

パフォーマンス評価における教員の評価と 学生の自己評価・学生調査との関連[†]

斎藤有吾^{*1}・小野和宏^{*2}・松下佳代^{*1}

京都大学高等教育研究開発推進センター^{*1}・新潟大学歯学部^{*2}

近年、大学教育では、学習成果の直接的指標と間接的指標との関連が活発に議論されている。本研究では、ある歯学系のコースのパフォーマンス評価を事例とし、教員の評価と学生の自己評価（直接評価）と、学生の学生調査用アンケート項目への自己報告（間接評価）との関連を検討し、そこからそれぞれの評価が担うべき役割と射程を議論する。

キーワード：直接評価、間接評価、パフォーマンス評価、ルーブリック、自己評価

1. 問題と目的

1.1. 問題背景

米国の高等教育界において、学習成果における直接評価と間接評価に相関が見られるかという議論は古くからなされている（*e.g.*, ANAYA 1999, ASTIN 1993, PIKE 1996）。なお、直接評価とは学生の知識や技能などの表出から学習成果を直接的に評価すること（何ができるか）である。それに対し、間接評価とは学習行動や学習成果についての学生の自己報告から学習成果を間接的に評価すること（何ができるかと思っているか）である（PALOMBA and BANTA 1999, 松下 2012）。これら2つの評価の分類に関して論点になるのは、標準化されているかどうか、真正性があるかどうかや評価者が誰であるかといったことではなく、テストやレポート、発表など、その能力の表出が期待できるような提示をもとに評価しているかどうかである。学生調査などで多用されるアンケート形式の項目による評価はそのような提示を伴わないため、間接評価に該当する。一方、提示をもとに評価しているのであれば、教員の評価のみならず、学生自身の自己評価や学生同士のピア評価

も直接評価となる。そして、これら直接評価と間接評価の相関に関する議論は、近年、日本の大学教育においても活発になってきている。

例えば、大学教育学会の課題研究「学士課程教育における共通教育の質保証」（2013～2015年度）のサブテーマ3では、学習成果の直接評価と間接評価を併用する方策を模索するため、直接評価による評定結果と間接評価による評定結果の間にどの程度の相関が見られるのかを検討している（松下ほか 2014, 山田ほか 2015, 山田 2015）。

具体的には、大学生を対象にジェネリックな学習成果を評価するような小テスト（直接評価）と、学習成果の獲得に関して自己報告を求める学生調査用の項目（間接評価）の両方を盛り込んだ調査票を開発・実施し、それらの相関関係から、間接評価は一定の整合性がある評価指標として利用できるかと報告している（山田 2015）。ただし、ここで報告されている相関関係の解釈には慎重になる必要がある。有意と報告された直接評価と間接評価の相関係数は効果量でいえば高いとはいえず、10%に満たない分散説明率である。また有意な関連があったと報告されたのは複数の項目の一部の組み合わせである。

このような直接評価と間接評価の関連の度合いは、ANAYA (1999) の米国の調査においても共通しており、SAT（大学進学適性試験）やGRE（大学院進学の共通試験）といった直接評価と、学生の自己報告による間接評価との相関は弱く、やはり分散説明率は10%に満たない。もしこのような傾向が頑健なものであるならば、直接評価と間接評価は緩やかな関連を見せるものの、決して代替可能ではなく、「何ができるか」と「何ができるかと思っているか」という学習成果の2つの異なる

2016年4月4日受理

[†] Yugo SAITO^{*1}, Kazuhiro ONO^{*2}, and Kayo MATSUSHITA^{*1} : Correlations of Direct Measures Based on Performance Assessment and Indirect Measures Based on Student Self-report

^{*1} Center for the Promotion of Excellence in Higher Education, Kyoto University Yoshida Nihonmatsucho, Sakyo-ku, Kyoto City, 606-8501 Japan

^{*2} Faculty of Dentistry, Niigata University 2-5274, Gakkocho-dori, Chuo-ku, Niigata, 951-8514 Japan

側面をそれぞれ測定していると捉えることが妥当であろう。これは、テスト得点に代えて学生の自己報告を用いることは正当化できないという PIKE (1996) の指摘に合致する。

以上のように、直接評価と間接評価の関連を議論する際には、関連の有意性だけではなく強さ（効果量の大きさ）も考慮しながら、それぞれの評価や指標が担う役割や射程を慎重に解釈していく必要がある。

英米を中心とした「エビデンスに基づく教育 (evidence-based education)」を求める動きは昨今の日本でも顕著になってきており（石井 2015）、効果の検証にはなにかしらの評価手法によって得点化した量的指標が多く用いられることになると予想される。その際、当該の指標によって何が捉えられるのかという議論を置き去りにしたままでは、本来射程外とされるべきことにまで一般化の範囲を広げてしまう危険性ははらむと考えられる。例えばアクティブ・ラーニング型授業の効果検証として学生調査のような問題解決能力の自己報告を指標に用いた場合、それが「何ができるか」の代替として使用できると保証されない限り、解釈の範囲は「何ができているか」に対する効果に限定すべきであろう。

また、評価する能力が異なれば、直接評価と間接評価の関連の強さも異なってくる可能性も考えられる。さらに、同じ直接評価であっても、主観的評価であれば、エキスパートである教員の鑑識眼を通すか、ノービスの学生の鑑識眼を通すかで、その指標のもつ意味が異なってくる可能性も考えられる。このような議論を発展させるためには、さまざま評価手法から得られる指標の関連の強さに関する知見を多く積み重ねていく必要がある。まさに山田 (2015) の調査は、複数の評価を組み合わせた調査票によって、プログラムレベルの学習成果を多面的に捉えるための方策を提示し、今後の直接評価と間接評価の関連に関する議論をさらに発展させる知見を示しており、その点で意義があるといえる。

さて、山田 (2015) の調査はプログラムレベルにおける学習成果の直接評価と間接評価の関連に焦点を当てていたが、コース（授業）レベルではどのようなのだろうか。特に、昨今の高等教育において育成が求められている汎用的で高次の認知的能力に関して、教員の評価と学生の自己評価はどの程度の関連の強さをみせるのだろうか。また、それらコースレベルの直接的指標と、学生調査のようなプログラムレベルの学習成果に関する自己報告という間接的指標はどの程度の関連の強さをみせるのだろうか。本研究では、複数の評価から得られる指標同士の関連の議論に資するた

め、以上のことを特定の授業を事例として検討することを目的とする。

1.2. 目的

新潟大学歯学部の新学年次生を対象としたパフォーマンス評価の実践を事例として、パフォーマンス評価における教員の評価と学生の自己評価（どちらもコースレベルの直接評価）との関連、およびそれら直接評価とプログラムレベルの学習成果についての学生による自己報告（間接評価）との関連について、それぞれ相関係数から検討し、これらの異なる評価が担う役割と射程を議論する。

2. 方 法

2.1. 調査対象者と手続き

新潟大学歯学部において2015年度に開講された初年次教育科目「大学学習法」のパフォーマンス評価から得られたデータを使用する。調査対象者は「大学学習法」を履修した新潟大学歯学部の新学年次生59名である。データの収集・使用にあたり、新潟大学歯学部の倫理委員会の承認（承認番号 26-R54-03-11）を得たうえで、調査対象となる全ての学生から同意を得た。この科目はレポートの書き方やプレゼンテーションの仕方などについて講義・演習を行うものであった。

2.2. 直接評価（コースレベル）

学生には講義・演習で身につけた知識やスキルをもとにレポートを作成する課題（図1）が与えられた。そして学生のレポートから、アカデミック・ライティングにおける複合的な能力を捉えるため、「背景と問題」「主張と結論」「根拠と事実・データ」「対立意見の検討」「全体構成」「表現ルール」の6つの観点からなるライティング・ルーブリック（表1）を用いて教員4名で評価し、レベルに応じて得点に換算した。この4名は複数年この科目の指導と評価に関わっており、評価に先立ってルーブリックの記述語の共通理解を目指し、さらに前年度以前の学生のレポートのうち特徴的な3本を採点事例として共有するなどしてキャリブレーションを行っている。また、このルーブリックは、松下ほか (2013) で報告されたルーブリックをもとにしており、著者を含む担当教員と教育評価の専門家による信頼性と妥当性の検討を通して数回の修正が加え

図1 大学学習法 パフォーマンス課題

医療や科学の進歩にはめざましいものがあります。少し前までは治療することが難しかった病気にも対処できるようになりました。また、便利な道具も世の中に溢れており、日々、私たちは恩恵を受けて暮らしています。しかしその一方で、進歩がもたらしたさまざまな問題も抱え込んでしまいました。皆さんもテレビや新聞で見聞きしていることでしょうか。その解決のために、多くの議論がなされていますが、どの問題もしかるべき解決策が示されていないのが現状ではないでしょうか。

そこで、与えられたテーマ「医療や科学の進歩がもたらした諸問題」から話題を設定し、主体的に調査・学習を行い、自分の考えをレポートにして提出してください。

表1 大学学習法 ライティング・ルーブリック

観点	問題解決		論理的思考			文章表現
	背景と問題	主張と結論	根拠と事実・データ	対立意見の検討	全体構成	表現ルール
観点の説明	与えられたテーマから自分で問題を設定する。	設定した問題に対し、展開してきた自分の主張を関連づけながら、結論を導く。	自分の主張を支える根拠を述べ、根拠の真実性を立証する事実・データを明らかにする。	自分の主張と対立する意見を取り上げ、それに対して論駁(問題点の指摘)を行う。	問題の設定から結論にいたる過程を論理的に組み立て、表現する。	研究レポートとしてのルール・規範を守り、適した文章と正しい回しを用いてレポートを作成する。
	与えられたテーマから問題を設定し、論ずる意義も含め、その問題を取り上げた理由や背景について述べている。	設定した問題に対し、展開してきた自分の主張を関連づけながら、結論を導いている。結論は一般論にとどまらず、独自性を有している。	自分の主張の根拠が述べられており、かつ根拠の真実性を立証する信頼できる複数の事実・データが示されている。	自分の主張と対立するいくつかの意見を取り上げ、それらすべてに対して論駁(問題点の指摘)を行っている。	問題の設定から結論にいたる論理的な組み立て、記述の順序、パラグラフの接続が整っている。概要は本文の内容を的確に要約している。	・研究レポートとして適した文章と正しい回しを用いてレポートを書いている。 ・引用部分と自分の文章の区別を明示し、引用部分については、レポートの最後に出所を確認できる形で参考文献を記載している。 ・概要、本文ともに字数制限が守られている。
レベル2	与えられたテーマから問題を設定し、その問題を取り上げた理由や背景について述べている。	設定した問題に対し、展開してきた自分の主張を関連づけながら、結論を導いている。	自分の主張の根拠が述べられており、かつ根拠の真実性を立証する信頼できる事実・データが少なくとも一つ示されている。	自分の主張と対立する少なくとも一つの意見を取り上げ、それに対して論駁(問題点の指摘)を行っている。	問題の設定から結論にいたる論理的な組み立て、記述の順序、パラグラフの接続がおおむね整っている。	・概要、本文ともに字数制限が守られている。 ・3つの条件をすべて満たす場合は「レベル3」、2つの場合は「レベル2」、1つの場合は「レベル1」とする。>
	与えられたテーマから問題を設定しているが、その問題を取り上げた理由や背景の内容が不十分である。	結論は述べられているが、展開してきた自分の主張との関連づけが不十分である。	自分の主張の根拠は述べられているが、根拠の真実性を立証する信頼できる事実・データが明らかにされていない。	自分の主張と対立する意見を取り上げているが、それに対して論駁(問題点の指摘)がなされていない。	問題の設定から結論にいたるアウトラインはたどれるが、記述の順序やパラグラフの接続に難点のある箇所が散見される。	
レベル0	レベル1を満たさない場合はゼロを割り当てること。					

られたものである。このコースの詳細、ルーブリックの詳しい開発の経緯や理論的根拠、信頼性と妥当性の検討は松下ほか(2013)を参照されたい。

さらに、課題の提示と同時に、学生へのルーブリックの提示と評価基準の解説が行われた。そして、同じルーブリックを用いて、学生も自分のレポートを自己評価した。ここでは、教員4名の6観点の平均点を「アカデミック・ライティング能力(教員)」、学生の自己評価の6観点の平均点を「アカデミック・ライティング能力(自己)」と操作的に定義する。各観点の一貫性に関して、 ω 係数は「アカデミック・ライティング能力(教員)」が.76、「アカデミック・ライティング能力(自己)」が.70であった。

2.3. 間接評価(プログラムレベル)

大学教育学会の課題研究「学士課程教育における共通教育の質保証」サブテーマ3が開発した「新入学生学習調査2014(JFLS2014)」のうち、ジェネリックな学習成果を測定するための20項目を間接評価として使用した。「現在身につけている能力や知識」を、この授業に限らず問うものであった(「とても少ない」～「とても多い」の5件法)。これらの項目は、国際比較もできるように設計された日本版大学生調査研究プログラムJCIRP(Japanese Cooperative Institutional Research Program)をもとにしたものであり、すでに多くの大学・短期大学で実施され十年以上の実証的な研究蓄積があることに加え、現在も大学IRコンソーシアムの学生調査で利用されるなど、学習成果の間接評価として大きな影響力をもっている(cf. 山田 2012, 山田ほか

2015)。

3. 結果と考察

それぞれの指標の要約統計量と相関係数を表2に示す。まず、直接評価である「アカデミック・ライティング能力(教員)」と「アカデミック・ライティング能力(自己)」とでは、自己評価の方が高くなる傾向にあった($t(57)=14.07$, $g=2.54$, $p=.00$)。そしてその相関係数は、.076とほぼ無相関であった。つまり、絶対的にも相対的にも、教員の評価と学生の自己評価はずれていることになる。また、「アカデミック・ライティング能力(教員)」と間接評価であるJFLS2014の20項目との相関係数は-.231～.149であり、ほぼ無相関～弱い相関である。

それに対し、「アカデミック・ライティング能力(自己)」と、JFLS2014の「分析や問題解決能力」「批判的に考える能力」「文章表現能力」といったアカデミック・ライティングと直接関連すると考えられる項目との間には、小～中程度の正の相関関係が見られた。ただし、「リーダーシップの能力」「人間関係を構築する能力」「数理的な能力」といった、アカデミック・ライティングと直接関連すると考えにくい項目との間にも中程度の正の相関関係が見られており、学生側に起因する何かしらの第3の変数による疑似相関の可能性が示唆される。

これらのことから、教員4名の評価をエキスパートの鑑識眼を通したものとして「何ができるか」のより妥当な指標とした場合、それと学生の自己評価および学生調査のような自己報告との間に、代替可能性や整

表2 アカデミック・ライティング能力とJFLSの要約統計量と相関係数

変数名	平均値	SD	アカデミック・ライティング能力 (教員)との相関係数 [95%CI]	アカデミック・ライティング能力 (自己)との相関係数 [95%CI]
アカデミック・ライティング能力(教員)	1.32	0.29	—	—
アカデミック・ライティング能力(自己)	2.17	0.38	-.076	—
JFLS1 一般的な教養	3.14	0.80	-.050	-.122
JFLS2 分析や問題解決能力	2.97	0.95	.001	.464 **
JFLS3 専門分野や学科の知識	2.12	0.90	.117	.101
JFLS4 批判的に考える能力	3.09	1.00	-.018	.387 **
JFLS5 異文化の人々に関する知識	2.66	0.89	.110	.270 *
JFLS6 リーダーシップの能力	2.78	1.14	.107	.531 **
JFLS7 人間関係を構築する能力	3.48	0.96	-.070	.483 **
JFLS8 他の人と協力して物事を遂行する能力	3.67	0.85	.149	.328 *
JFLS9 異文化の人々と協力する能力	2.86	0.93	.049	.259
JFLS10 地域社会が直面する問題の理解	2.90	0.83	.057	.341 **
JFLS11 国民が直面する問題の理解	3.03	0.79	.092	.327 *
JFLS12 文章表現の能力	2.98	1.08	.130	.338 *
JFLS13 プレゼンテーションの能力	2.78	1.12	.027	.348 **
JFLS14 数理的な能力	3.02	0.85	.103	.486 **
JFLS15 コンピュータの操作能力	2.97	1.14	.094	.203
JFLS16 卒業後に就職するための準備の程度	2.62	1.06	-.049	.104
JFLS17 コミュニケーションの能力	3.48	1.01	.059	.344 **
JFLS18 時間を効率的に利用する能力	3.03	0.92	-.139	.241
JFLS19 グローバルな問題の理解	2.78	0.86	-.186	.378 **
JFLS20 外国語の能力	2.72	0.91	-.231	.379 **

** $p < .01$, * $p < .05$, CI = confidence interval

合性があるといえるような関連は見出せなかった。よって、それぞれの指標は異なる学習成果の側面を測定していると捉えるべきであろう。また関連が見出せなかった原因として、学生の鑑識眼の未熟さ（自己評価能力の低さ）の影響や、学生の自己認識を通すことにより生じるバイアスの交絡が考えられる。

4. 課題と展望

本研究の結果は、あくまで一事例の結果であることに加え、使用した評価手法は直接評価・間接評価ともに限定的なため、過度な一般化は控えるべきである。また仮に本研究で得られたような直接評価と間接評価の関連が頑健であったとしても、「何ができると思っているか」それ自体を学習成果と捉え、それを測る目的であれば、学生調査のような自己報告による間接評価は妥当な指標となりえるだろう。さらに、ルーブリックを用いたとしても教員の評価と学生の自己評価がずれることが明らかになったが、そのずれの度合いから学生の自己評価能力を捉え、それをフィードバックし、修正を促していくことも可能だろう。本研究は初年次生から得られた知見であるが、そのようなフィードバックを行うことで、年次が上がるにつれて学生の自己評価能力が高まり、教員の評価と学生の自己評価のずれが小さくなり、関連が強くなることも考えられる。また、その際には、直接評価と間接評価の関連の様相が変化するかもしれない。今後もこのような検討を重ね、さまざまな評価それぞれが担うべき役割や射程を明確にしていきたい。

参考文献

ANAYA, G. (1999) College Impact on Student Learning: Comparing the Use of Self-Reported Gains,

Standardized Test Scores, and College Grades. *Research in Higher Education*, 40(5) : 499-526

ASTIN, A.W. (1993) *What matters in college? : Four critical years revisited*. San Francisco : Jossey-Bass

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, 82(2) : 216-228

松下佳代 (2012) パフォーマンス評価による学習の質の評価—学習評価の構図の分析にもとづいて—。京都大学高等教育研究, 18 : 75-114

松下佳代, 小野和宏, 斎藤有吾, 白川優治 (2014) 学士課程教育における共通教育の質保証—直接評価と間接評価の開発と統合について—。大学教育学会誌, 36(2) : 17-21

松下佳代, 小野和宏, 高橋雄介 (2013) レポート評価におけるルーブリックの開発とその信頼性の検討。大学教育学会誌, 35(1) : 107-115

PALOMBA, C., and BANTA, T. (1999) *Assessment essentials: Planning, implementing, and improving assessment in higher education*. San Francisco : Jossey-Bass

PIKE, G. R. (1996) Limitations of using students' self-reports of academic development as proxies for traditional achievement measures. *Research in Higher Education*, 37(1): 89-114.

山田礼子 (2012) 学士課程教育の質保証へむけて—学生調査と初年次教育からみえてきたもの。東信堂, 東京。

山田礼子 (2015) 共通教育における直接評価と間接評価における相関関係 : 成果と課題。大学教育学会2015年度課題研究集会要旨集, 35-37

山田礼子, 森利枝, 亀倉正彦, 斎藤有吾, 高野篤子 (2015) 学士課程教育における共通教育の質保証 : 評価データの併用と質保証のマネジメント。大学教育学会誌, 37(2) : 96-100

(Received April 4, 2016)