

〔研究論文〕

スタンダード準拠評価論の成立と新たな展開

—ロイス・サドラーの所論に焦点を合わせて—

石 田 智 敬

(京都大学大学院 日本学術振興会特別研究員)

はじめに

第一章 スタンダード準拠評価論(1980年代)の成立とその背景

- (1) ラドフォード制度から ROSBA 制度へ
- (2) ROSBA におけるスタンダード準拠評価の考え方とその特徴
- (3) スタンダード準拠評価論の理論的土台
- (4) サドラーの抱えていたジレンマと彼の所論の受容

第二章 スタンダード準拠評価論の新たな地平

- (1) ルーブリック批判における二つの論点：分析性と成文化への疑い
- (2) スタンダードの外的表現とそれによる質的判断の調整
- (3) スタンダード準拠評価論における変化とその示唆

おわりに

はじめに

教育の焦点は、事実に知識や個別のスキルの習得から、認知的に複雑なコンピテンスをどう育成できるのかに移行してきた。このような能力は、知識の統合、複雑な問題解決、批判的思考などに要求されるもので、高次の認知能力、または、単に高次の学力と呼ばれている⁽¹⁾。高次の学力の質的レベル(熟達の度合い)は、主体的に知識やスキルを統合するようなパフォーマンスを要求する複雑な課題を通じて質的に評価される。そこでの成果物は、正・誤で評価することはできず、その質的レベルは素朴なものから洗練されたものまでグラデーションとして現れる。その際、複雑に絡み合ったクライテリア群が質的判断に用いられるため、総括的判断を導くアルゴリズムを定式化することは、一般に困難である。

伝統的に、複雑なパフォーマンスや作品(work)の質的レベルは、権威ある評価者の絶対的判断によって評価されてきた。しかし、権威による絶対的判断が有する不透明性や主観性への疑念から、評価における透明性や客観性が歴史的に要求されてきた。このような背景のもと台頭しているアプローチは、ルーブリックと呼ばれるスタンダード(水準)とクライテリア(規準)のマトリックスに準拠して、質的判断を行うことである⁽²⁾。本稿では、このような言語表現に基づく評価基準表(マトリックス)をルーブリックと広義に呼称することとする。

ルーブリックを用いることは、外的に固定された評価枠組み(スタンダードとクライテリア)に準拠した教師の質的判断を可能にする。ゆえに、ルーブリックのような明示的な評価基準表を用いることは、質の評価における透明性や客観性、妥当性や信頼性を高め、評価を公正にすると考えられている。外化された評価枠組みに準拠して成果物全体の質的レベルを教師の質的判断によって評価しようという、このような考え方の理論的淵源の一つは、豪州の教育評価学者であるロイス・サドラー(D. Royce Sadler)⁽³⁾が1987年に提起した「スタンダード準拠評価(standard referenced assessment)」論にあるとされる⁽⁴⁾。

これまで、サドラーが提示したスタンダード準拠評価の考え方は、1987年の論文の内容を中心として、C. ギップス (C. Gipps)、国内では鈴木秀幸や石井英真らによって紹介されてきた⁽⁵⁾。彼らは、サドラーの所論を、ドメイン準拠評価と称される J. ポファム (J. Popham) らの所論と対比しながら、高次の学力の熟達度合いを評価する最適な方法を提起するものとして捉え、その特徴や意義を浮き彫りにした。しかし、近年サドラーは、ルーブリック等に準拠して質的判断を下すという考え方に対し、ラディカルな批判を展開するようになっていく⁽⁶⁾。こうした状況は、サドラーの所論の理解や受容に何らかの問題があった、または、彼の考え方に何らかの変化が生じたことを示唆している。いずれにせよこのことは、彼の所論全体や、近年展開されている批判の論理について注意深く検討する必要性を提起している。なお、上述の先行研究では、スタンダード準拠評価 (1987年) が提唱された背景や文脈、そして、2000年代後半以降に展開される彼の批判的論調は、基本的に扱われていない。

そこで本稿は、こうした状況について説明するため、(1) サドラーがどのような背景や文脈のもと、どのようなスタンダード準拠評価論 (1987年) を提起し、その考え方がどう受容されたのか、(2) 2000年代後半以降に、サドラーは、どのようなルーブリック批判を展開し、どのようなオルタナティブを提起していったのか明らかにすることを目的とする⁽⁷⁾。この試みを通じて、スタンダード準拠評価における論点を析出し、さらなる展望への示唆を得たい⁽⁸⁾。

留意すべき点として、サドラーは、2000年に教授 (高等教育) に異動しており、以降の所論は、上述の批判的論調を含め、高等教育を主な文脈として発表されているということがある。けれども、そこでの議論は、彼も言及するように、必ずしも高等教育に限定されるものではない⁽⁹⁾。彼の論考は一貫して、複雑で高次の学習の文脈における評価の方法論に向けられており、複雑で高次の学習は高等教育を特徴付けるものであるが、その他の教育段階でも重視されているものである。

第一章 スタンダード準拠評価論 (1980年代) の成立とその背景

スタンダード準拠評価の考え方は、サドラーが豪州クイーンズランド州における教育評価制度改革 ROSBA (Review of School-Based Assessment) に取り組む中で提唱されたものであり、それを学術的に論述したのが論文 Specifying and Promulgating Achievement Standards (1987年) である。本章では、スタンダード準拠評価が提起された文脈である ROSBA 改革の展開を跡付けながら、どのような評価論が構想されたのかを明らかにする。そして、その理論的土台となった彼の博士論文に遡り、彼の所論のどの側面が、スタンダード準拠評価論として、どのように具現化されたのかを描き出す。

(1) ラドフォード制度から ROSBA 制度へ

初めに、ROSBA 制度導入に至るまでの流れを簡単に概観しておこう。19世紀以来、豪州クイーンズランドでは、中等教育の修了認定と大学入学の資格認定は、公的試験と呼ばれる外部試験の成績に依拠してきた⁽¹⁰⁾。こうした外部試験は、1970年のラドフォード制度 (Radford scheme) 導入に伴い廃止され、学校ベース (school-based) の相対評価制が導入された。こうした学校ベースの評価制度は、豪州全土においても初めての試みであった。ただし、制度転換から数年後、早くも相対評価に伴う深刻な諸問題が顕在化するようになる。このような背景のもと、新たに導入されたのが ROSBA 制度である。

ROSBA 改革の主眼は、学校ベースの評価制度を維持しつつも、相対評価を改めてクライテリオン準拠評価 (criterion referenced assessment) を導入することに置かれた。1980年、政府は ROSBA 制度への移行を開始しようとしたが、クイーンズランド教員組合 (Queensland Teachers Union) の強い反発もあり、その移行は円滑には進まなかった。というのも、ROSBA への移行に際しては、政治的判断が先行して

いたため、ROSBA 制度が実際にどのようなもので、どのように機能するものなのかといった具体的様相は全く不明瞭なままであり、よって、学校現場は混乱に陥っていたのである。

こうした混迷を背景に、クイーンズランド政府は、ROSBA 制度への移行を支援するための追加予算を提供する旨を表明し、この政府予算によって中等教育委員会に評価部門 (assessment unit) が設置された。この部門には、ROSBA 制度の理論的な基盤と制度運用に関する実施方針を確立させる役割が求められていた。そして、この部門長 (director) に就任したのが、サドラーである。

彼を筆頭とする評価部門は、1985年から87年にかけて、ROSBA 制度における評価の基本的な考え方や、制度実施に関する基本的な枠組みを描き出す「検討報告書 (discussion paper)」を計21本公表した⁽¹¹⁾。この検討報告書で提示されたのが、スタンダード準拠評価の考え方である。その内の約半数にあたる10本を執筆したサドラーは、ROSBA における評価理論とその実践の核となる考え方を描き出し、他のメンバーは、それを補完する周辺の議論を提供した。これらの検討報告書は、ROSBA 実施に多大な影響を与え、以後、クイーンズランド州の評価システムを実質的に形成するものになったと指摘されている⁽¹²⁾。

(2) ROSBA におけるスタンダード準拠評価の考え方とその特徴

ROSBA においてサドラーが示したスタンダード準拠評価の考え方はどのようなものだったのか、以下にその内実を見ていこう。当時、彼が示したスタンダード準拠評価の考え方は、クライテリオン準拠評価の中に位置づくものであるが、その米国における解釈・発展形態 (クライテリオン準拠テスト) とは全く異なるものであった⁽¹³⁾。なお、ここでいうクライテリオン準拠評価とは、相対評価のように他の生徒と比較するのではなく、設定された評価枠組みに照らし合わせて、生徒を評価する試みを包括的に指すものである。彼は、米国でのクライテリオン準拠評価へのアプローチには、以下の特徴があったと指摘する。それは、行動目標 (objectives) の明細を作成し、その明細事項の習得可否を確認する客観テストを実施し、それらの正答数を尺度として評定を行うというものである。こうしたアプローチは、客観性が重視され、分割・合算という測定的手順が踏まれるという意味で要素主義的な性格を持つ。彼は、こうしたアプローチに対し「人の直接的な質的判断が生徒の作品の質を最も良く評価できる多くの科目で、不適切である。客観テストへの没頭は、真正なものに対する代理的・間接的な尺度への代用を推奨してしまう」と批判した⁽¹⁴⁾。そこで、こうしたアプローチに対するオルタナティブとして提唱されたのが、高次で複雑な学習成果をも評価の射程に捉えるスタンダード準拠評価の考え方である。

スタンダード準拠評価では、力量ある教師の質的判断 (qualitative judgement) がその中心的役割を果たす。要は、生徒の評定を行う際に、パフォーマンスや作品に対する教師の質的判断が直接的に用いられるのである。スタンダード準拠評価では、パフォーマンスや作品の全体的な質が重視されるという点で、クライテリオン準拠テストとは判断における粒度 (granularity) が異なる。彼は、スタンダード準拠評価の中心にある質的判断の特徴を、表1のように説明する。こうした特徴は、すべての質的判断が主

表1. 質的判断の5つの特徴

- ・質的判断で用いられるクライテリアの少なくとも一部は、シャープでなくファジーである。
- ・質的判断では、それが正しいかどうかを確認する独立した方法はない。
- ・質的判断では、クライテリア同士が連動するため、全体的構成が重要となる。
- ・質的判断では、使用される可能性がある無数の潜在的クライテリア群が背後に存在するが、一度の評価ではその内のいくつか少数部分のみが適用される。
- ・質的判断の後にのみ、数値的表現は割り当てられる。

Sadler, D. R. (1986b). *Discussion papers 5: Subjectivity, objectivity, and teachers' qualitative judgments*, pp.21-22. を元に著者作成。

観的であることを意味する。では、教師の質的判断はどのような条件の下で、独善的とならず妥当で信頼できるものとなるのか。彼は、「明確な評価枠組みの中で評価が行われる場合、適切に訓練された教師は、一貫して信頼でき妥当な評価を行う、[中略] 教師の質的判断は、優れた教育と優れた評価の中心にあり続けている。[中略] 判断を下す人が『キャリブレート [基準合わせ] (calibrate)』されているという条件で、質的判断を信頼することは合理的である」と指摘した⁽¹⁵⁾。つまり、教師間で主観性が達成されている場合、教師らの質的判断はある程度一致し、評価者間信頼性を保証できるというわけである。

ROSBA 制度では、学校ベースで評価が行われるものの、修了資格の認証は州の中央当局機関によって行われる。したがって、各学校は評価に関して一定の自由を享受するものの、その成績評価が州内の他の学校と十分に比較可能である必要があった。そこで、彼は学校間における比較可能性を達成するために、州内で共通のスタンダードを構成し、教師の質的判断をその共通枠組みに準拠して行えるように調整することを提案した⁽¹⁶⁾。また、このような共通の評価枠組みが構成され公表される場合、教師や学校だけでなく、専門家集団の外部者、つまり、保護者や生徒らもこの評価枠組みにアクセスできるようになるという利点もあった。

そこで、スタンダードを外化し共有化する方法として提案されたのが、作品事例 (exemplar) によって補完された言語記述 (verbal description) による評価枠組みである⁽¹⁷⁾。この評価枠組みにおける評価基準表は、教科ごとにいくつかのクライテリアが指定され、それぞれのクライテリアに対して数段階のスタンダード・レベルが設定されるものである。スタンダード準拠評価における教師の質的判断は、このような作品事例に補完されたスタンダード記述に準拠して行われる。彼は、スタンダードが明示される必要性を、表2のように指摘した。このような方法が適切に機能するためには、教師がこうした評価枠組みに準拠して評価を行えるように、質的判断を洗練・調整していくことが肝となる。なお、教師が経験を蓄積しこの枠組みに慣れていくと、スタンダードは徐々に教師に内化されていくが、外化されたスタンダード記述が最終的・決定的な評価枠組みを構成するとされた。

表2. 評価枠組みが明示されるべき理由

- ・明快さは、より明晰な思考を可能にする。
- ・クライテリアとスタンダードは、より客観的に機能する。
- ・保護者に対して評価判断を正当化する際に役立つ。
- ・目標が明確となり、評価に対する生徒の困惑が少なくなる。
- ・生徒は、教師ではなく、スタンダードによって評価される。
- ・教師の経験不足を補う上で役立つ。

Sadler, D. R. (1986c). *Discussion papers 2: The case for explicitly stated standards*, pp.6-7. を元に著者作成。

以上、ROSBA 制度では、教師の質的判断を評価の中心に据えつつも、以上のように外化された共通評価枠組みを構成することで、学校間・教師間の比較可能性を一定達成し、質的判断の妥当性と信頼性を公的な資格認証レベルまで引き上げることが目指された。こうした評価の考え方の大枠は、以後細かな修正が行われつつも、現行の制度に至るまで続いていくものとなる。

(3) スタンダード準拠評価論の理論的土台

ROSBA においてサドラーが提示したスタンダード準拠評価の考え方はどこからきたものか。簡潔に言えば、こうした考え方は、彼の博士論文における評価論をその理論的土台として構想された。

計12本の学術論文から構成された博士論文 *Theory and Practice in Educational Evaluation: a Methodical Inquiry* (1985年) において、彼の教育評価論は一定の着地点を示す⁽¹⁸⁾。この博士論文における試みの概要やその背景は、概ね次の通りである。1970年代以前の教育評価研究と実践では、明快さ、客観性、効率性に特徴付けられる教育測定や行動目標を軸としたアプローチ、すなわち工業生産的のパラダイムが支配的であった。ただし、1970年代よりそれに対する批判が顕在化し、人類学や社会学にルーツを持つ新

しい評価の考え方がオルタナティブとして登場していった。例えば、M. パーレット (M. Parlett) や D. ハミルトン (D. Hamilton) の所論を引き合いに出しながら、彼はこうした立場に対して、評価における価値判断を躊躇する傾向があるために非判断的 (non-judgmental) だと指摘した。要は、彼は、工業生産的パラダイムを拒否しつつも、評価を「説明」や「記述」として再解釈する非判断的な評価論の動向に歩調を合わせようとはしなかったのである。彼は、評価はプログラムやカリキュラムであろうと生徒の作品であろうと、その価値の判断であるとし、価値判断としての評価の立場を堅持する⁽¹⁹⁾。そこで彼は、両立場に対するオルタナティブとして人の質的判断を中心に据えた評価パラダイムの可能性に焦点を合わせ、その評価原理や方法論について思弁的に論究し記述した。こうした人の質的判断を中心とする評価のパラダイムが、まさに ROSBA において描き出されたスタンダード準拠評価論の根幹となったのである。

博士論文の中で、とりわけスタンダード準拠評価論に関連深い箇所は、スタンダードを外的に表現し共有化する方法論について論じた「第4章：スタンダード」である。彼は、スタンダードは概して四つの方法：数値打切スコア (numerical cutting score)、評価経験 (evaluative experience)、暗黙的定義 (implicit definition)、明示的定義 (explicit definition) で共有化 (promulgate) されるとした⁽²⁰⁾。スタンダード準拠評価では、この内の後者二つの方法を組み合わせることによって、評価枠組みが構築された。

まず、暗黙的定義による方法は、あるレベルの質または能力の典型的な事例の集合体によって、スタンダードを示そうとするものである。この場合、あるレベルと他のレベルの区別を成文化する試みは行われない。ただ、多様なクライテリアが複雑に絡み合う場合、少数の作品事例ではスタンダードを適切に定義できず多大な作品事例が必要となる。

次に、明示的定義による方法は、あるスタンダード・レベルを正確に言葉で記述するという形式をとるものである。これは、スタンダードを客観化する試みであり、評価者の主観的判断に依存することからの解放を意味する。ただし、こうした記述が可能かどうかは、スタンダードを構成するクライテリアの性質に依存する。ファジーなクライテリアは、実在物によってのみ定義されるため、それを含む場合、完全な明示的定義は構成できないとした。

ここでいうファジーなクライテリアとは、ある状態から別の状態へのグラデーションを内含するものである。また、実在物の指示対象を理解の前提としており、それらがないと理解できない。すなわち、これらは帰納的に認識されるものである。例えば、一貫性・説得性・明快さ等のクライテリアがこれに当たる。これに対し、シャープなクライテリアは、ある状態から別の状態への非連続的な跳躍を含み持ち、合意された解釈体系 (例えば、数学の公理系) を前提とするものを指す。この最も単純な形は、正・誤の二分法で表現される。

彼は、以上の検討を踏まえ、暗黙的定義と明示的定義の両者を補完的に用いることが、スタンダードを外化する上で、最も効果的な方法であるとした。要は、ある質レベルのスタンダード記述が作品事例に付随する場合、事例の数を減らすことができ、また、抽象的なスタンダードの言語表現は、解釈されるために実在物の指示対象を必要とするというわけである。こうしたスタンダードの外化 (共有化) の方法論が、ROSBA の枠組みにおいて具現化されたのである。

(4) サドラーの抱えていたジレンマと彼の所論の受容

ただ、実のところサドラーは、部門長に就任する前の1982年に、ROSBA の試みは失敗する可能性が高いとの指摘をしていた⁽²¹⁾。その際、論拠としたのは、先に挙げたファジーなクライテリアの存在である。教育におけるスタンダードを構成する際に、ファジーなクライテリアを軽んじたり排除したりする

ことはできないものの、これによってスタンダードを言語で定義することが不可能になると指摘していたのである。当時、彼は、この問題を解決する唯一の方法が、作品事例と鑑識眼（connoisseurship）の考え方を拡張していくことにありと結論づけていた。以上のように、スタンダードを言語記述で定義することの困難さに関しては、その後の博士論文でも、1987年の論文でも、その論証に多大な頁数が割かれている。こうした彼の姿勢からは、作品事例による補完という解決案を導きつつも、スタンダードの言語表現による定義に、一貫して慎重であり続けていたことを窺い知ることができる。

加えて注目すべきことは、博士論文における彼のクライテリアに対する考え方である。彼は、クライテリアの本質的な根源は、行われた価値判断を正当化し説明する過程にあるとした。つまり、認識することが第一義的な評価行為であり、如何なるクライテリアにも先立つというのである（一度顕在化したクライテリアは、以後の判断において規範性を帯びる）。また、ディスコースの世界は、経験の世界と同程度の広がりをも有していないため、全てのクライテリアを明示することはできないとした。以上より、彼は「教育評価におけるクライテリアは一般的に明示されるべきであるが、完全に事前指定することは原理的に望ましくない」と結論づけた⁽²²⁾。要は、目標として組織されたクライテリアは、^{もくろみ}目論見として機能するが、そのみで評価を行うことは自己制限的だという訳である。なお、スタンダードやクライテリアに対するこのような彼の捉え方は、言語による知の表現・記述の限界性を論じた M. ポランニー（M. Polanyi）の認識論からの影響を受けるものである。事実、彼は自身の教育評価論に最も影響を与えた人物として、ポランニーの名をあげている⁽²³⁾。

以上のように、ROSBA 導入期前後のサドラーの所論からは、クライテリアやスタンダードは明示されるべきだが、それを完全に行うことはできないというジレンマを抱えていたことが窺える。実際のところ、彼自身、ROSBA は困難な取り組みであったと回顧している⁽²⁴⁾。

では、以上に見てきたスタンダード準拠評価の考え方は、以後の教育評価研究や実践に対して、どのような影響を与え、どのように評価されていったのか。こうしたサドラーの所論が注目されたのは、ギブズがその著書 *Beyond Testing*（1994年）で、サドラーのスタンダード準拠評価論を新たな教育評価のパラダイムとして紹介したことの影響が大きい⁽²⁵⁾。その後、サドラーの示したスタンダード準拠評価論やそれに類する考え方は、スタンダード記述に照らし合わせた教師の質的判断を重視すること、そして、スタンダードとクライテリアのマトリックス（いわゆるループリック）を開発していく世界的動向において重要な役割を果たしていったと指摘されている⁽²⁶⁾。ただし、こうした動向は彼の提案によって直接的に引き起こされたというよりも、そのような歴史的動態の中の一つに彼の試みがあったものと解釈する方が妥当であろう。以上、第一章では、ROSBA 改革や博士論文での議論を参照しながら、彼が提唱したスタンダード準拠評価論（1987年）の内実について明らかにした。

第二章 スタンダード準拠評価論の新たな地平

本章では、2000年代後半以降に展開されるサドラーの所論に焦点を合わせる。2000年代に入ると、1980年代とは異なり、クライテリアとスタンダードの言語記述による評価基準表（いわゆるループリック）が、単一課題に対するものからカリキュラムに対するものまで、教育評価の世界において台頭していく。ところがサドラーは、こうした評価枠組みに対してラディカルな批判を行い、1980年代（第一章）とは異なる形のスタンダード準拠評価を提唱していく。

なお、こうした批判の背景として、豪州ではループリックが採点装置として強い規範性と厳格性を有しているということがある。彼によれば、豪州をはじめ英語圏の大学では、一般的にループリックを用いた成績評価が求められており、義務化されている場合すらあるという。また豪州では、高校の成績評

定によって大学入学選抜が行われるため、成績評定における高度な学校間比較可能性が要求される。そこで、導入されているのが中央当局機関 QCAA (Queensland Curriculum and Assessment Authority) による「承認 (endorsement)」と「確認 (confirmation)」と呼ばれる手続きである⁽²⁷⁾。各学校での評価課題とその採点指針 (ルーブリック等) は、QCAA による事前の承認と事後の確認をパスする必要がある。このように、ルーブリックは厳格な採点装置という性格を帯びているのである。では、彼はどのような視角や論理でルーブリックを批判し、どのようなオルタナティブを提案しているのだろうか。第一章における彼の所論と対比させながら以下に見ていこう。

(1) ルーブリック批判における二つの論点：分析性と成文化への疑い

サドラーは、どのような論理でルーブリックを批判するのか。彼の批判は、概して二つの論点から展開される。第一の論点は、ルーブリック等で示される事前設定 (pre-set) クライテリアに基づいて分析的評価を行うことの是非についてである。第二の論点は、言葉によるスタンダードの成文化 (codify) は可能なのかという点である。まず、前者について見ていこう。

一般的に、全体的判断を扱いやすい部分に分割することは、評価の客観性や透明性を高めると考えられており、事前設定クライテリアによる評価枠組み (分析的ルーブリック) が幅広く支持されている。しかし、彼はこうした評価方法は、複雑な作品を評価する上で不適切であり、この方法が有する欠陥は、この方法に固有だとする⁽²⁸⁾。分析的ルーブリックによる評価手順は一見容易に理解され、合理的でシンプルに思われるが、その実際のプロセスは複雑である。彼は、分析的評価を行う際に以下の現象が出現すると指摘する⁽²⁹⁾。評価者は、複雑な作品の評価を行う際、全体的な構造や様相をホリスティックに捉えたり、各クライテリアに焦点化したりしながら作品の質を見定めていく。この時、全体的な質が最高だと判断されても、各クライテリアが最高だと判断されない時がある。同様に、全てのクライテリアが良い評価であっても全体的な質はそれほど良くない時がある。これが、「分析的・全体的評価の不一致」という問題である。また、評価を行う際に、事前に指定されていないクライテリアが浮上することがある。これが「創成クライテリア (emergent criteria)」の出現である。この場合、創成クライテリアを無視するか、それを超規則的に採用することになる。以上の問題に加え、指定されたクライテリア群に概念的な重複があったり、クライテリアの意味が一意に定まらなかったりする問題が生じる場合もある。

こうした現象は、なぜ出現するのだろうか。これらの問題は、第一に、明示されたクライテリア群が、あくまで部分的に選択されたものであること、第二に、一部のクライテリアがファジーであるということによって現れる⁽³⁰⁾。そして、こうした諸現象の存在は、事前設定された明示的クライテリアのみで評価を行うこと、またクライテリアごとに判断を行ってそれらを合算することの限界性を示唆している。ルーブリック等で示されたクライテリアに基づいて評価が行われる場合、評価される作品群に対して同一のクライテリア群が等しく適用される。これは一見、公平・公正であるかのように思われるが、実際のところ、画一的な適用は不適切なものであり、歪んだ評価判断を導く可能性があるというわけである。

なお、1980年代の所論では、創成クライテリアの出現など、既存の評価枠組みで適切な評価判断が行えない場合に、つまり「ルーブリックの破れ」⁽³¹⁾が明らかとなった時に、適宜、評価者が潜在的なクライテリアを持ち出すことは、正当に認められるべき行為であるとしていた⁽³²⁾。ただし、2000年代以降は、ルーブリックという採点装置の厳格性や規範性が高まっている背景もあり、「それを認めることは、指定されたクライテリアだけが使用されるという教師と学生間の暗黙的契約に違反する」と指摘している⁽³³⁾。

次に、スタンダードを言葉で定義すること、すなわち成文化は可能かという第二の論点について見ていこう。教育のスタンダードを成文化する試みは、課題に対するものからカリキュラムに対するものま

で、世界的に広がっている。しかし、彼は、言語によるスタンダードの成文化は本質的に不可能であるため、スタンダードを適切に外化し共有化するためには、根本的に異なる表現形式が必要であると主張している⁽³⁴⁾。なぜなら、スタンダード記述は直感的に理解できるように思われるものの、記述文で用いられる語句や修飾詞は、具体的な文脈無くして解釈することはできず、よって決定的な意味を持つことはないからである。

一般的にスタンダード記述は、相対的な修飾語を用いるために、相対主義に陥りがちであると彼は批判する。「スタンダード記述の修飾語は、絶対的ではなく相対的に解釈される。ループリックの場合、修飾語が意味することはコホートに限定されうる」⁽³⁵⁾。例えば、要求される質の水準は明確に異なるのにもかかわらず、同じループリックが、学部1年、4年、大学院生の作品に等しく適用できてしまう場合がある。このようなループリックでは、評価におけるクライテリアは指定できても、質の水準、つまりスタンダードは固定できていないのである。

なお、1987年の論文では、記述文を作品事例によって補完することでこうした問題は解決可能であるとしていた。しかし、彼はこの考え方を後に否定する。彼は、「スタンダード記述とその関連する作品事例の組み合わせには、二段階の解釈が必要である。一般的に、スタンダード記述は特定の作品を正確に説明することはできない。[中略]教育におけるスタンダード記述は、あるレベルの一般的な特徴の記述なのである。作品事例によっては、質の水準を正しく判断するのに重要な側面がスタンダード記述に全く含まれていない」として、スタンダード記述は、あくまで特定レベルの一般的な特徴の記述であるため、それを補完するいくつかの作品事例とは、一対一で対応するものではないと指摘する⁽³⁶⁾。「認識論的理由から、スタンダードを純粹に命題的・宣言的形式で表現することはしばしば不可能である」⁽³⁷⁾。

ただ彼は、スタンダードを成文化する試みは、その過程において確かに利点もあると指摘する。彼によると、スタンダードを成文化することは「カリキュラムの計画、学術プログラムの開発、指導、評価を導くための貴重な手段として確かに機能する。その策定過程における協議プロセスは、学術的価値としての共有されたスタンダード概念に対する合意とコミットメントを促進し、学術的内容と統合的で高次な成果の追求に貢献する。けれども、それ以上の利益をもたらすことはなく、スタンダードの成文化は学術スタンダードを保証することも、学生の成績判断において高い比較可能性をもたらすこともできない」とした⁽³⁸⁾。

以上、二つの論点からのループリック批判の論理は、彼が1980年代の所論で提案した、質的判断を間主観化する装置としての「スタンダード記述とそれを補完する作品事例」という考え方を否定する。それでも彼は、スタンダードの外的表現は必要なのかという問いに、はっきりと必要であると答える⁽³⁹⁾。つまり、1980年代とは違った形式での外的表現を要求するのである。それでは彼は、どのような形態の外的表現を提案しているのか。

(2) スタンダードの外的表現とそれによる質的判断の調整

サドラーによれば、教育におけるスタンダードの概念は「共同体アカデミーの財産であり、個々の教師らによって個人的に決定または保持されることはない。大学教師は、スタンダードの開発に参加し[中略]それらを適用する方法を知るための専門的な責任を負う」⁽⁴⁰⁾。スタンダードは、合意に基づいて適切に外化され、評価者間で共有知として保持されることが望ましく、この場合、質的判断は間主観的となり一貫して信頼できるようになるという。

彼によれば、教育におけるスタンダードを評価者間で共有化するには、四つの要素—作品事例、論拠説明 (explanation)、ディスコース、暗黙知—の組み合わせが必要になる⁽⁴¹⁾。このうち、はじめの二つ

がスタンダードの外的表現を構成する。この外的表現は、あるスタンダード・レベルを満たす作品事例の集合によって構成され、それぞれの作品事例には、その判断に対する論拠説明が付加されている。各々の作品の論拠説明は、同一のクライテリア群が画一的に適用されるのではなく、それぞれの作品を評価するのに最も適切なクライテリアが用いられて記述される。ここで重要なことは、スタンダードを定義する出発点が言葉による論拠説明ではなく、具体物である作品事例にあることである。そして、論拠説明で用いられる言葉が、指示対象としての作品事例と一対一の対応関係を形成していることである。なお、ループリック等におけるスタンダード記述は、ある質レベルの一般的な特徴の記述であるため、その記述とそれぞれの作品事例が一対一で対応しているわけではない。

こうした外的表現を規範として、残る二つの要素によってスタンダードはギルド、つまり、教師という専門職集団によって保持される。第三の要素は、作品事例とそれに付随する論拠説明に関連するディスコースである。これは、評価における共通語彙を確立し、曖昧さを最小限に抑えて評価者間のコミュニケーションを可能にするものである。最後の要素は暗黙知であり、これは経験的に人々が発達させ共有するものである。暗黙知は、その定義により、表現することが不可能であるが、鑑識 (connoisseur) はそれに依存する。

彼は、このようなギルド知識を形成し、比較可能性を達成するため、上述したような外化されたスタンダードを用いて評価者の質的判断をキャリブレーション (calibration) することを重視する⁽⁴²⁾。キャリブレーションとは、外化された評価枠組みによって教師の質的判断を調整するプロセスを指す。彼は、計器の調整に擬えながら、次のように説明する。物理的計器は、絶対的な基準 (例えば、kg 原器) に照らし合わせ適宜調整される。教育の文脈では、外化されたスタンダード (作品事例集) がそのような原器 (標準器) となる⁽⁴³⁾。彼によれば、高品質の計量器は、正確な測定値を何度も生成するため信頼できるが、正確な値が無限に生成されるとは想定されていない。よって計量器は、標準化された原器により定期的に、または必要に応じてキャリブレーションされる。「評価者は、合意されたスタンダードに対する評価の責任を受け入れ、[中略] 定期的、しかし常時的ではない再キャリブレーションを行うことで、意思決定の大部分を独立して実行できるようになる」⁽⁴⁴⁾。

(3) スタンダード準拠評価論における変化とその示唆

サドラーの所論における決定的な変化はどこにあり、それが示唆することは何か。彼が示したスタンダード準拠評価の考え方は一貫して、間主観化された教師の質的判断をその中心に据えるものであった。また、質の固定的なレベルを指すスタンダードを外化しようとする点が一貫していた。スタンダードの外的表現が評価者をキャリブレーションする装置となるのである。ただし、外的表現の形態については明確な変化が見られた。スタンダード記述 (あるスタンダード・レベルの一般的な特徴の記述) とそれを補完する作品事例という形式だったものが、あるスタンダード・レベルの作品事例集 (論拠説明が付加されたもの) という形式に改められた。つまり、言語表現と具体物の両者を補完的に用いることでスタンダードを外化するという点は通底するが、そのどちらが決定的にスタンダードを定めるのかという点が改められたのである。新たに提起された外化形式は、言語表現ではなく具体的な作品事例を定義の出発点に据えるもので、事例主義的にスタンダードを定義しようとする志向性が現れてきていると言える。

では、こうした変化は何を意味し、何を可能にするのか。一般的に、ループリック等におけるスタンダード記述では、その特性上、あるレベルの作品を共通に特徴付けるクライテリア (共通項) に焦点が向けられる。けれども、共通的に出現するクライテリアが、必ずしも各作品の評価判断において最も重要なクライテリアではなく、このことは最も注目すべきクライテリアへの意識を埋没させ、評価に歪み

を生じさせうる。これに対して、論拠説明の記述では、それぞれの作品を評価するのに最もふさわしいクライテリアが用いられる。要は、このような外的表現の形式では、作品事例と言語表現の間に一対一の対応が形成されるため、その作品の質を判断する際にどのようなクライテリアに焦点が向けられるべきか、そして、抽象的な言語記述が意味することが何かがより明確に定まるのである。

この変化は、どのようにスタンダードを外化すれば、より適切に評価者をキャリブレーションできるのかその装置の形態を模索した結果であると言える。新たな方法では、質についての一般的な叙述であるスタンダード記述を読解することではなく、論拠説明が付された具体的な作品事例への知覚と認識が中心に据えられることで、より確かなキャリブレーションが可能になると考えられたのである。彼のこうした試みは、評価行為における権威（authority）をループリック等における言語記述から、その元になっている実際の作品事例へと立ち戻らせることを求めるものであると言えよう⁽⁴⁵⁾。

こうした事例主義的な方法には、それがコンパクトでなく可搬性に優れていないため、あまり現実的ではないとの指摘が向けられがちである。ただし、ループリック等の場合でも、それを適切に解釈するためには、記述文を補完するいくつかの作品事例は必要なのである。ところが、ループリックのようにスタンダード・レベルが宣言的形式で提示されると、ある種の利便性から作品事例が伴わず、記述文それのみが一人歩きしてしまうことが少なくない。彼の新たな提案では、このようにスタンダード記述のみが一人歩きし、その記述文をただ当てはめていくような評価行為が横行することを抑制しようとする意図があるものとも捉えられる。

作品事例を中心にギルド的にスタンダードを保持しようとする場合—ループリックの場合も同様だが—、専門家集団としての教師らに各事例を見極める一定の鑑識眼が要求される。けれども、全ての教師にそれなりの水準の鑑識眼を要求することは容易ではない。彼は、より確かなキャリブレーションを可能にする方法論を上述のように見出したが、事例主義的なスタンダードの外的表現によって教師の質的判断が精緻に調整でき、問主観化できることを過信しているようにも思われる（十分な議論が提供されているとは言えない）。彼の提案は、評価者が十分な鑑識眼を備えている専門教育や、特定の分野領域で特に親和性が高いように思われるが、全ての教育段階や分野領域で、有効かつ効果的に機能するとは言いきれない。また、彼はスタンダードを外化することに妥協しないが、明文化が行われない事例主義的な方法では、ギルド集団以外からのスタンダードへのアクセスが比較的難しくなるという意味で一定の閉鎖性が生じる。以上の課題点については、さらなる検討が求められると言えよう。

おわりに

サドラーは、力量ある教師の質的判断によってパフォーマンスや作品の全体的な質の水準を見取る評価、すなわちスタンダード準拠評価をどう構想するかを、1980年代から現在に至るまで論究し続けてきた。本稿では、このような彼の所論に即して、(1) 彼がどのような文脈においてスタンダード準拠評価論（1987年）を提唱し、(2) 2000年代後半以後にどのような批判を展開し、どのようなオルタナティブを提唱したのかを明らかにしてきた。

教育における評価が、教師の専門的力量に依存せずに、どんな人がどんな時でも同じ結果を導くことができることの追求によって、定式化・分割化されたアルゴリズムへと矮小化され、結果として評価に歪みが生じていくことを危惧した点で、彼の批判は80年代以来通底している。彼の教育評価論では、ホリスティックな質的判断を、機械的なアルゴリズムに定式化せずに、評価の中心に位置付けることが目指されてきたのである。彼のスタンダード準拠評価論で一貫して問われていることは、人の質的判断をどう洗練させ調整するか、そして、その過程や装置の形態をどのように構想するかということである。

ホリスティックな質的判断では、作品のゲシュタルト全体としての質を、ある種直観的に認識する行為が肝となる。ただし、直観的と言っても、これは質的判断を評価者の気まぐれに任せることを意味していない。それは、教師というギルド集団の中で堅牢にキャリアレーション（間主観化）されたものである。ホリスティックな質的判断は、具体物であるさまざまな作品事例への意識的、無意識的な参照・比較を通して行われる。ホリスティックな質を認識することの重視が、作品事例を中心に据えるという新たなアプローチを導いたのである。

このように、スタンダード準拠評価という考え方は、ルーブリックという評価基準表ではなく、具象に対する人間の質的判断や解釈をその根底原理とするものである。「ルーブリック評価」という言葉に代表されるような、ルーブリックがまずありきで、その記述をただ単に当てはめていくような評価は主客転倒というわけである。けれども、ルーブリックが強い拘束性や規範性を帯びてくると、その両者の関係が逆転していく危険性を孕む。彼の主張は、このような傾向性に対する強い警鐘として受け止められるだろう。

日本でも、昨今、高次の学力の評価方法として、ルーブリック等を用いた評価実践が、研究・政策・実践レベルで大きく注目されている。学校現場レベルでは、それが標準的というほどに普及しているとは言えないものの、ルーブリック等を用いた評価実践は今後さらに広がっていくと想定される。彼の所論は、高次の学力を対象とする評価実践の原理的側面に対して深い示唆を提供するものであり、彼のルーブリックへの警笛は、今後ルーブリックの普及が見込まれる中でそれが生じさせる諸問題を先見的に示すものとして示唆的である。彼の問題提起や提案、それらが示唆することを踏まえ、今後我々は、高次の学力の評価方法をどのように構想すべきか、ルーブリックとどう向き合うべきか、その可能性と限界はどこにあるか、さらにはポスト・ルーブリックの評価はどう構想されるかという問いに答えていく必要があるだろう。本稿は、その萌芽的な試みの一つである。

〔注〕

- (1) 石井英真『今求められる学力と学びとは』日本標準、2015年。
- (2) 本稿では、スタンダードとは、“あるもの”がどの程度優れているかという水準やレベルのことを指す。一方、クライテリアとは、“あるもの”が評価される時の価値次元や観点を指す。
- (3) 高校や工科大での数学教師を経て、1973年クイーンズランド大学講師。1985年同大より博士（教育学）。1991年グリフィス大学教授、2000年同大高等教育研究所所長。2010年クイーンズランド大学名誉教授。2019年、日本カリキュラム学会第30回記念大会（京都大学）にて基調講演を行う。
- (4) Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13 (2), pp.191-209. 石井英真「第1章：教育評価の立場」西岡加名恵・石井英真・田中耕治編『新しい教育評価入門』有斐閣、2015年、pp.45-46。
- (5) Gipps, C. (1994). *Beyond Testing: Towards a theory of educational assessment*. Routledge. 鈴木秀幸『スタンダード準拠評価』図書文化、2013年。石井英真、同上論文。
- (6) Sadler, D. R. (2009a). Indeterminacy in the use of preset criteria for assessment and grading in higher education. *Assessment and Evaluation in Higher Education*, 34, pp.159-179. Sadler, D. R. (2014). The futility of attempting to codify academic achievement standards. *Higher Education*, 67 (3), pp.273-288.
- (7) 本稿の目的を達成するため、1980年以降現在に至るまでのサドラーの著作物を包括的にレビューし、2019年2月に彼の元を訪問し、論文に書かれていない背景や意図に関するインタビューを計4日間実施した。なお、Personal Communication としての引用は、全て本人からの許諾を得ている。

- (8) 彼の所論は、形成的アセスメント論の視点からも参照・検討されてきた。ただし、本稿は、スタンダード準拠評価論、すなわち、教師の質的判断を重視する評価の方法原理に関する論点に焦点化するものである。彼の形成的アセスメント論に関する議論については別稿で記したい。
- (9) この点に関して、サドラーは、「私自身の見解と希望は、学習評価について私が言ったこと、考えたこと、書いたことは、幼児から博士に至るまで、すべての学習の段階で適用可能であるということです」(Personal Communication, 8th, January 2021 by email) と答えている。
- (10) 以下、クイーンズランドにおける教育評価制度の史的展開については、Clarke, E. (1987). *Assessment in Queensland secondary schools: Two decades of change 1964-1983*, Department of Educationを参考にした。
- (11) Assessment Unit of the Board of Secondary School Studies (1986-1987). *Discussion papers on assessment No.1-21*. なお、以下に本文献を引用する際は、各報告書の著者名、タイトル、出版年のみを付す。
- (12) Maxwell, G. S., & Cumming, J. J. (2011). Managing without public examinations: Successful and sustained curriculum and assessment reform in Queensland. in Yates, L., Collins, C., & O'Connor, K. ed., *Australia's curriculum dilemmas: State perspectives and changing times*, Melbourne Univ. Publishing, p.193.
- (13) Sadler, D. R. (1986a). *Discussion papers 1: ROSBA's Family Connections*, pp.1-5.
- (14) Sadler, D. R. (1987)., p.192.
- (15) Sadler, D. R. (1986b). *Discussion papers 5: Subjectivity, objectivity, and teachers' qualitative judgments*, p.23.
- (16) *Ibid.*, pp.24-25.
- (17) Sadler, D. R. (1986c). *Discussion papers 4: Defining Achievement Levels*, pp.14-19. McMeniman, M. (1986). *Discussion papers 3: A Standards Schema*, pp.11-13.
- (18) Sadler, D. R. (1985a). *Theory and Practice of Educational Evaluation: a methodological inquiry* (unpublished doctoral dissertation), the University of Queensland.
- (19) 以上の内容については、*Ibid*, pp.3-12. を参照。
- (20) *Ibid*, pp.66-85. なお、数値打切スコアによる方法とは、一次元尺度における閾値の指定である。一方、評価経験による方法とは、評価活動への徒弟的な共同的参画によるものである。
- (21) Sadler, D. R. (1982). Why the explicit definition of standards is not always possible?. *Ideas in Education*, 1 (2), pp.12-13.
- (22) Sadler, D. R. (1985b). The origins and functions of evaluative criteria. *Educational Theory*, 35 (3), p.296. なお本論考は、博士論文「第3章：クライテリア」の元となった投稿論文である。
- (23) Personal Communications, 21st, February 2019 at the University of Queensland.
- (24) *Ibid*.
- (25) Gipps, C. (1994).
- (26) Maxwell, G. S., & Cumming, J. J. (2011)., p.193.
- (27) 詳しくは、QCAAの公式webサイトを参照。<https://www.qcaa.qld.edu.au/senior/certificates-and-qualifications/qce-qcia-handbook/9-internal-assessment-qa>. [2020年9月29日最終確認]
- (28) Sadler, D. R. (2009a)., p.159.
- (29) *Ibid*, pp.164-168.
- (30) *Ibid*, pp.168-172.
- (31) 筆者による造語。分析的・全体的評価の不一致や、創成クライテリアが出現するなどして、既存の評価枠組み(ループブリック)に則ると適切な評価が行えないことが明らかとなった状況を指す。

この現象は、ルーブリックが、あくまで評価知の部分的で不完全な表現物（representation）であることによって生じる。

- (32) Sadler, D. R. (1983). Evaluation and the improvement of academic learning. *Journal of Higher Education*, 54, pp.72-73.
- (33) Sadler, D. R. (2009b). Transforming holistic assessment and grading into a vehicle for complex learning. in Joughin G. ed., *Assessment, learning and judgement in higher education*, Springer, p.54.
- (34) Sadler, D. R. (2014).
- (35) Ibid, p.281.
- (36) Ibid, p.282.
- (37) Sadler, D. R. (2009c). Grade integrity and the representation of academic achievement. *Studies in Higher Education*, 34, p.820.
- (38) Sadler, D. R. (2014)., p.287. ここでの彼の指摘は、以下のことを示唆している。ルーブリックの作成過程に参画した教師らは、その過程で評価知をお互いに共有し合意を形成することで、質的判断の洗練や調整の利益を享受できる。ただ、結果として生み出されたルーブリックの記述それ自体が、質保証の装置として有効に機能するのは難しい。というのも、ルーブリックの記述それのみでは、質的判断を洗練したり調整したりすることはできないからである。ルーブリックの記述をただ外部から与えられた評価者は、ルーブリックを飼い慣らせず、むしろルーブリックに飼い慣らされてしまう危険性を孕む。こうした事態は、ルーブリックを運用するための背景や文脈に関する知識、そして、ルーブリック自体を問い直すような知識を十分に持ち合わせない場合に生じてしまう。
- (39) Sadler, D. R. (2014)., pp.283-284.
- (40) Sadler, D. R. (2009c)., p.819.
- (41) Ibid, p.821.
- (42) Sadler, D. R. (2013). Assuring academic achievement standards: From moderation to calibration. *Assessment in Education: Principles, Policy and Practice*, 20, pp.5-19.
- (43) なお、外化されたスタンダード（作品事例群）は、標準器に擬えられるものとはいえ、完全に不変なものではなく、コミュニティの同意のもと必要に応じて適宜改定されるものである。
- (44) Sadler, D. R. (2013)., pp.17-18.
- (45) 第二章で扱った彼のオルタナティブの提案は、豪州の現状に対して一石を投じるようなものであり、それが既に豪州の教育実践において広く一般的に受け入れられているというわけではない。

A Formulation and New Developments of Standard-referenced Assessment: A Critical Review on the Works of Royce Sadler

Tomohiro ISHIDA

(JSPS Research Fellow (DC1), Graduate School of Education, Kyoto University)

The emphasis in education has shifted from promoting the memorization of factual knowledge to developing cognitively complex competence. In general, higher-order cognitive capabilities are assessed through complex and divergent assessment tasks which require extended performance (works), actively integrating and synthesizing knowledge and skills. Such students' works have traditionally been evaluated by the absolute judgement of authoritative assessors. In recent years, however, in response to doubts for in-transparency and subjectivity in authoritative judgments, an approach to use a matrix of standards and criteria as a grading scheme (well known as a rubric) in teachers' qualitative judgments, has emerged and strongly endorsed. The use of an explicit grading scheme, such as a rubric, is considered to significantly increase the transparency, objectivity, validity and reliability of teachers' qualitative judgements; make grading process fairer. One of the theoretical foundations of this idea can be traced back to the idea of 'standards-referenced assessment' proposed by the assessment scholar D. Royce Sadler in 1987. Recently, however, Sadler has begun to develop radical criticism of the use of rubrics for teachers' qualitative judgements. To elucidate why this paradoxical situation has arisen, this article critically reviews the earlier (about 1980s) and later (about 2010s) works of Sadler regarding standards-referenced assessment and identifies the key essences of his proposals.

In the first chapter, it traced the historical development of the ROSBA reform in Queensland, which is the context wherein "standards-referenced assessment" was proposed and, in so doing, clarifies key ideas of Sadler's earlier works. In removing norm-referenced assessment, but without falling into criterion-referenced testing (elementalism measurement), Sadler envisaged standards-referenced assessment that emphasizes the calibrated qualitative judgements on the holistic quality of student works. To make teachers' qualitative judgements inter-subjective (achieve inter-rater reliability), he proposed a common assessment framework of verbal descriptions of standards, supplemented by exemplars of students' works, as a way to calibrate teachers' qualitative judgements. His propositions contributed significantly to the subsequent theorization and worldwide dissemination of a standards-criteria matrix assessment framework in teachers' qualitative judgements.

The second chapter focused on Sadler's later works. His criticism towards grading scheme such as rubrics is developed mainly from two perspectives: (1) problems in the use of pre-set criteria, and (2) problems in the codification of academic achievement standards. The logic of Sadler's two arguments rejects the ideas he proposed in the 1980s. However, he emphasized the importance of external representation, the material form of educational standard. Sadler modified external representation form of standards to a exemplars' collection of student's works satisfying a standard (with explanations of the evaluative judgement for each exemplar), rather than the verbal codification of standards (generic descriptions of a certain standard level) supported by exemplars. In other words, the complementary use of both verbal expression and concrete works to externalise educational standards is a commonality, but which of them is primarily decisive for defining standards was revised.

In conclusion, this paper showed that Sadler's revision to the form of standards expression was motivated by his concern that rubrics, which are an imperfect representation of teachers' evaluative knowledge in the first place, could potentially not only straightjacket but also distort teachers' professional judgements. Sadler's later works called for reclaiming evaluative authority from rubric as a verbal representation to the actual student's works. Sadler's scholarship has consistently explored ways to place holistic qualitative judgements at the centre of educational assessment without formulating them into simplistic, and mechanical algorithms as well as ways to calibrate teacher's qualitative judgments by the external representation of standards.