

どのような観点が探究的な学習の全体評価に影響するか

—探究発表大会2ヶ年データの分析から—

立命館大学教育開発推進機構

蒲生 諒 太

1. 問題と目的

2022年度より年次進行で実施される高等学校学習指導要領において、総合的な学習の時間は「総合的な探究の時間」と衣替えし、中心的な学習活動である「探究的な学習」（以下、探究）が今以上に強調されることになる。この学習方法は2020年度から始まる高大接続改革でも重要な役割を担うものと考えられ（蒲生、2019）、2019年度現在、教員向けの指導法をまとめた実践書（酒井、2013、がもう、2017、2018）や学習教材（岡本、2017）が刊行され、さらに教員向け研修会⁽¹⁾が実施されてきている。

このような探究を巡る動きの中でも、学習者の学習成果（探究成果）発表の機会が盛んになってきている。それは、スーパーサイエンスハイスクールなど、地域の拠点校が中心となる発表会⁽²⁾、各学協会が学術研究大会とともに実施するジュニアセッション⁽³⁾、民間企業などが主催するイベント⁽⁴⁾などである。このような発表イベントでは学習成果の評価や表彰が行われることが多い。また、学習指導を行う上でも探究の学習成果について評価を行うことが必要になる。

評価に関してはルーブリックの作成等、学習成果＝パフォーマンスをどのような基準・観点で評価するかが大切な論点となるが、そのためには探究成果についての全体評価と各種観点がどのような関係であるのか（どのような観点が探究成果の全体評価に影響を与えるのか）を明らかにしておきたい。

本稿では筆者が実行委員会の代表を務める「IBL ユースカンファレンス」という探究成果発表大会における2カ年の評価データをもとに、どのような

観点が探究成果の全体評価に影響を与えるのか、明らかにする。

IBL ユースカンファレンスは2019年3月の大会をもって、3回目を数える探究成果発表大会である。当日発表と展示発表があり、北は北海道から南は大分県まで参加者が集う全国大会となっている⁽⁵⁾。本稿では2018年3月29日の第2回大会、2019年3月21日の第3回大会における当日発表部門の評価データを対象とする。

2. 方法

(1) IBL ユースカンファレンスにおける評価について

IBL ユースカンファレンスではすべての当日発表・展示発表について、4段階の総合認定評価（金・銀・銅・選外の認定）とともに当日発表については観点別評価を行っている（表1）。第2回大会ではルーブリックが作成され、毎回、認定評価委員会のもと大学教員・大学院生・卒論を書き終え修士課程に進学を予定する学部4回生からなる審査員団が結成される。評価においては、1つの発表につき、1審査員でルーブリックによる評価を行っており、すべてのデータは事務局に集約され、発表者には最終的な総合認定評価（全体評価）がフィードバックされる。

評価体制が確立されていくまでの経緯を示したい。まず、ルーブリック作成以前の第1回大会では審査員団と学生スタッフの混成チームによって、1発表につき1審査員の評価に2人程度のスタッフがサポートをするかたちで総合認定評価のみ実施していた。その後、高大接続改革を念頭に、評価方法の明確化を行うため、6観点からなるルーブリックが作成され、第2回大会以降の審査ではルーブリックによる観点別評価とそれを踏まえた総合認定評価が行われるようになった。総合認定評価については審査員の専門性を信頼し、各審査員に一任しており、観点別評価の積算点等で総合認定評価がなされているわけではない。ルーブリックに関しては第3回大会で1項目追加された。今回は第2回大会・第3回大会で共通する6つの観点別評価と総合認定評価について検討した。

表1 評価用ルーブリック

<p>1. 探究成果について</p> <p><u>(1) 自分たちの探究（研究）の価値を説明できているか《「価値説明」》</u> 金 十分に説明しており、意図も理解できる 銀 説明しており意図も理解できるが、疑問点が残る 銅 説明しようとしているが、意図が理解できない 選外 説明していない</p> <p><u>(2) 探究は論理的に行われているか《「論理的」》</u> 金 方法、結果の読み取り、考察ともに十分に論理的で道筋が通っている 銀 全体的に筋が通っているが、一部破綻している部分が指摘できる 銅 大筋で説明に破綻が認められる 選外 そもそも筋道の通った説明を試みようとしていない</p> <p><u>(3) 多角的な研究・議論が行われているか《「多角的」》</u> 金 複数の研究・議論が行われ、1つの課題意識にまとまっている 銀 複数の研究・議論が行われているが、まとまりきれていない 銅 複数の研究・議論は行われていないが、多角的な検討は認められる 選外 探究全体に極端にまとまりがない</p> <p>2. ポスターの形式について</p> <p><u>(1) ポスターは見やすいか《「見やすさ」》</u> 金 レイアウトや文字の大きさ等に注意を向けられ、見やすい 銀 レイアウトや文字の大きさ等に注意を向けているようだが、見にくさが多少ある 銅 ポスターが全体的に見にくく、根本的な改良の余地が指摘できる 選外 何が書かれているのか、判別できない</p> <p><u>(2) 適切なデータ・議論が可視化されているか《「適切なデータ・議論」》</u> 金 図表が分かりやすく、内容の理解を助けている 銀 一部に不適切ないしは分かりにくい図表があり、理解の妨げが指摘できる 銅 説明に重要な図表に誤りが指摘できる 選外 重要な図表に根本的な誤りがあったり、そもそも図表等がなく、分かりにくい</p> <p>3. 発表態度について</p> <p><u>(1) 聴衆に対して十分に働きかけているか《「聴衆働きかけ」》</u> 金 聴衆の興味や疑問を引き出すような話し方をしている 銀 聴衆の興味や疑問を引き出すような話し方をしているがうまくいっていない 銅 聴衆の方を見て話している 選外 聴衆の方を見て話していない</p> <p>4. 最終評価 金/銀/銅/選外</p>
--

(2) 研究承諾について

評価データの研究利用については、第2回大会では各参加者に書面で説明され、参加校代表者の署名にて承諾とした。第2回大会審査員団・第3回大会審査員団については分析に際して改めてSNS等を通して連絡され承諾された。

第3回大会参加者についても同じタイミングで参加校代表者にWebメー

ルにて連絡され承諾された。承諾に際しては、評価データの匿名化を行い、発表に際してはどの成果物を誰がどのような評価をしたのかは分からないようにすること、承諾については自由に拒否することができること、分析成果については各校代表者及び審査員団にフィードバックされ、可能であるなら、論文が学会発表等で一般に公開される旨が伝えられた。

(3) データの概要

分析対象となる当日発表は2ヶ年で160件であるが(当日配布プログラムよりカウント)、当日欠席などもあり認定証発行カウントでは154件、そのうち、評価データとして分析可能なものとして蓄積されているものが138件となっていた(欠損値があった場合や二重投稿など削除。二重投稿についてはイベント当日には認定のため、口頭でスタッフが確認しているが今回は削除した)。

(4) 分析の方法

まず、単純集計を行い全体の傾向を把握した。

次に探究の特性を考えるため、著者の現場指導経験や現場との交流から「理数系探究 vs. 非理数系探究」、「単著 vs. 共著」の点から総合認定評価と観点別評価をクロス集計し、 χ^2 検定した。

理数系探究か非理数系探究かは審査員振り分けの経験をもとに、発表タイトルから分類した。「理数系探究」は理科(物理・化学・生物・地学)と数学に強い関連を認められるもので「非理数系探究」はそれ以外のものと考えた。学際的な印象を与えられるものに関しては後者として扱った。たとえば、「マグネシウム電池の開発」は理数系探究とし、「自然と共存できる街づくり」は非理数系探究とした(ともに実例を参考に作成したダミー)。

「単著」と「共著」は当日配布プログラムの表記をもとに発表者が1名なら単著、そうでないなら共著とした。

最後に総合認定評価と観点別評価の関係を明らかにするため、相関分析を行った。データの特性からここではスピアマンの順位相関係数を出し、有意な相関であるかを確認しながら、それぞれの相関の強さを検討することにした。

統計解析は IBM SPSS Statistics23を使用した。

3. 結果

(1) 単純集計

2カ年の総合認定評価及び観点別評価を表2にまとめた。以下、金・銀・銅・選外の評価については金=Level.4・銀=Level.3・銅=Level.2・選外=Level.1と表記している。総合認定評価の結果はL.4とL.3で全体の85.5%となっている。ちなみに2ヶ年での割合の差があるか、差があるかどうかを判定するために別途 χ^2 検定を行ったが、有意差はなかった。

表2 単純集計結果

総合認定			価値説明			多角的			データ・議論		
評価	件数	割合	評価	件数	割合	評価	件数	割合	評価	件数	割合
L.4	57	41.3	L.4	68	49.3	L.4	60	43.5	L.4	46	33.3
L.3	61	44.2	L.3	61	44.2	L.3	46	33.3	L.3	79	57.2
L.2	20	14.5	L.2	7	5.1	L.2	30	21.7	L.2	7	5.1
L.1	0	0	L.1	2	1.4	L.1	2	1.4	L.1	6	4.3
合計	138	100	合計	138	100	合計	138	100	合計	138	100
		(%)			(%)			(%)			(%)

論理的			見やすい			聴衆		
評価	件数	割合	評価	件数	割合	評価	件数	割合
L.4	65	47.1	L.4	57	41.3	L.4	82	59.4
L.3	64	46.4	L.3	69	50	L.3	39	28.3
L.2	8	5.8	L.2	12	8.7	L.2	15	10.9
L.1	1	0.7	L.1	0	0	L.1	2	1.4
合計	138	100	合計	138	100	合計	138	100
		(%)			(%)			(%)

(2) 「理数系探究」と「非理数系探究」、どちらが高い評価を得たか

次に「理数系探究 vs. 非理数系探究」での比較検討を行った。全体を分類し集計したものが表3である。割合としては理数系探究が多いが、非理数系探究も45.7%と半分近くになっている。

この分類をもとに総合認定評価及び観点別評価とクロス集計、 χ^2 検定を行った。その際、分析の都合上、L.4か、それ以外かで区分した。

表3 理数系・非理数系集計

分類	件数	割合
理数系	75	54.3
非理数系	63	45.7
合計	138	100
		(%)

表4 理数系・非理数系＊観点別評価

価値説明		L.4	L.4未満	合計
理数系	件数	39	36	75
	割合	52%	48%	100%
非理数系	件数	29	34	63
	割合	46%	54%	100%
合計	件数	68	70	138
	割合	49.3%	50.7%	100%

$\chi^2 = .488, p = .485$

データ・議論		L.4	L.4未満	合計
理数系	件数	24	51	75
	割合	32%	68%	100%
非理数系	件数	22	41	63
	割合	34.9%	65.1%	100%
合計	件数	46	92	138
	割合	33.3%	66.7%	100%

$\chi^2 = .131, p = .717$

論理的		L.4	L.4未満	合計
理数系	件数	42	33	75
	割合	<u>56%</u>	44%	100%
非理数系	件数	23	40	63
	割合	36.5%	<u>63.5%</u>	100%
合計	件数	65	73	138
	割合	47.1%	52.9%	100%

$\chi^2 = 5.21, p = .022$ (5%水準で有意)

聴衆		L.4	L.4未満	合計
理数系	件数	37	38	75
	割合	49.3%	<u>50.7%</u>	100%
非理数系	件数	45	18	63
	割合	<u>71.4%</u>	28.6%	100%
合計	件数	82	56	138
	割合	59.4%	40.6%	100%

$\chi^2 = 6.932, p = .008$ (1%水準で有意)

多角的		L.4	L.4未満	合計
理数系	件数	34	41	75
	割合	45.3%	54.7%	100%
非理数系	件数	26	37	63
	割合	41.3%	58.7%	100%
合計	件数	60	78	138
	割合	43.5%	56.5%	100%

$\chi^2 = .230, p = .631$

総合認定		L.4	L.4未満	合計
理数系	件数	31	44	75
	割合	41.3%	58.7%	100%
非理数系	件数	26	37	63
	割合	41.3%	58.7%	100%
合計	件数	57	81	138
	割合	41.3%	58.7%	100%

$\chi^2 = .000, p = .994$

見やすい		L.4	L.4未満	合計
理数系	件数	29	46	75
	割合	38.7%	61.3%	100%
非理数系	件数	28	35	63
	割合	44.4%	55.6%	100%
合計	件数	57	81	138
	割合	41.3%	58.7%	100%

$\chi^2 = .471, p = .492$

結果は表4であるが、 χ^2 検定の結果、理数系探究と非理数系探究の間で総合認定評価において有意差はなく、どちらのテーマを選ぶかによつての全体評価の高低は見受けられなかった。

観点別評価についても表4にまとめた。 χ^2 検定の結果、「論理的」と「聴衆働きかけ」の2点でそれぞれ5%、1%水準で有意差が認められた。前者に関しては理数系探究がL.4の割合が高く、後者に関しては非理数系探究の割合が高かった。

(3)「単著」と「共著」、どちらが高い評価を得たか

さらに「単著 vs. 共著」での比較検討を行った。全体を分類し集計したものが表5である。割合としては共著が8割と圧倒的であるが、単著も2割程度と一定存在する。単著はそもそも全体の割合が少ない(件数が少ない)こともあり、構成比について示しておく。学校別では26件中9件(34.6%)が1校、3件(11.5%)が1校、2件(7.7%)が3校、1件(3.8%)が8校だった。理数系か、非理数系かでは、理数系探究が26件中5件(19.2%)、非理数系探究が21件(80.8%)だった。

表5 単著・共著集計

分類	件数	割合
単著	26	18.8
共著	112	81.2
合計	138	100

(%)

この分類をもとに総合認定評価及び観点別評価とクロス集計、 χ^2 検定を行った。

結果は表6であるが、 χ^2 検定の結果、単著と共著の間で総合認定評価に有意差はなく、探究を1人で行うか、そうでないかで評価の高低は見受けられなかった。

観点別評価についても表6にまとめた。 χ^2 検定の結果、「聴衆働きかけ」が5%水準で有意差が認められた。ともにL.4が割合として多いが、その比率として単著のほうが高いことが分かった。

表6 単著・共著*観点別評価

価値説明		L.4	L.4未満	合計
単著	件数	15	11	26
	割合	57.7%	42.3%	100%
共著	件数	53	59	112
	割合	47.3%	52.7%	100%
合計	件数	68	70	138
	割合	49.3%	50.7%	100%

$\chi^2 = .908, p = .341$

データ・議論		L.4	L.4未満	合計
単著	件数	8	18	26
	割合	30.8%	69.2%	100%
共著	件数	38	74	112
	割合	33.9%	66.1%	100%
合計	件数	46	92	138
	割合	33.3%	66.7%	100%

$\chi^2 = .095, p = .758$

論理的		L.4	L.4未満	合計
単著	件数	10	16	26
	割合	38.5%	61.5%	100%
共著	件数	55	57	112
	割合	49.1%	50.9%	100%
合計	件数	65	73	138
	割合	47.1%	52.9%	100%

$\chi^2 = .960, p = .327$

聴衆		L.4	L.4未満	合計
単著	件数	21	5	26
	割合	<u>80.8%</u>	19.2%	100%
共著	件数	61	51	112
	割合	<u>54.5%</u>	45.5%	100%
合計	件数	82	56	138
	割合	59.4%	40.6%	100%

$\chi^2 = 6.055, p = .014$ (5%水準で有意)

多角的		L.4	L.4未満	合計
単著	件数	10	16	26
	割合	38.5%	61.5%	100%
共著	件数	50	62	112
	割合	44.6%	55.4%	100%
合計	件数	60	78	138
	割合	43.5%	56.5%	100%

$\chi^2 = .328, p = .567$

総合認定		L.4	L.4未満	合計
単著	件数	14	12	26
	割合	53.8%	46.2%	100%
共著	件数	43	69	112
	割合	38.4%	61.6%	100%
合計	件数	57	81	138
	割合	41.3%	58.7%	100%

$\chi^2 = 2.079, p = .149$

見やすい		L.4	L.4未満	合計
単著	件数	13	13	26
	割合	50%	50%	100%
共著	件数	44	68	112
	割合	39.3%	60.7%	100%
合計	件数	57	81	138
	割合	41.3%	58.7%	100%

$\chi^2 = .999, p = .318$

(4) 「総合認定評価」と「観点別評価」の関係はどのようなものか

最後に総合認定評価と観点別評価の関係を明らかにするため、相関分析を行った。

まず、総合認定評価と観点別評価の相関を表7にまとめた。

表7 総合認定評価*観点別評価

項目	価値説明	論理的	多角的	見やすい	データ・議論	聴衆
総合認定	.718**	.708**	.573**	.414**	.630**	.485**

**は1%水準で有意(両側)、*は5%水準で有意(両側)
下線は強い相関

総合認定評価は観点別評価の全項目と相関していることが分かった(1%水準)。相関の強さを見ると、「価値説明」と「論理的」が強く相関しており、その他は比較的強い相関となっている。

次に観点別評価同士の相関を調べて表8にまとめた。

表8 観点別評価全項目

項目	価値説明	論理的	多角的	見やすい	データ・議論	聴衆
価値説明	1	.516**	.452**	.269**	.395**	.376**
論理的	.516**	1	.440**	.300**	.607**	.261**
多角的	.452**	.440**	1	0.166	.392**	.319**
見やすい	.269**	.300**	0.166	1	.248**	.215*
データ・議論	.395**	.607**	.392**	.248**	1	.214*
聴衆	.376**	.261**	.319**	.215*	.214*	1

**は1%水準で有意(両側)、*は5%水準で有意(両側)
下線は比較的強い相関

全項目間で相関が見られたが、とくに下線部は比較的強い相関が認められた。

4. 考察

(1) IBL ユースカンファレンスの傾向について

単純集計結果を見ると最高評価のL.4が4割程度とかなり高いものになっていることが分かる。ループリックの各観点L.1に当たるものの多くがそも

そも探究として成り立たない水準のものであることを考えると、実質的に金・銀・銅の3段階評価となるが、そう見ても（期待値が33.3%と考えると）全体の評価は高い傾向にある。

この点は審査員が甘い評価をしているというよりも、そもそも参加校自体がSSH（スーパーサイエンスハイスクール）やSGH（スーパーグローバルハイスクール）などの研究指定校が多く、そのため、探究成果の水準がもとから高いからだと理解できる。発表レベルの高さはこの種のイベントとして誇れることにも感じられるが、その一方でIBLユースカンファレンス自体、「入門段階から」を謳っていることを考えると一定水準以上の探究成果の発表ばかりになっているのは探究入門層への訴求が十分でないということでもある。ただ、そもそも「入門段階」の学習者が外部での発表に臨むことが本稿執筆時現在、一般的とは言えない現状において、学校外の発表大会である以上、一定以上の学習成果になることは必然かもしれない。

この点は分析結果やその解釈に大きく影響を与えることに注意したい。つまり、全国で行われている探究のうち、ここでの学習成果は一定水準以上のものであるということを以下、念頭に考えないといけないのである。

（2）どのような探究が高い評価であったか

本稿では、まず、「理数系探究 vs. 非理数系探究」の観点から評価の検討を行ったが、総合認定評価に関しては差が見られなかった。この点は探究テーマよっての有利不利が少なくとも理数系探究かそうでないかという視点では認められなかったことを意味している。一方で理数系探究は非理数系探究よりも「論理的」について優位性を持っており、この点を重視した場合、非理数系探究は理数系探究に劣る部分が出てしまう。ただ、非理数系探究は理数系探究よりも「聴衆働きかけ」で高い評価になる割合が多く、この点が総合認定評価では相殺している可能性もある。

次に単著と共著の観点別評価に目を向けると、単著のほうが「聴衆働きかけ」において共著より高い評価を得る傾向が見受けられた。ただし、殆どが「非理数探究」であり、構成比でも示したように3割近い割合を示す学校もあることから、そもそもこの分析結果が純粹に「単著 vs. 共著」の比較結果を示唆しているのか、微妙なところであり、結果に基づく判断を留保したい。

総合認定評価と観点別評価の相関分析から、「価値説明」と「論理的」が強く総合認定評価に関係していることが明らかになった。「価値説明」は「論理的」及び「多角的」と関係が強く見られ、また「論理的」は「価値説明」と「多角的」、「適切なデータ・議論」と関係が強く見られた。

これらのことを総合すると次のようなことが言える。テーマが理数系か、そうでないかは探究成果の総合認定評価（全体評価）に関係しているとは言えそうにない。観点別で見ると探究の「価値説明」と「論理的」の総合認定評価との関係が強そうで、「多角的」や「適切なデータ・議論」も一定の強い関係がありそうである。「見やすさ」や「聴衆働きかけ」は総合認定評価に一定の関係性もあるのだが、相対的にはそれほど高いものではない。

ループリックに立ち返るなら、自分たちの探究（研究）の価値を「十分に説明しており、意図も理解できる」こと（「価値説明」）、「方法、結果の読み取り、考察ともに十分に論理的で道筋が通っている」こと（「論理的」）が重要であり、また、「複数の研究・議論が行われ、1つの課題意識にまとまっている」こと（「多角的」）、「図表が分かりやすく、内容の理解を助けている」こと（「適切なデータ・議論」）も意義があるようである。

こう考えると、高い評価を得る探究成果というのは研究としての基本的な構成要件である、研究の価値を説明するとともに明瞭な課題意識を示した「問題設定」と分かりやすい「結果」の提示、それらが論理的であること、そして、探究全体が一貫した課題意識のもと行われていることが重要となるようである。探究成果の見せ方や聴衆へのアピールというレトリカルなテクニックは総合的な評価に一定の影響を与えるものの、総合認定評価に大きく影響するような探究の本質とはいいいにくそうである。

以上が本研究の結論であるが、多くの課題と限界点を持っている。既に示したように本研究のもとになる IBL ユースカンファレンスは「入門段階から」を謳いつつも、一定水準の探究成果が集まる状況であり、今回のデータのサンプルとしての価値は高いというわけではない。本研究では触れなかったが観点別評価の一部では天井効果が認められるなど、データに歪みがあり分析にも限界があった。今後は IBL ユースカンファレンスの裾野を広げながら、探究の学習機会提供という公共的な目的とともに万遍ないデータの収集に努めながらよりクリアな分析ができるように尽力したい。

註釈

- (1) 高校教員向け探究学習指導セミナー2019夏(大阪大学高等教育・入試研究開発センター主催 <https://chega.osaka-u.ac.jp/event/201906151822/> 最終確認：2019年11月24日、以下 URL の確認日については同じ) など
- (2) 京都探究ポスターセッション2019 (京都市立堀川高等学校主催 開催案内 [https://www.city.kyoto.lg.jp/kyoiku/cmsfiles/contents/0000259/259031/191018\(kouhou\)horikawassh.pdf](https://www.city.kyoto.lg.jp/kyoiku/cmsfiles/contents/0000259/259031/191018(kouhou)horikawassh.pdf)) など
- (3) 日本天文学会ジュニアセッション (<http://www.asj.or.jp/jsession/>) や、ジュニア農芸化学会 (日本農芸化学会主催 http://www.jsbba.or.jp/science_edu/event_junior.html) など
- (4) サイエンスキャッスル (株式会社リバネス主催 <https://s-castle.com>) など
- (5) 各大会の参加者数及び発表件数については以下。第1回 (2016年度) 109人、ポスター発表24件、展示発表19件／第2回 (2017年度) 241人、ポスター発表69件、展示発表33件／第3回 (2018年度) 352人、ポスター発表91件、展示発表44件

文献

- 岡本尚也 (2017) 『課題研究メソッド』 新興出版社啓林館
- がもうりょうた (2018) 『探究実践ガイドブック』 七猫社
- がもうりょうた (2017) 『「探究」カリキュラム・デザインブック：アクティブ・ラーニング＝主体的・対話的で深い学びはじめました。』 ヴィッセン出版
- 蒲生諒太 (2019) 「今求められる高大接続改革」(香川七海・福若真人・蒲生諒太『教育原理』 七猫社、136-140頁)
- 酒井聡樹 (2013) 『これから研究を始める高校生と指導教員のために：研究の進め方・論文の書き方・口頭とポスター発表の仕方』 共立出版

謝辞

紙幅の都合上、それぞれの校名・人物名については省略させていただきますが、評価データを提供いただきました各学校の皆さま、審査員の皆さま、

さらにイベントに参加いただきました生徒の皆さまに深謝申し上げます。また、IBL ユースカンファレンス実行委員会および事務局の皆さま、イベント開催のためご協力いただいたスタッフ・関係者の皆さまに心より感謝申し上げます。