

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 —

松下佳代・小野和宏・斎藤有吾・白川優治
(京都大学) (新潟大学) (京都大学) (千葉大学)

〔キーワード：共通教育，質保証，直接評価，間接評価，統合，PBL，学生調査〕

企画者：松下佳代，山田礼子

発表者：松下佳代，小野和宏，斎藤有吾，白川優治

司会：高橋哲也

評価および「授業用学習質問紙（CLQ）」の開発・実施について，サブテーマ3からは，「大学生学習調査2013年（JSLS2013）」の内容および千葉大学での実施結果についてそれぞれ報告する。また，直接評価と間接評価の相関について，米国における研究のレビューを行い，本課題研究での方向性を検討する。（松下佳代）

1. 企画趣旨

課題研究「学士課程教育における共通教育の質保証」（研究代表者：高橋哲也）は，分野別質保証の議論では欠落しがちな共通教育の質保証を主要な対象とする研究であり，2013年度より3年計画で実施されている。研究対象が広範にわたるため4つのサブテーマを設定し，図1のような相互関係の下で研究を進めている。2013年度の課題研究集会では，各サブテーマの研究目的・計画・進捗状況について報告した（『大学教育学会誌』36巻1号，pp.59-77）。今回のラウンドテーブルでは，相補的な関係にあるサブテーマ1と3にしぼって進捗状況を報告する。

本ラウンドテーブルで設定した問題は以下の2つである。①学生の学習成果を明らかにするために，直接評価と間接評価についてそれぞれ，どのような方法を開発し，その結果をどう分析するか，②直接評価と間接評価の統合は可能か，また両者はどう関連づけられるのか。

サブテーマ1からは，新潟大学歯学部のパブリックでの直接

2. 報告

(1) PBLでの直接評価の開発，および間接評価との関連性（小野和宏・斎藤有吾）

近年，大学教育の質的転換の要請から，問題解決を目的としたアクティブラーニングであるProblem-Based Learning (PBL) が共通教育においても注目を集めている。しかし，PBLを通して育成された能力を，妥当性や信頼性，さらには実行可能性をも考慮した上で適切に評価することは容易ではない。今回，PBLにおける問題分析・解決能力，自己学習能力を評価するために考案された，3ステップからなる「トリプルジャンプ」（Blake et al., 1995）の問題点を改良し，新たな直接評価方法を開発した。

改良版トリプルジャンプの特徴の一つは，これまでのトリプルジャンプのステップ1からステップ3を，ステップ1・2として，口頭で代えて文書で評価すること，さらにその評価にあたってはルーブリックを用いることにある。ステップ1では，これまでと同様にシナリオから問題を見つけだし，解決策を立案し，学習課題を設定するが，その過程を60分間でワークシートに記述させる。ステップ2では，学習課題を調査し学習するだけでなく，その結果をもとに解決策を検討し，最終的な解決策を提案するまでを含めて1週間とし，その過程もワークシートに記述させる。それに加え，改良版では，シナリオの状況を再現して，教員を相手にした解決策の実行までをルーブリックを用いて評価することとし，その評価結果のフィードバックを含め，15分間のステップ3を設けている。すなわち，改良版トリプルジャンプは，ワークシートによる筆記課題とロールプレイという実演課題を組み合わせ，2つの異なるタイプのルーブリック（ステップ

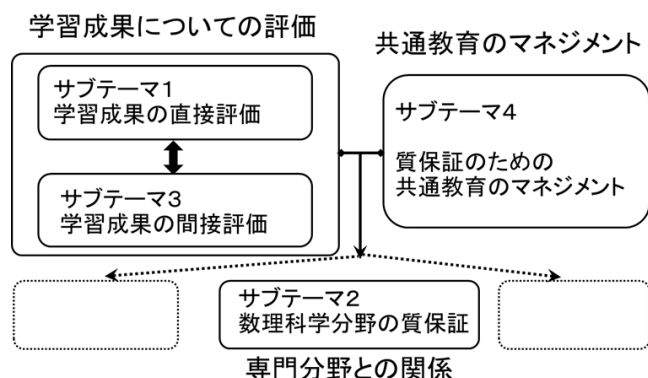


図1 サブテーマ相互の関係
(出典) 高橋 (2014, p.60)

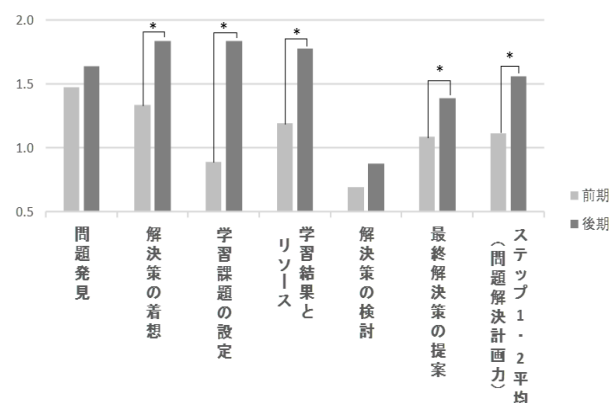
1・2は一般的ループリック、ステップ3は課題特殊的ループリック)を用いたパフォーマンス評価である(小野他, 2014)。

この直接評価をPBLで学ぶ新潟大学歯学部口腔生命福祉学科2013年度2年生24名を対象として、前期と後期の2回、評価者3名で実施したところ、ループリックの観点別の級内相関係数は、おおむね0.7弱～0.9弱の範囲にあり、学生の能力を一定の信頼性をもって評価できることが明らかになった。また、ワークシートを導入したことにより、従来型に比べ、教員の評価負担の軽減がなされることが示された。さらに、改良版トリプルジャンプ終了後の学生アンケート調査の結果から、改良版トリプルジャンプは、単に学習の評価であるだけでなく、それ自体が学生の学習経験にもなるような「学習としての評価」になりえると考えられた。

プログラムレベルでの間接評価の開発はサブテーマ3の課題であるが、授業科目レベルでの間接評価と直接評価の統合の可能性を探るために、サブテーマ1でも「授業用学習質問紙 (Course Learning Questionnaire: CLQ)」を作成した。これは、ある授業における学生の学習への取り組み方や態度を捉えるための間接評価である。本報告では改良版トリプルジャンプによる能力の直接評価と、CLQによる学習プロセスの間接評価の両データを用いた分析を行い、統合の可能性を示した。

具体的には、まず改良版トリプルジャンプのステップ1・2 (ループリックでは6つの観点) の平均点を「問題解決計画力」と操作的に定義した。2013年度は、前期と後期で、課題は異なるが同じループリックを使用していた。したがって、2013年度は学生の「問題解決計画力」を2回評価していたといえる。この2回の評価の結果を、図2に示す。この結果から、「問題解決計画力」は平均的に有意な上昇が見られた。ただし、24名の学生が一樣に同じような変化をしているわけではない。大きく伸びている学生もいれば、そうでもない学生がいるなど、変化にも個人差がある。このような変化の個人差は、学生それぞれが後期のPBLの授業における学習にどのように取り組んでいたか、すなわち学習プロセスに大きく依存していると考えられる。どのような学習をしていた学生ほどポジティブに変化していたのかを明らかにできれば、今後の実践における有力な示唆となるであろう。

そこで、後期のPBLの直接評価の実施と同時期に学生にCLQを配布し、後期のPBLの授業における学習にどのように取り組んでいたのかを尋ねた。このように、前期と後期の2回の直接評価(改良版トリプルジャンプ)によって得られた「問題解決計画力」の変化の個人差に



*は t 検定 (対応あり) で 5% 水準で有意

図2 改良版トリプルジャンプ ステップ1・2 前後期比較

対し、間接評価 (CLQ) によって得られた学習プロセスがどのように影響を与えているのかを探索的に検討することで、直接評価と間接評価の統合を試みた。

後期の「問題解決計画力」を従属変数、前期の「問題解決計画力」と学習プロセスを独立変数として階層的重回帰分析を行った結果を表1に示す。

表1 階層的重回帰分析結果

投入の 順序	変数	最終的な推定値			
		<i>B</i>	<i>SE</i>	β	偏決定係数
1	前期 問題解決計画力	.51	.17	.47 *	.34
2	リハーサル方略	.31	.11	.48 *	.30
	精緻化方略	.42	.20	.60 ($p=0.052$)	.19
	批判的思考方略	-.26	.19	-.34	.10
	主体的な授業態度	-.08	.13	-.13	.02

* $p < .05$

*B*は非標準化偏回帰係数、 β は標準化偏回帰係数を表す

この結果から、暗記をしようとする際に使用される「リハーサル方略」が「問題解決計画力」の変化に大きな影響力を持っていることが示唆された。問題解決という高次の認知的能力の変化に、このような暗記の方略が影響を与えていたのは、幅広い専門知識が要求される医歯学系の学問領域の特徴を反映していると考えられる。すなわち、この分野における問題解決のためには、幅広い専門知識の下支えのうえで解決を計画する必要がある、その下支えをつくるために暗記の方略が有効であったという解釈が可能であろう。またこの結果から、今後の実践においては、この授業で目標としている能力の涵養のためには幅広い知識を覚えることが有効であることを学生に示し、そのためにはどのような学習をしたらよいかというアドバイスをすることで、効果的な学習支援が行える可能性が示唆された。

ただし、サンプルサイズが小さく、推定結果が不安定

である可能性があり、母集団への過度な一般化は危険である。引き続き検討をしていく必要がある。

(2) 共通教育の間接評価の試行結果とそこから見えた課題 (白川優治)

共通教育の到達目標は、包括的であり、多様で広範な内容が設定されることが一般的である。そのため、共通教育の学習成果を把握するには、個々の授業科目の学習成果をみるだけでなく、大学教育全体での経験や、修得が求められる各種の能力項目の自己認識等を学生自身に尋ねることが必要となる。このことから、本課題研究では「学習成果の間接評価」のための手法開発と直接評価との連携の在り方をサブテーマとしている。2013年度は、このうちの前者について、これまでの大学生調査(JCSS)の蓄積(山田, 2012他)を活かしつつ、大学の各種授業・教育方法の経験、大学での学習経験、一週間の生活時間、19の能力項目の大学入学時と比べたときの自己評価など、17項目92問を尋ねる調査票「大学生学習調査2013年(JSLS2013)」を開発した。そして、2013年12月から2014年2月に千葉大学において試行調査を実施し、295(うち1年次生207)の回答を得た。本報告では、このような経過を踏まえ、①「間接評価」はどのように活用することができるか、②「間接評価」と「直接評価」を効果的に結び付けるためにはどのような調査票の開発が必要か、という2つの検討課題の検討を行った。

まず、試行調査の結果の一部を紹介すると、例えば、能力項目の成長認識について、大学入学時と比べたときの自己評価として、能力が高まった(「大きく増えた」+「増えた」)とする項目をみると、上位には「専門分野の学科の知識」(81.0%)、「一般的な教養」(73.0%)、「分析や問題解決能力」(61.4%)、下位には「リーダーシップの能力」(30.6%)、「異文化の人々と協力する能力」(31.8%)、「卒業後に就職するための準備の程度」(31.9%)があり、項目による差がみられた。また、入学後のさまざまな受講経験とこれらの能力項目の成長認識の関係をみると、「学生が自分の考えや研究を発表する」授業の受講経験のある学生は、そうでない学生よりも各能力項目の成長認識が高い傾向が示された。

このような結果に基づいて、本報告の検討課題の観点から、間接評価の3つの可能性と2つの限界を考察した。可能性の第一は、間接評価は、多元的な教育目標等の(主観的な)到達度評価として、教育課程・学習環境・実施している教育上の取組に対する総括的な評価として位置づけられることである。第二は、教育改善を目的に項目間の関連性を分析・検討することによって、大学教育の効果検証に役立てることが可能であることである。第三

は、各種調査項目を用いた「学生群」の形成とその特徴の検討を通じて、学生の受講経験・能力状況の(主観的な)可視化につなげることができることである。他方、その限界は、第一に、間接評価が集合的評価であることから個々の学生と個々の教育実践との関係がみえないこと、第二は、回答者の自己評価に基づくことからデータの信頼性・妥当性の検証が課題となることである。

このような間接評価の可能性と限界を踏まえた今後の方向性として、個々の学生の能力段階の測定を精緻化するための「ループリック型設問」と個々の学生の能力を間接評価のなかで測定する「小テスト型設問」をそれぞれ開発し、これらの項目を含んだ「間接評価」と実際の授業科目(=「直接評価」)の連結を図ることを提示した。

(3) 直接指標と間接指標の相関に関する米国での研究 (松下佳代)

学習成果の直接指標(direct measures)と間接指標(indirect measures)は相関するか、すなわち直接指標の表す学習成果が高い学生ほど間接指標の表す学習成果も高いのか、という問いは、本課題研究の重要な論点の一つである。両者の対比は、〈直接指標 vs. 間接指標〉以外にも、〈テスト vs. サーベイ〉〈客観的指標 vs. 主観的指標〉〈テストされた知識 vs. 自己申告による知識〉といった形で議論されてきた。

この問いに答えようとする研究は米国の高等教育研究においてかなり蓄積されてきているが、その結果には多様性がある。Pike(1996)は、知識獲得に関する学生の自己評価と試験の成績との相関に関する研究のレビューを行っている。それによれば、「有名人に関する知識」について両者の間に中程度の正の相関があるとしたもの(Berdie, 1971)、自己評価とCOMP(College Outcomes Measures Program)の成績との間に弱い正の相関があるとしたもの(Dumont & Troelstrup, 1980)、自己評価とGRE(Graduate Record Examination)および全米教員資格試験の成績との間に弱～中程度の正の相関があるとしたもの(Astin, 1993)などがある。これらの研究から、Pike(1996)は、間接指標は直接指標の代理指標たりうるかという問いに対し、学習達成の一般的指標として学生の自己報告を用いることは正当化できるが、テスト得点に代えて特定の自己報告を用いることは正当化できない、と論じている。さらに、DiRamio & Shannon(2011)は、学生の態度と達成の関係について研究を行い、NSSE(National Survey of Student Engagement)で得られる学生の関与の程度と6年以内の卒業率や卒業時のGPAとは相関しないという研究結果を発表している。このように、直接指標と間接指標の

結果の相関の有無や強さをめぐっては、いまだ諸説ある状態である。

こうした研究に加えて、間接指標そのものについて検証する研究も繰り返し行われている。表2は、「教科書やノートのまとめを作る」頻度について、表現を変えて尋ねたときの回答分布である。他の学修行動でも同様の結果が得られている（Pace & Friedlander, 1982; Laird et al, 2008）。問い方によって、回答にかなりのばらつきがあることがわかる。

表2 「教科書やノートのまとめを作る」の頻度 (%)

	occasionally	often	very often
一度もない	10	3	0
年に1～2回	21	6	2
年に3～6回	26	12	6
月に1～2回	24	27	18
週1回程度	13	32	30
週1回以上	6	20	44

(出典) Pace & Friedlander (1982, p.271) より訳出、一部修正。
(注) 最初に頻度を表わす副詞で尋ね、後に回数を尋ねた場合の回答分布。
最初の質問はneverを含む4件法。

間接指標の中でもとりわけNSSEをめぐるのは、その影響力の大きさゆえに、多くの議論がなされてきた。NSSEの質問項目で使われている用語(頻度表現や「批判的思考」などの語句)が多義的で妥当性に欠けるという批判に対して、NSSEの側からは、NSSEは完全ではないが効率的であり、また単独の数値ではなく経年比較や大学間比較で有効だという反論が行われているが（Jaschik, 2009）、その後も議論は続いている。

以上から、本課題研究へのインプリケーションとして、次の2点をあげておきたい。第一に、間接評価（質問紙調査）における用語の選択である。なるべく具体的な数値や表現を用いて曖昧さを避ける必要がある。第二に、直接評価と間接評価の統合に有効な手法の開発である。ID番号が付与できない場合もあることを考えると、同じ調査票で質問紙調査と小テストを実施できるような仕組みも検討する必要がある。

(付記)以上の報告の作成にあたっては、研究分担者である森利枝氏（大学評価・学位授与機構）の協力を得た。

3. ディスカッション

(1) 相関をみることの意味

以上の報告の後、参加者もまじえたディスカッションに移った。まず、「今回の報告はいずれも、直接指標と間接指標の相関を検討し、相関が高い方がよいとする前提

に立っているようにみえる、むしろズレにこそ意味があるのではないか」ということが議論になった。

これに対しては、「確かに複数の評価の統合には、質的研究のトライアングレーション（複数の方法やデータを組み合わせるやり方）と似たところがある。データが収斂すること（相関が高くなること）ばかりが望ましいのではなく、むしろ異なる結果が出るからこそ複数の評価方法で学習成果をみることに意味があるともいえる」という意見が出された一方で、「データの収集や分析の仕方が悪いために本来現れるべき相関が現れない場合もあるので、そうした可能性も含めて、ズレの原因について検討する必要がある」という意見も出された。

(2) 間接評価の信頼性・妥当性

次に議論になったのは、間接評価の信頼性と妥当性の問題である。これについては、フロアのサブテーマ3のメンバーから「NSSEについては、AAHE (American Association for Higher Education) でも高等教育の質を反映していると判断していることから、一定の妥当性が担保されているといえるのではないか」という補足があった。

一方、サブテーマ1で開発したCLQの信頼性係数(α 係数)があまり高くないこと(一般に求められる0.7に達していないものがあること)についての疑義に対しては、報告者から「今回評価した学生数が少なく分散が小さかったため、引き続き他大学でも用いることによって尺度としての信頼性を検討したい」という回答があった。

(3) ルーブリックの教育的意義

さらに、ルーブリックの教育的意義について、「ルーブリックを学生に提示すれば、学生は適応してくるのでレポートの得点は高くなってくるが、それでよいのか」という問題が投げかけられた。

報告者からは、「学士課程段階のレポート・ライティングは一定の型を学ばせることが目標なので、それでよい」という意見とあわせて、「新潟大学の事例では、一般的ルーブリックの場合は学生に提示しているが、課題特殊的ルーブリックの場合は(「傾向と対策」的な学習になることを避けて)学生に提示せず、代わりに教員からの口頭のフィードバックを返している」という補足があった。一般に、ルーブリックは学生の自己評価を促すよう事前に提示すべきだとされているが、常にそれが適切だとは限らない。領域やルーブリックの性格によってケース・バイ・ケースで対応しながら、原則を見つけていく必要がある。

(4) その他

「間接評価は学習成果を予測する独立変数で、直接評価

は従属変数ではないか」という質問に対しては、「間接評価には〈学習過程の評価〉と〈学習成果の間接評価〉という2つの意味があり、前者の意味では、間接評価＝独立変数、直接評価＝従属変数となるが、後者の意味では必ずしもそうはならない」という応答があった。

また、「直接評価とマージするために間接評価でもID番号を記入させると、学生が望ましいと思われる回答をしやすくなるおそれがないか」という疑義も出された。これについては、ID番号を記入させることのメリット・デメリットをあらためて慎重に検討する必要がある。

このように、短時間ながら、根本的な問題からテクニカルな問題まで、途切れることなく議論がかわされ、実り多いディスカッションとなった。（松下佳代）

4. まとめ

本課題研究では当初、図1に示したように、サブテーマ1では直接評価、サブテーマ3では間接評価の開発を行い、両者を統合する予定であった。しかしながら、研究を進めるなかで見えてきたのは、サブテーマ1で開発した授業科目レベルでの直接評価と、サブテーマ3で開発したプログラムレベルでの間接評価を統合することの困難さであった。そこで、現在では、サブテーマ1において、授業科目レベルでの直接評価（パフォーマンス評価）と間接評価（CLQ）の統合、また、サブテーマ3において、プログラムレベルでの間接評価（JSLS2013）と直接評価（小テストなど）の統合をそれぞれ試みるにいたっている。

今後は、授業科目レベルとプログラムレベルの統合をどう図るかが、サブテーマ1と3の連携における最大の課題となるだろう。また、Pike (1996) では、間接指標は直接指標の代理指標たりうるかという問いが立てられていたが、学生の自己報告が単に直接指標の代理指標ではなく固有の意味をもつような学習成果にどんなものがあるかも、それぞれの独自性を明らかにしていく上で検討する必要がある。例えば、「自己効力感」などは、学生の自己報告によるのが適した学習成果といえるだろう。

その他、質疑応答を通じて明らかになった論点や問題についても、残り1年半の課題研究期間の中で最大限答えていきたい。

なお、当日の配付資料はすべて、本学会ウェブサイトの課題研究のページからダウンロードできる (<http://daigakukyoiku-gakkai.org/site/research/projects/>)。参加者は61名であった。（松下佳代）

参考文献

- Astin, A.W. (1993). *What matters in college?: Four critical years revisited*. San Francisco: Jossey-Bass.
- Berdie, R. F. (1971). Self-claimed and tested knowledge. *Educational and Psychological Measurement*, 31, 629-636.
- Blake, J. M., Norman, G. R., & Smith, E. K. (1995). Report card from McMaster: Student evaluation at a problem-based medical school. *The Lancet*, 345, 899-902.
- DiRamio, D., & Shannon, D. (2011, April). Is NSSE messy?: An analysis of predictive validity. Paper presented for the Annual Meeting of the American Educational Research, New Orleans, LA.
- Dumont, R. G., & Troelstrup, R. L. (1980). Exploring relationships between objective and subjective measures of instructional outcomes. *Research in Higher Education*, 12, 37-51.
- Jaschik, S. (2009, November 9). Engaged or confused? *Inside Higher Ed*. (<http://www.insidehighered.com>)
- Laird, T. F. N., Korkmaz, A., & Chen, P.-S. D., (April, 2008). How often is “often” revisited: The meaning and linearity of vague quantifiers used on the National Survey of Student Engagement. Paper presented for the Annual Meeting of the American Educational Research, San Diego, CA.
- 小野和宏・松下佳代・斎藤有吾 (2014) 「PBLにおける問題解決能力の直接評価—改良版トリプルジャンプの試み—」, 『大学教育学会誌』 36-1, pp.123-132.
- Pace, C. R., & Friedlander, J. (1982). The meaning of response categories: How often is “occasionally,” “often,” and “very often”? *Research in Higher Education*, 17(3), 267-281.
- 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.
- 高橋哲也 (2014) 「学士課程教育における共通教育の質保証に向けて」, 『大学教育学会誌』 36-1, pp.59-61.
- 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。