設定され、これが達成できない場合一定の改善策を取ることが義務づけられている。これに対しては困難な状況にある学校に「ペナルティー」を与え、そのような学校を「チャータースクール化」することで公教育を解体するものであるという批判も提起された。¹²

そのような状況を背景にKohnの主張のポイントを確認する。Kohnの立場はあくまでも教育・学習の側面から「（より厳しい）標準化 Tougher Standards」「説明責任 Accountability」を批判するものである。ここでは“The Case Against ‘Tougher Standards’”¹³によって、Kohnの標準テスト（そして標準化）批判の論点を紹介する。ここでは5点に整理している。

① 誤った動機づけとなる。

標準についての議論は、生徒が成績（performance）の向上を常に考えるべきであるという前提を持っている。結果のみに関心を持たせることは単純化しすぎである。結果がすべてであると見なすことは、どれだけ上手くやっているかに焦点をあてることと、行っている事柄に焦点をあてることとは異なるという心理学の研究を無視している。さらに、成績に気を取られることは、学習の質や難しい問題に挑戦したいという意欲を損なうことが多い。

② 誤った教え方につながる。

より厳しい標準化の主張者は「基礎に立ち戻れback-to-basics」という方針、より一般的には、子どもを受け身の物体であるかのように扱う教授法を好むが、それは「基礎技能」や「コアとなる知識」などの混合物を用意しそれを子どもに飲ませようとするものである。とりわけ州の標準の文書には、すべての生徒が一定の学年で習得することが求められている特定の事実や技能の長いリストが書かれている。これは時代遅れのモデルであると言えるだろう（もっともこのようなことが成功した時代などこれまでになかったのではあるが）。現在の認知科学はそれが常に十分なものではない理由を

より体系的に説明している。

③誤った評価が行われる。

「卓越性」「より高い標準」そして「規準を引き上げる raising the bar」はすべて標準テストの点数について語られているが、標準テストの多くは多肢選択式、集団準拠的であり、それ以外の欠陥も持っている。実際、現代の教育についての議論の多くは「テストの点数が低い、だから引き上げろ」というレベルに留まっている。このような試験方式の限界と問題によって、これらの試験に依存する学校改革の質に対する強い批判が出されることになる。

④誤った学校改革が行われる。

「より高い標準」の主張者は、何が教えられなければならないかを厳密に限定することによって、つまり特定の種類の教育を義務化することによって、改善を強要しようとする傾向を持っている。学校教育を変える方法が、単に教師と生徒に行いを変えることの要求であるとは考えられない理由は十分にある。「説明責任」は教室で起こる事に対して、そこにはいない人により強い統制力を持たせる符牒であることが通例である。それは首を絞める縄が息を塞ぐのと同じ効果を学習に対して与える。

⑤改善の方向が誤ったものとなる。

これらの考えを貫いているのは「厳格さ rigor」と「困難 challenge」についての暗黙の前提である。つまり「より難しいものが常により良い」というものである。試験、教科書、そして教師はすべて、難しさのレベルという単一の基準で判断することができるという還元主義的な（そして愚かな）考えが「レベルの低下している」教育への非難と、「基準を引き上げよ」という声高の要求の底流にあるのだ。それに付随して、物事（例えば価値の疑わしい宿題をするように生徒に求めること）があまりうまく行かなければ、もっと多くの同じことを要求することで問題は解決されるという考えもある。Harvey Danielsが述べているように、学校改善について今日主導的な哲学は、まとめれば「私たちのしていることはよいことである。それをより懸命に、より長く、より力強く、より声高に、より巧妙にするだけでよいのだ」と言うものである。

以上の中で、③の評価についてKohnは先に触れた *The Case Against Standardized Testing* において、「測定可能な成果は、最も重要度の低い学習の結果であり得る」¹⁴、「試験では正解が求められる。正解がないとは言えないとしても、正解が必ずしも理解を示すものでもなく、逆に誤答が理解不足を示すものでもない。標準テストは学生が正解に達するプロセスを無視しており、単純な計算間違いと理解不足とを区別できない」¹⁵、「正解とされる選択肢は明らかに正しいと分かるものであるが、現実の社会の問題は決してそのようなものではない」¹⁶と指摘している。さらに標準テストの成績は社会経済的地位 (socioeconomic status SES) によって大きな影響を受け、問題に貧困層の生徒に不利になるようなバイアスがあるとも述べている。¹⁷

以上のようにKohnは「より高い標準」を求める動きと、それによる標準テストの用いられ方を批判する。それに対して彼が提案する試験・評価は、数値ではなく文章による記述 (narratives)、パフォーマンス評価、とりわけ「ポートフォリオ」によるものである。¹⁸ ただこれらが無条件で望ましいとする訳ではない。画一的な標準テストよりは優れてはいても、Kohnにとっては評価それ自体が問題であるからである。

2-2 ルーブリック評価批判

次にルーブリック評価についてKohnの見解を検討する。

一般的にルーブリック評価については、例えば以下のように説明されている。

米国で開発された学修評価の基準の作成方法であり、評価水準である「尺度」と、尺度を満たした場合の「特徴の記述」で構成される。記述により達成水準等が明確化されることにより、他の手段では困難な、パフォーマンス等の定性的な評価に向くとされ、評価者・被評価者の認識の共有、複数の評価者による評価の標準化等のメリットがある。¹⁹

日本の授業での事例であるが、例えば高校の古典の授業（『蜻蛉日記』）において、「和歌に込められた登場人物の心情を理解することができる。（読む能力）」と「実演を経て、和歌の解釈を深めることができる。（関心・意欲・態度）」という二つの尺度（観点）を設定する。

そして前者については以下の評価規準（特徴の記述）によって評価を行う。²⁰

- A（あるいは3） 和歌の解釈として、作者の、それまでのいきさつを正しく踏まえた兼家に対する感情を、和歌に用いられた表現に絡めて述べることができる。
- B（あるいは2） 和歌の解釈として、作者の、それまでのいきさつを正しく踏まえた兼家に対する感情を述べることができる。
- C（あるいは1） 和歌の解釈として、作者の、兼家に対する感情を述べてはいるが、それまでのいきさつを正しく理解できない。または、感情を述べていない。

後者については以下の通りである。

- A（あるいは3） 実演を見て、和歌の解釈に沿って登場人物の心情理解を深めた解釈の書き直しをすることができる。
- B（あるいは2） 実演を見て、和歌の解釈の書き直しをすることができる。
- C（あるいは1） 和歌の解釈をすることはできるが、実演を見ても解釈を書き直すことができない。または、解釈ができない。

この例は一時間の授業についてであるが、ある教科の年間指導、あるいは学校教育全体についてのルーブリックも作成されている。

「用語集」でも述べられているように、ルーブリック評価は点数化することの難しい実技やパフォーマンス（レポートやプレゼンテーションなど）を比較的客観的に評価するのに有効であるとされている。

このようなルーブリック評価に対してKohnは以下のように述べる。²¹

「ルーブリックの魅力は、評価者の間の信頼性（interrater reliability）にあり、特に国語（language arts）で有効性を発揮する。エッセイを評価する際に最高の点数が与えられるべきものの規準（criteria）のリストはルーブリックを作成した人たちのほぼ総意を反映し、どのエッセイがそれらの規準に合うかを考える（つまり、決定するというよりは発見する）ためにリストを活用するすべての人の手助けになると想定されている。

（判断規準が曖昧で教師の裁量次第であると批判する人もいるが、）私はむしろルーブリックが上手くいく場合を心配している。標準テストの点数を上げるために、教育課程を解体して（gut）、学校をテストの準備工場にしてしまうように、ある課題に対してどのような成績をつけるかについて何人か

の意見を一致させるためには、自ら進んで、誰か他人が作った、その成績にふさわしい狭い規準を受け入れて適用することが必要である。私たちが一旦自分自身の判断を放棄すれば、まったく同一のものに4（最高点）を与えることになるのである。」²²

つまり、教師が与えられたルーブリックによる評価を行うことで、ルーブリックが評価の基準としてだけでなく、「教育の質の判定者」や「教えられ評価されることを支配する存在（agents of control）」として機能するようになるのである。

もう一つKohnが指摘するのは、生徒が事前にルーブリックの内容を知ることの問題である。ルーブリックを嫌うある生徒の言葉が次のように引用されている。「もし何かうまくできないと、あなたは何をすべきだったのか分かっているね、と先生が言える。」またある教師が生徒に「この課題にはルーブリックがあるのですか？」と尋ねられたことに対して、その教師が「彼らは、課題の表にすべての必要事項が明示され、点数が与えられていなければ学ぶことができず、さらに自らの思考力や文章力に自信が持てず、敢えて挑戦をしようとしな」とコメントしたことを紹介している。²³

そして次のように指摘する。「答案用紙の上に書かれたB+という評価は、答案の質について何も教えてくれないが、ルーブリックは複数の規準によるより詳しい情報を与えてくれる点で、素晴らしい評価法である、と評価の専門家（その多くは技術者）は言うかもしれない。しかしこの論理の決定的欠陥は、自分たちがどのくらいうまくできるかを絶えず気にしている生徒は、自らが取り組んでいる対象への関心を失ってしまう、という教育心理学の知見から明らかである。自分が読んでいる物語の内容を考えることと、自分の読解力について考えることは大きな違いがあるのだ。」²⁴「学ぶこと（learning）」と「成績を取る（結果を得る）こと（achieving）」とは異なることであり、評価の価値がどれだけ多くの情報を与えるかと直接的に関わっていると考えるのは誤りであるとKohnは述べる。

さらに、エッセイや文章をルーブリックという細分化された規準で評価することは、評価から「良い文章に生気を吹き込む複雑さ」を奪ってしまうという専門家の言葉を引用している。

2-3 標準準拠評価批判

標準準拠評価（Standards-Based Grading）は、前もって明示されている複数の目標がどの程度達成されたのかを評価しようとするものである。Kohnは標準準拠評価をそれ自体として批判している訳ではない。その傍証となるのは例えば、この評価方法を提唱している *Rethinking Grading: Meaningful Assessment for Standards-Based Learning*（ASCD 2015年）を著したCathy Vattero が、この本の中で自らに影響を与えた人物として第一にKohnを挙げていることである。Kohnが *The Case Against Grades* の中で述べているのは、この方法では十分でないということである。すなわち以下のように指摘しているのである。

「標準準拠評価を用いるのは十分ではない。この言葉は多くのことを示唆するように思われる。例えば、成績を決めるときの一貫性の向上、つまりより精巧な方式を活用すること、評価される課題や技能の数の増加である。これらは、良くて、評価の基本的問題に対処することとは無関係であり、最悪の場合は問題を拡大する。より多くのデータを得ることが常に良いという単純化された前提に加えて、学習は要素に分割でき、そして分割されるべきであって、各々が別々に評価されるものであるというかつての行動主義の信奉者と同じ傾向が見られるからである。そして成績をつける頻度が増えることで、（学習を犠牲にして）成績に強すぎる関心を持つようになるが、それは逆効果であると研

究者が認めている。」

「標準準拠評価は目標に準拠した評価（criterion-based testing）と類似のものとされることもあり、その意図は分布曲線に沿った評価を避けようとするものである。（中略）確かにこれは最高点の数を非常に少なくして、生徒を互いに競わせる状況からの改善ではある。しかしこれはタマネギの外の皮（競争）をむいて、内側のより害悪をもたらす皮を見つけるようなものである。それはすなわち外発的動機づけ、点数評価、学習を犠牲にして成績を向上させようとする傾向、である。」²⁵

このようにKohnは、ループリックや標準準拠評価を批判した上で、評価自体の問題へと進むのである。

2-4 評価批判

Kohnは評価自体への批判を繰り返しているが、基本的には*Punished by Rewards*で展開された議論が基になっている。ここでは評価を報酬（reward 褒め言葉、良い成績、賞品／賞品など）の一つの典型として捉え、報酬についての問題が論じられている。Kohnの報酬への批判の論点は大きく5つある。²⁶

第一は報酬が罰になり得るという点である。これは報酬によって他人（上司や教師や親）からコントロールされるということと、もし期待した報酬が得られなかった場合、それは罰を受けるのと同じであるということの意味する。

第二は報酬が人間関係を壊すという点である。これにも二つの側面があり、報酬を与えられる者同士の水平的関係と、報酬を与える者と与えられる者との垂直的關係がある。水平的関係については、（すべての人が報酬を得られるのではない以上）他人は障害物と見なされることになる。垂直的關係については、報酬を与える側が常に優位に立ち、場合によっては「懲罰者 punisher」ともなることで、協力的・協調的なつながりが持てないことになる。

第三は報酬が行動の背後にある理由を無視するという点である。例えば夜寝る時間になっても寝室から出てくる子どもに、報酬や罰で対応する（「ベッドに入ったら明日テディ・ベアを買ってあげるよ」「寝なければ明日からテレビを見せないよ」）ことは、そもそも何故子どもが寝室から出てくるのかを考えないことになる。行動の背後にある理由・原因を考えずに、行動をコントロールしようとするのが報酬（あるいは罰）なのである。

第四は報酬が危険を冒す（risk-taking）姿勢を妨げるという点である。報酬を得るために働く（勉強する）のであれば、求められたこと、あるいは必要なことだけをしておけばよいと考えるようになるのである。敢えて難しいことに挑戦をして失敗をすることで報酬を失うことは無駄なことだと見なされる。

第五は、そしてこれが最大の問題であるが、報酬が内発的動機づけ（intrinsic motivation）を失わせるという点である。報酬は基本的に外発的動機づけ（extrinsic motivation）であり、それを与えることで、行動の内容への関心度が低くなるのである。

以上をKohnは次のようにまとめている。つまり、「これをすれば、あれがもらえる。（Do this and you'll get that.）」と言われたら、人は「これ」ではなく「あれ」に関心を持つようになる。「報酬は動機づけとなるか？もちろん動機づけとなる。報酬を得たいという動機づけになるのだ。」

2-5 まとめ

Kohnは成績評価を報酬（あるいは罰）の一つと見なし、教師が生徒をコントロールする手段であると考えている（同時に行政が教師や学校をコントロールする回路でもある）。どのような方法を採用したとしても、基本的な問題は同じであり、外発的動機づけを与えることで内発的動機づけを弱め、学習内容への興味や関心を損ない、生徒同士の競争意識を高め協力関係を持たなくさせる、などの弊害があるとしているのである。

Kohnは著作の中で「doing to sb（一方的に行う）」のではなく「working with sb（ともに行う）」ことの重要性を繰り返し説いている。彼にとっては評価も「doing to」の一つなのである。教師や親がすべきことは、assessmentやgradingではなく、学習のためのfeedbackである。

例えばブログ記事“Progressive Labels for Regressive Practices: How Key Terms in Education Have Been Co-opted”（2015年1月）²⁷の中で以下のように述べている。

「形成的評価formative assessmentの目的が、生徒が次のテストでどれだけよくできるかを見るものになってしまい、生徒が自分の関心のある問題について深く考えることを促すようなフィードバックを与えるものになっていない。」（なおブログ記事“Why the Best Teachers Don't Give Tests”2014年10月30日²⁸では“kidwatching”というYetta Goodmanによる造語「子どもの読解力を知るために一緒に読むこと」を紹介している。²⁹）

教師や親が生徒・子どもとともに学び、学習の内容に関心を持ち考えを深めていけるようにすることこそ必要なことなのである。

おわりに

現在様々な「評価」が行われている。学校の「教員評価」「学校評価」、ネット上でのオークション出品者や飲食店などへの評価、職場での「人事評価」、国や自治体の「政策・行政評価」など、何らかの形で評価を行ったり、されたりする場面に出会う。ただ議論は「どのように評価をするのか」という方法論に集中しがちで、そもそもの目的やその問題点についてはあまり意識されていない。それは「評価自体は必要で、有効である」という前提があるからであろう。

本稿で取りあげたAlfie Kohnは、評価を含めた「賞罰」の問題に正面から取り組んでいる人物である。ここでは議論のごく一部しか取り上げることができなかった。またもちろんKohnの議論に対する批判も数多く存在する。³⁰ これらを踏まえて、評価の問題は何であるのか、教育においてどのように考えていけばよいのか、今後とも検討を続けたい。

なおKohnは*The Case Against Standardized Testing*の中で日本について触れている。それは「日本の小学校教育の水準は高く、それは教師が標準テストの圧力から自由であることが一因である」³¹ というものである。全国学力・学習状況調査はこの本が刊行された7年後（2007年）に始まったが、これをKohnはどう見るであろうか。³²

注

- 1 文部科学省初等中等教育局教職員課（2017）『暫定版 教職課程認定申請の手引き（教員の免許状授与の所要資格を得させるための大学の課程認定申請の手引き）（平成31年度開設用）【再課程認定】』p.91
- 2 『小学校学習指導要領解説 総則編』（2017年6月）pp.92-94 なお中学校学習指導要領でも同趣旨の記述がなされている。
- 3 「現代教育界の思想を振り返る— :Alfie Kohnの『子育て論・家庭教育論』批判をめぐる—」昭和女子大学近代文化研究所『学苑』917号（2017年3月）
- 4 <http://www.alfiekohn.org/essays-standards-testing/>（2017年8月27日参照）
- 5 後にJoe Bower and P.L. Thomas ed. *Detesting de-grading schools* (Peter Lang 2013年) の第10章として再録されている。本稿での引用はこちらによる。
- 6 gradingあるいはgradeを「評価」と訳してよいのかどうかは問題である。日本の学校の用語では「評定」の方が適切と思われる。
- 7 <http://edglossary.org/standardized-test/>（2017年7月23日参照）
- 8 末藤美津子「アメリカにおける学力調査の位置づけと役割—NAEP、TIMSS、PISAに注目して—」（比較教育学研究 第40号 2010年 130頁）
- 9 元はScholastic Aptitude Test（大学適性試験）の略語であったが、1990年にScholastic Assessment Testと改称され、現在では略語とはされていない。
- 10 元はAmerican College Testingの略語であった。
- 11 1965年に制定されたElementary and Secondary Education Act（ESEA 初等中等教育法）は5年毎に見直し（reauthorization）が行われることになっており、2000年の見直しによるものがNCLBと呼ばれる。次のESSAも同様である。
- 12 土屋恵司「2001年初等中等教育改正法（NCLB法）の施行状況と問題点」（『外国の立法』No.227, 2006, 2, pp.129-136. <<http://www.ndl.go.jp/jp/diet/publication/legis/227/022707.pdf>> 2017年8月6日参照）
- 13 <http://www.alfiekohn.org/standards-and-testing/case-tougher-standards/>（2017年8月10日参照）
- 14 *The Case Against Standardized Testing* p.4
- 15 *ibid.* pp.7-8
- 16 *ibid.* p.8
- 17 *ibid.* pp.20, 35
- 18 *ibid.* pp.41-42
- 19 中央教育審議会答申「新たな未来を築くための大学教育の質的転換に向けて～生涯学び続け、主体的に考える力を育成する大学へ～」の「用語集」（2012年8月）
- 20 中央教育審議会教育課程部会総則・評価特別部会「学習評価に関する資料」2016年1月
- 21 *The Trouble with Rubrics* (*English Journal* March 2006 — vol.95, no.4)
なお、本稿での引用は<http://www.alfiekohn.org/article/trouble-rubrics/?print=pdf>（2017年9月4日参照）による。引用元のpdfファイルにはページ数がないため引用者でp.1～p.5を付した。
- 22 *The Trouble with Rubrics* p.2
- 23 *ibid.* pp.2-3
- 24 *ibid.* p.3

- 25 The Case Against Grades p.148
- 26 以下の5点についての説明は、*Punished by Rewards* (pp.49-76) の要約である。翻訳は引用者によるものである。
- 27 <http://www.alfiekohn.org/blogs/regressive/> (2017年8月29日参照)
- 28 <http://www.alfiekohn.org/blogs/no-tests/> (2017年8月29日参照)
- 29 同趣旨の内容は Valerie Strauss が Washington Post 紙電子版 (2014年10月30日) で引用している Kohn のブログ記事 “The problem with tests that are not standardized” でも述べられている。
https://www.washingtonpost.com/news/answer-sheet/wp/2014/10/31/the-problem-with-tests-that-are-not-standardized/?utm_term=.23d76e6cd34e (2017年8月10日参照) なおこのブログ記事は Kohn のサイトには掲載されていない。
- 30 例えばルーブリックについては以下の反論がある。Michael Livingston The Infamy of Grading Rubrics *The English Journal* Vol.102 No.2 pp.108. 113 (2012年12月) Livingston は Kohn のルーブリック批判が、成績一般に対する批判と混同されており、ルーブリックの有効性が無視されていると反論している。
- 31 *The Case Against Standardized Testing* p.62 ここではかつて日本の教師たちが、「全国中学校一斉学力調査」の導入に反対したことにも触れられている。
- 32 なお本稿では触れられなかったが、以下の書籍も試験をめぐるアメリカの現状を知る上で参考になる。
Anya Kamenetz *The Test: Why Our Schools Are Obsessed with Standardized Testing—But You Don't Have to Be* (Public Affairs 2015年)