中学校における特別活動（学校行事）としての合唱活動と、その競技性4
― 合唱活動に絶対評価、相対評価を導入するための基礎研究 ―

柴田　篤志

Choral activity, as special activity (school events), and its competitive nature 4: Basic research to introduce absolute evaluation and relative evaluation into choral activities

構成

0. はじめに（合唱コンクールを採点すること、その意味について）
1. 模擬合唱コンクール
　　1-1　課題曲及び自由曲について
　　1-2　採点方法について
2. 模擬合唱コンクール結果
　　2-1　採点結果（絶対評価）
　　2-2　採点結果（相対評価）
3. 考察
　　3-1　順位判断の基準としての「順位計」
　　3-2　評価項目ごとの標準偏差から見る絶対評価、相対評価
　　3-3　評価者の採点傾向から見る絶対評価、相対評価
4. まとめ

0. はじめに（合唱コンクールを採点すること、その意味について）

本論は、継続研究の第4稿に当たる。

2020の同タイトル研究（3）では「まとめ」において

> 集計と検証を終えて今更のように「競技性を導入するとして、そこに順位は必須なのか」を改めて悩ましく感じる。今回の集計結果は、1位と7位は根拠のあるものとして納得できたが、それ以外の順位はまさしく僅差であって、算術平均の差を順位に用いていいのか、納得できなかった。ましてや、順位だけを結果として聞く生

徒たちはなおさらだろう。

と述べた。順位は必要なのか、その順位は投じられた点を合計して投票者数で割る、いわゆる算術平均の差で判断すべきなのか…この疑問に対する一つの挑戦すべき方法論として、「絶対評価による順位と、相対評価による評価」、いずれが教育亭な妥当性を持つのかについて論ずることが本稿の目的となる。

昨年までの研究において、採点の方法は「絶対評価」を用いている。前提となる評価方法が、合唱コンクール参加者による「相互採点」であったたである。

特別活動（学校行事）として行われる合唱コンクールに順位を出す場合、多く行われるのが審査員による採点となる。知見に優れた専門家が、独自の審美眼に基づいて順位付けるもので、独善的ではあるが「その人の見立てであるのだから」という意味でのフォローも容易となるため、実際によく用いられており、独善性を排するために複数審査員を採用したり、教員(複数)がそこに配置されたり、専門性の高さによって持ち点が変えられたり、という細かなレギュレーション上の調整は頻繁に行われる。

このやり方の欠点は、採点に「教育的配慮」が必ずしも用いられるとは限らないこと、用いられるとして、日々教育活動に関わっている学校側の見立てと一致するとは限らないことであろう。

私が過去の研究において採用してきた「相互採点」は、専門性には欠ける反面、学校での生活を長時間ともにする学友同士の評価となるため教育的効果が大いに期待できるところに利点がある。しかし、当然のことながら、合唱イベントに参加する当事者でもあるため思い込みや強い主観が現れやすく、極端な採点結果を導きやすいという危険性をはらむ。継続研究の（1）～（3）まではその辺りを見極めてきたつもりである。

本稿に至り、模擬合唱コンクールに参加しているのが大学生、それも日々音楽の専門的な学習に勤しむ音楽大学の学部学生であるということの特殊性を排除するべきではないかと考えた。過去三稿における審査結果は（中間順位の不透明さはさておき）優秀な演奏、問題を多く有する演奏に関しては文字通り『妥当』な評価結果が出ていた。しかし、果たして一般の中学生にこのやり方を求めた場合同じ結果を期待していいのだろうか。

そこで今回は「予備研究」という立場をとり、絶対評価に基づく審査員方式とは別に、「どのグループが一番上手でしたか」と順位だけを問う相対評価に基づく順位の決定法を同時に求めた。二つの評価方法の間にどのような差異が生じるのか、それぞれの利点と欠点はどこにあるかを考えるための基礎データを提示したい。

1．模擬合唱コンクール

1-1　課題曲及び自由曲について

本稿において、「曲の選定」は考察対象としていない。課題曲はこちらから与え、自由曲は23曲の候補曲をリストとして渡し、そこから選択させている（第三希望までの候補を提出させ、第一候補優先でもし重複した場合は第二候補、第三候補への変更を許す）。

《連絡文》

課題曲と、選曲リストです。自由曲は、リストの23曲から一曲を選んでください。ただ、他のチームとダブる可能性もあります。重複もあり、と考えていますが、重複した場合の第二候補、第三候補を選んでおいてもいいです。

《課題曲と自由曲》…「曲名」作曲者

◆課題曲

「明日の空へ」山崎朋子

◆自由曲リスト

1「IN TERRA PAX」荻久保和明

2「omnibus star 光年の旅」鹿谷美緒子

3「unlimited」若松歓

4「青葉の歌」熊谷賢一

5「聞こえる」新美徳英

6「今日は君のbirthday」若松歓

7「決意」鈴木憲夫

8「初心のうた」信長貴富

9「そのひとがうたうとき」木下牧子

10「旅立ちの時～asian dream song」久石譲

11「地球の詩」三浦真理

12「春に」木下牧子

13「ヒカリ」松下耕

14「ひとつの朝」平吉毅州

15「ふるさと（嵐）」youth case

16「ぼくはぼく」三宅悠太

17「予感」大熊崇子

18「若い翼は」平吉毅州

19「君と見た海」若松歓
20「信じる」松下耕
21「名付けられた葉」飯沼信義
22「明日へ」富岡博士
23「翔る川よ」瑞木薫

2021年度の教職実践演習履修者は53名。参加グループは五つ。「1」グループ12名、「2A」グループ8名、「2B」グループ9名、「2C」グループ12名、「2D」グループ12名。人数に偏りあり。

各グループの自由曲（２B、２Cで同じ曲で競合があったがCグループが第二希望に変更）は以下のように決定。

1→旅立ちの時
2A→ふるさと
2B→信じる
2C→IN TERRA PAX
2D→初心のうた

グループごとの男声女声は1、2A、2Cにはそれぞれ3人ずつ男声が含まれ、2Bには男声一名のみ。2Dは女声のみ。2Bの「信じる」はメンバーの女声数人を男声の音域で歌うことで補い（一部編曲あり）、2Dは女声版を用いての演奏となった。

なお、課題曲「明日の空へ」は混声でも同声二部でも歌えることを選曲の理由としているため、グループごとの男声女声比は演奏に影響しないと判断している。

1-2　採点方法について

グループごとの練習に入る前（11月中旬）の段階で以下の連絡文を全員に送付している。

《連絡文》

模擬合唱コンクールの評価基準について

A　音程

B　ハーモニー、バランス

C　表現の工夫（強弱、速度の変化など）

D　歌詞の理解

声量、伴奏は単独の項目としては採点に含めません。（B、C、Dには影響する可能性はあります）

★グループによって人数も違えば、歌唱を得意とするメンバーの数も、ピアノを専門とするメンバーの有無も異なります。突出した個人のパフォーマンスが評価

に反映しにくいようにしてあります。

★実際の中学校での合唱コンクールでは、ドレスコードや声量、歌っているときの表情なども採点することがありますが、「模擬」ではそこは得点に反映させません。ただ、表現の工夫や歌詞の理解の一環として服装、表情、動きなどをパフォーマンスの一環として用いることは構いません（その場合、CやDの項目で評価される可能性はあります）。

★本番の合唱は先に課題曲、その後に自由曲を演奏します。自由曲を演奏する前に、グループが選んだ自由曲に関する短いスピーチをしてください。C、Dの項目への加点を左右しますのでしっかり準備してください。

模擬合唱コンクールが行われたのは1/9、昨年同様、一つの空間に多人数が入ることを避けるためグループごとに演奏時間を定め、持ち時間10分で課題曲、自由曲に関するスピーチ、自由曲、という順番で演奏、それを録画し、Webで公開。授業履修者は録画映像を後日視聴し、飢えに定めた四項目で採点する。

コンクール開催の二週間前、12月中旬に以下の《連絡文》を全員に送付し、「絶対評価と相対評価の二つの採点を求めることについて申し渡している。

《連絡文》

模擬合唱コン採点の方法について

（略）各グループの演奏は録画し、履修者全員が閲覧できるように公開します。

その動画を閲覧し、全員が審査員となって採点してください（自分の所属グループは採点しません）。その採点結果の提出をもって後半課題の完了と見なします。

採点の項目ですが、すでに公開している四つ、

A　音程

B　ハーモニー、バランス

C　表現の工夫（強弱、速度の変化など）

D　歌詞の理解

各々を**「6点満点」**で採点します。全部満点なら24点になります。

「表の」順位は、この各項目の点を合計して算出します。

この審査は「絶対評価」です。審査員それぞれが“このくらいならば◯点”という基準に照らして採点しますので、例えばBのハーモニー、バランスなら、「ともに文句なければ6点」「どちらかが今ひとつなら5点」「両方今ひとつなら4点」のようなルールを各自で設定して点をつけてください。

逆に言えば、「ハーモニー、バランスとも、すべてのグループが文句なかった」となれば全グループが6点になることもあります。

この「採点方法」私の研究としての意味合いもあるのでぜひ「自覚と誇りを持って」お願いします。

ただし、基準は個人の自由なので、ほとんどすべて6点出すような甘い点を付ける人と、良くても4点5点までしか出さず1点2点も容赦なく付ける人がまじります。全員の審査点を合計しますので、最終的には点の厳しい人によって順位が決まるケースが多くなります。

そこで！今年は「裏」の順位を出す試みにチャレンジいたします。

上記「音程」「ハーモニー・バランス」「表現の工夫」「歌詞理解」の各項目について、採点する4グループを比較して“順位”づけしてください。

例えば、あなたがAグループだとして、「一限」「B」「C」「D」の4つのグループを採点するわけですが、

音程　一位　一限、2位　C、3位　B、4位　D

ハモ・バラ　一位　D、2位　C、3位　B、4位　一限

…のように、項目ごとの順位を出します。

極端な話、音程はすべて「6」と点をつけたとしても、その中で順位を示してもらうことが狙いです。

これは「相対評価」になります。音楽演奏を相対的に評価したときと、絶対的に評価したときのばらつきを検証したいので、こちらもぜひしっかりお願いいたします。

採点に用いる「採点シート」はこちらで作成します。来週までには公開できると思いますのでお待ち下さい。

2. 模擬合唱コンクール結果

模擬合唱コンクール終了と同時に、以下の《採点フォーム》を全員に送付している。

《採点フォーム》

採点は二通り行います。このフォームの太字部分をコピーペーストして、チャットで送ってください。

◆その一

自分のグループ以外の4つに対し、以下の4項目を六点満点で採点してください。下の太字五行をコピペして、採点を書き加えてください。

1　　音程　H＆B　表工　詞理解

2A　音程　H＆B　表工　詞理解

2B　音程　H＆B　表工　詞理解

２Ｃ　音程　Ｈ＆Ｂ　表工　詞理解
２Ｄ　音程　Ｈ＆Ｂ　表工　詞理解

例（自分がＡグループだった場合）
１　　音程５　Ｈ＆Ｂ６　表工４　詞理解５
２Ａ　音程　　Ｈ＆Ｂ　　表工　　詞理解
２Ｂ　音程６　Ｈ＆Ｂ６　表工４　詞理解４
２Ｃ　音程３　Ｈ＆Ｂ５　表工６　詞理解６
２Ｄ　音程５　Ｈ＆Ｂ５　表工５　詞理解５

◆その二
項目ごとの順位を出してください。
下の太字四行をコピペして、審査対象の四グループを書き入れてください。

音程　　一位　二位　三位　四位
Ｈ＆Ｂ　一位　二位　三位　四位
表工　　一位　二位　三位　四位
詞理解　一位　二位　三位　四位

例（自分がＡグループだった場合）
音程　一位Ｂ　二位Ｄ　三位１　四位Ｃ
Ｈ＆Ｂ　一位１　二位Ｂ　三位Ｃ　四位Ｄ
表工　一位Ｄ　二位Ｃ　三位Ｂ　四位１
詞理解　一位Ｃ　二位１　三位Ｄ　四位Ｂ

参加者53名のうち、51名が審査結果を提出。以下、その審査結果を示す。

2-1　採点結果（絶対評価）

破線で囲まれる四列が一つのグループの審査結果となる。項目は左から音程、ハーモニー・バランス（HBと略）、表現と工夫（表工と略）、歌詞の理解（詞理解と略）。総計は51名の審査点の各項目別合計。グループ成因が8～12名なのでデータ数で割って平均を出し、その平均の高いものから順に①～⑤で項目別順位を出している。

グループ	1				A			
審査項目	1音程	1HB	1表工	1詞理解	2A音程	2AHB	2A表工	2A詞理解
総計	143	145	157	177	188	189	162	195
データ数	39	39	39	39	45	45	45	45
平均	3.666667	3.717949	4.025641	4.538462	4.177778	4.2	3.6	4.333333
順位	⑤	⑤	④	③	④	④	⑤	⑤
順位計	17				18			

B				C			
2B音程	2BHB	2B表工	2B詞理解	2C音程	2CHB	2C表工	2C詞理解
209	194	195	191	211	179	200	175
42	42	42	42	39	39	39	39
4.97619	4.619048	4.642857	4.547619	5.410256	4.589744	5.128205	4.487179
②	②	②	②	①	③	①	④
8				9			

D			
2D音程	2DHB	2D表工	2D詞理解
187	196	177	192
39	39	39	39
4.794872	5.025641	4.538462	4.923077
③	①	③	①
8			

最下行にある「順位計」は各項目の順位（丸数字）を合計した値で、これが少ないほど高く評価されたと判断できる。

この「順位計」で判断できる順位は、BD→C→1→Aになる。

2-2　採点結果（相対評価）

こちらの総計は、各評価項目の一位を4点、二位を3点、三位を2点、四位を1点と換算したものの合計になる（自グループ以外の4つを評価するため四位までになる）。

グループ	1				A			
審査項目	1音程	1HB	1表工	1詞理解	2A音程	2AHB	2A表工	2A詞理解
総計	51	58	80	94	89	95	70	89
データ数	39	39	39	39	45	45	45	45
平均	1.307692	1.487179	2.051282	2.410256	1.977778	2.111111	1.555556	1.977778
順位	⑤	⑤	④	③	④	④	⑤	⑤
順位計	17				18			

B				C			
2B音程	2BHB	2B表工	2B詞理解	2C音程	2CHB	2C表工	2C詞理解
126	119	131	124	136	109	127	92
43	43	43	43	39	39	39	39
2.930233	2.767442	3.046512	2.883721	3.487179	2.794872	3.25641	2.358974
②	③	②	②	①	②	①	④
9				8			

D			
2D音程	2DHB	2D表工	2D詞理解
107	128	101	118
38	38	38	38
2.815789	3.368421	2.657895	3.105263
③	①	③	①
8			

平均を見ると、最も高い数値であるCグループの音程は3.49点でほぼ一位として推されていることがわかる一方、最も低い数値である1グループの音程は1.30点でほぼ五位の評価が固定していることがわかる。

各評価項目ごとに一位から五位までを丸数字で示してある。そしてその合計を「順位計」として最下行に示したが、これに基づくと順位はCD（同点）→B→1→Aとなる。
以下、これらのデータに基づいて考察する。

3. 考察

3-1　順位判断の基準としての「順位計」

順位計に基づいて順位を出すと、

絶対評価　BD→C→1→A

相対評価　CD→B→1→A

何が変わるかといえば、CとBの順位である。この違いをもたらした要因を、各項目ごとの順位に求めることができる。

1グループとAグループの各評価項目合計の順位は、絶対評価でも相対評価でも動いていない。音程、HB、表工、詞理解の順に1グループは⑤⑤④③、Aグループは④④⑤⑤で順位計も17、18と変わっていない。つまり、この二つのグループにおいて、絶対評価で審査しても相対評価で審査しても、今回の五グループの中では四位、五位との評価が妥当であるといえる。

各評価項目合計に関してはDグループも絶対③①③①、相対③①③①となっており、不変である。非常に高い評価で、絶対評価、相対評価双方で（同点であるが）一位と評価されている。

対して、B、Cの二グループは絶対評価と相対評価で各評価項目合計順位で移動がある。Bグループは絶対②②②②、相対②③②②、Cグループは絶対①③①④、相対①②①④、二番目のHB（ハーモニー・バランス）の評価項目において順位の変動があった。

ここで、B、C両グループの絶対評価での点を見てみる。B、Cグループのメンバーは自分のグループを採点していないので、1グループ、Aグループの採点だけを見ると

2BHB	5	6	6	5	5	3	4	5	4	5	5	4	4	5	4	4	4	3
2CHB	6	5	5	4	5	4	3	4	3	3	4	5	5	4	5	5	4	2

この18データとなる。BよりCが優る、と採点したもの（橙セル）6名、CよりBが優る、と採点したもの（黄緑、緑セル）10名。うち一名は二段階の差をつけている（緑セル）。つまり、絶対評価においてBのほうがCよりもHBに優るという採点が多かったことになる。ここに、HBにおいては全評価者から「最も優れている」と認められたDグループの採点を参照するとこうなる。

2BHB	5	5	5	5	5	6	5	4	5	5	3	3
2CHB	6	6	5	6	5	5	5	6	5	5	5	5

BよりCが優れる（橙、黄セル）が6名、逆にCよりBが優るとした一名を遥かに上回る。注目すべきは二段階の差をつけたもの（黄セル）が三人いる、ということである。

相対評価の場合、甲より乙が優れる、と判断したとして、その「優れ具合」は評価の中

に入らない。絶対評価はその優れ方の差異を数値に置き換えることが可能で、さらに、その突出の度合いをどうやって数値に乗せるかは完全に評価者の主観となる。評価者の数が増えれば増えるほど、絶対評価のほうが「評価のために利用される数値」の分散が大きくなる。

なお、HBにおいて明らかに優れていると認定されたこの3グループをすべて評価したデータ（1グループ、Aグループのみ）でBCDどのグループをHBにおける最上位と判定したかを表で表すとこうなる。

2BHB	5	6	6	5	5	3	4	5	4	5	5	4	4	5	4	4	4	3
2CHB	6	5	5	4	5	4	3	4	3	3	4	5	5	4	5	5	4	2
2DHB	6	6	5	5	6	5	5	4	6	6	6	4	4	5	4	6	3	5

18データ中12が、Dを最も優れているとする一方、同点で推す場合はBを評価しているのが3データある。BとCが同点でDを一段低く評価するものが1データ、CとDを同点としたものが1データとなる。あえて意味づけるのであれば、今回の採点者にとってBとDは同じ方向性のハーモニー構成を感じさせており、Cはそれとはアプローチが違う音楽であると受け取られた可能性を感じる。

ここには数値は示さないが、CのHBには非常に低い評価が混在していることも特徴である（2データ）。採点者によっては評価がばらつくのがCグループの特徴とも言える。

以上のデータから判断できることは、CグループとBグループはハーモニーやバランスに関して方向性の異なるアプローチから音楽を作ったため、評価者による点のばらつきやすい絶対評価ではその「差」を大きいと判断したものによりBが優り（Cがマイナスに評価され）、相対評価ではどちらが魅力的かと迫られた時に不完全性よりもパフォーマンスのエネルギーでCが優った可能性が指摘できる、となろう。

3-2　評価項目ごとの標準偏差から見る絶対評価、相対評価

評価項目ごとに、グループの評価数値を合計した一覧を、絶対評価と相対評価で比較してみたい。

◆音程項目について

	絶対評価					相対評価				
	1音程	2A音程	2B音程	2C音程	2D音程	1音程	2A音程	2B音程	2C音程	2D音程
総計	143	188	209	211	187	51	89	126	136	107
データ数	39	45	42	39	39	39	45	43	39	38
平均	3.67	4.18	4.98	5.41	4.79	1.31	1.98	2.93	3.49	2.82
順位	5	4	2	1	3	5	4	2	1	3
標準偏差	0.94	0.97	0.94	0.67	1.02	0.51	0.98	0.9	0.75	0.88
					0.92					0.91

下から二行目が標準偏差になる。最下行は標準偏差を平均した数値で、興味深いことに絶対評価も相対評価もだいたい同じ数値（0.9近辺）に収束した。これはかなり意外で、絶対評価は最大が6で最小が1、対して相対評価は最大が4で最小が1となり（自グループは採点しないため）、そもそもの数値区間の大きさが違う。これは、絶対評価では「1」がほぼ使われないため、実質的に6～2の5段階評価になると考えた。これに対し相対評価では、必ず各項目に1から4が用いられるので相対評価の標準偏差は絶対評価と比べてかなり小さくなることが想定されていた。

ところが実際は、両者に差が殆ど出ていない。これは、あまりに表が小さくなるため全体を本稿では示すことをしないが、2の使用頻度が非常に低く、17回しか用いられていない（51人、各16項目採点で総項目数816中）ためである。 これは2％に過ぎず、絶対評価が実質的に3、4、5、6の四段階評価になってしまい、相対評価との差がなくなったものであると結論する。

想定からは外れたが、そのため絶対評価と相対評価の標準偏差を同質の数滴データとして扱うことができることとなった。

さて、大きな特徴はCグループにある。標準偏差が偏差平均を大きく下回り（黄セル）、相対評価、絶対評価とも「Cの音程は優れて美しい」と評価されていることが示されている。

反面、標準偏差が大きく分散した（1.02、赤セル）のは絶対評価のDグループで、音程が優れていると感じるものと、劣っていると感じるものの二派に分かれる評価になったことを示している。それなのに相対評価のDグループはほぼ分散具合が平均になっており、妥当な順位になったと判断できる。

逆の意味で突出しているのが1グループ。絶対評価ではほぼ平均となっており、評価得

点はそれなりにばらついている（ばらつき方は平均的）。しかし相対評価になると標準偏差は0.51（黄緑セル）となり、これはもうほとんどこの順位以外はない、との共通見解が示されたことになる。

ここから言えるのは、音程（の正確さ、美しさの）順位を語る際には相対評価の方が「有無を言わせぬ結果」が示せる、ということであろう。今回の模擬実践では、音程に関しての順位は絶対評価、相対評価とも同じであったが、絶対評価の数値を用いることで「順位はこうであったが、いろいろな見方があった」ということを、特にDグループにはポジティブなフォローとして用いることができる。逆に、1グループには「五つのグループを並べると五番目にはなるけれど、あなた方が劣っているわけではない」という否（アンチ）ネガティブなフォローとして用いることができる。

→相対評価は順位決定に、絶対評価はフォローアップに効能あり

◆ハーモニーとバランスについて

	絶対評価					相対評価				
	1HB	2AHB	2BHB	2CHB	2DHB	1HB	2AHB	2BHB	2CHB	2DHB
総計	145	189	194	179	196	58	95	119	109	128
データ数	39	45	42	39	39	39	45	43	39	38
平均	3.72	4.2	4.62	4.59	5.03	1.49	2.11	2.77	2.79	3.37
順位	5	4	2	3	1	5	4	3	2	1
標準偏差	0.78	0.98	0.82	1.06	0.86	0.71	0.99	0.88	1.07	0.9
					0.9					0.91

この項目は、前項「順位計」でも述べたが、相対評価、絶対評価ともほぼ同じばらつきを示し、特に1グループの順位に関しては両評価の標準偏差が一致して偏差平均を下回り（黄セル）「問題を含む」ことを指摘している。

一方、BとCは絶対評価と相対評価で順位が変動しているが、特にCグループは絶対評価の数値のみならず、相対評価の数値でも標準偏差が1を超え（赤セル）、賛否が割れていることが特徴である。これは、両評価の数字をともに示し、Cグループへの称賛に代えることができると考えたい。「順位は三位だが、あなた方はものすごく支持される一方、これは嫌だと感じる人も多かったのだ」。順位以外の達成感をもたらすことを期待できる。

→データのばらつきの大きさは、絶対評価では起こりがちだが、相対評価で起こることは稀であり、両者が連動している時は表彰の対象とできる可能性あり

◆表現の工夫について

データのばらつき方はごく標準的、いずれの順位も「妥当」だと受け取れるが、唯一Dグループへの点の出方が特徴的。絶対評価の偏差平均0.86に対し0.98(赤セル)の標準偏差。相対評価の偏差平均0.91に対し1.11(赤セル)の標準偏差。両者とも、平均を大きく上回り、評価が割れたことを示している。

	絶対評価					相対評価				
	1表工	2A表工	2B表工	2C表工	2D表工	1表工	2A表工	2B表工	2C表工	2D表工
総計	157	162	195	200	177	80	70	131	127	101
データ数	39	45	42	39	39	39	45	43	39	38
平均	4.03	3.6	4.64	5.13	4.54	2.05	1.56	3.05	3.26	2.66
順位	4	5	2	1	3	4	5	2	1	3
標準偏差	0.86	0.9	0.78	0.79	0.98	0.96	0.83	0.83	0.84	1.11
					0.86					0.91

これは「表現の工夫」という項目そのものの抱える問題でもある可能性がある。Dグループには男声が含まれず、純然と女声合唱になったため、音の響き方が他のグループと異なり…有り体に言って女声合唱を聞き慣れていない評価者の評価点を割ったと考える。

ここで筆者の主観を混ぜることは本意ではないが、Dグループの合唱の完成度はかなり高かった（これは練習からの観察を通しても明らか）と評価しているので、ここで「三位」になったことには驚いたのだが、採点結果を分析したことでこの結果が出た理由が納得できた。ただし、演奏者に対してこの順位だけを返すのはかなり不親切だろう。あえて今回のデータを用いるのであれば、標準偏差が大きかった（評価が大きく割れていた）ことを示し、評価の数字には現れない「聞き手の状態」を理解する助けにはできると考える。

→絶対評価と相対評価の標準偏差がともに大きい場合、問題は審査される側にあるとは限らない

◆歌詞理解について

	絶対評価					相対評価				
	1 詞理解	2A 詞理解	2B 詞理解	2C 詞理解	2D 詞理解	1 詞理解	2A 詞理解	2B 詞理解	2C 詞理解	2D 詞理解
総計	177	195	191	175	192	94	89	124	92	118
データ数	39	45	42	39	39	39	45	43	39	38
平均	4.54	4.33	4.55	4.49	4.92	2.41	1.98	2.88	2.36	3.11
順位	3	5	2	4	1	3	5	2	4	1
標準偏差	1.08	1.01	0.91	1.11	0.8	1.13	1.02	0.99	1.19	0.94
					0.98					1.05

今までの三評価項目との大きな違いは、偏差平均が大きいことにある。およそ0.1ほど分散が大きくなっており、更にグループごとの標準偏差も偏差平均を上回るもの（赤セル）が多く出現している。常に賛否が割れてきたCグループのみならず、下位として高く評価されてこなかった1グループの評価が割れている。

実を言えば「歌詞理解」という項目は《評価をばらつかせる》効果を狙って設定したとも言える視座で、標準偏差が大きくなるのは狙い通りであった。ただし、音程やハーモニーに難あり、とされた1グループが割れたのには、「表現と工夫」と似た背景がある。1グループは課題曲「明日の空へ」の上パートを男声が、下パートを女声が歌うという「工夫」をしていた。しかし、これによって女声が上パートを歌う際には伸びる高音が潰れ気味に聞こえたのではないかと予想する。これも、練習段階での観察を通し、最も詞を読み込んでいたのは1とDだ、と評価していたので、ここで評価が割れるということは、一回の演奏を聞いただけでは歌詞の理解云々まで求めるのは困難か、と感じた。

Cグループが割れたのにも理由はあり、録画する際に「課題曲」「自由曲に関するスピーチ」「自由曲」という順番で演奏するところ、スピーチを飛ばしてしまい、自由曲を歌い終わったあとでスピーチが入ったというハプニングがあったため。録画視聴だったので、巻き戻し再生は可能だとはいえ、歌う前に「こういう工夫をしました」と聴くのと、歌を聞いてしまった後に詞をどう解釈して歌ったかを知らされるのではやはり評価は割れると思われる。

この評価項目からは、データとして絶対評価、相対評価の差異を読み取るには十分ではなかったと考える。

3-3　評価者の採点傾向から見る絶対評価、相対評価

相対評価の出方には、評価者個人による特徴があるのか。今回の調査では、項目ごとの相対順位は求めているが「総合順位」は求めていない。そこで、「各評価項目の一位を4点、二位を3点、三位を2点、四位を1点と換算」という2-2の算出法を用いて各評価者がそれぞれのグループにだした四つの評価項目順位を点数換算し、それを合計、4で割ってグループごとの平均順位を算出した。

このやり方だと、例えば四項目すべて一位、と評価されていれば合計は16点となり、平均は4点。読ん項目が一位、二位、三位、四位、とばらついていれば合計は10点となり、平均は2.5点となる。

この各グループの項目別順位点を合計平均した値を一覧にするとこうなる（次ページ左）。橙セルがその評価者が一番高く評価したグループであり、総合評価の一位と言えるものになる。赤セルはその一位グループが二つあるもの、臙脂セルは3つあるものになる。同様に絶対評価の項目別採点を合計した値の一覧がその右の表になる（次ページ右）。

相対評価による各チームの総合順位は、D→16票（橙）、3票（赤）1票（臙脂）で最多、以下、B→13橙3赤1臙脂、C→13橙2赤1臙脂で赤票一つの僅差でBが二位となる。これ（D→B→C）は相対評価の順位計による結果（CD→B）とは異なる。 注目できるのは少数とはいえ、順位系では全く「いいところ」が見えなかった1グループが2橙1赤、Aグループが1橙1赤と上位に推されていることが見える点であろう。順位計だけでは見えない「成果」はこうした分析結果を併せて示さない限り演奏者には伝わらない。

絶対評価による総合順位はD→12橙5赤1臙脂、BC→橙11赤8臙脂1となり、赤、臙脂をどう解釈するかで順位が変わる。仮に赤0.5、臙脂0.33と読むと、BとCがともに15.33で、14.83のDを逆転することになる。これ（BC→D）も順位計による結果BD→Cとは異なる。ここから、評価者個々の採点結果を用いる場合と、項目別合計の順位（及び順位の計）を用いた場合とでは判断できる順位が異なる、ということができる。

さて、次ページの左右の票を見比べると、多くの評価者は絶対評価、相対評価とも五つのチームの順位は変動していないことがわかる。これは51評価者中34人に言える。しかし、残り17人は絶対評価と相対評価では順位の出方が異なるのだ。多くは一位グループが複数→単独、もしくはその逆という異動だが、この17人中、「総合一位」が変わるのは二名のみである。

←相対

1	A	B	C	D
1.75	2	3	3.25	
	1.75	2	3.75	3.25
1.5	2		3.25	3.25
2.25	1.5	3.5	2.75	
	1	3.5	3.25	2.25
1		3.25	2.5	3.25
2.25		2.5	2.5	2.75
	2	3.5	2.5	2.5
2.25	1.5	3	3.25	
1.5	1.5	3		4
1.75	3	3.5	1.75	
2.25	2.5		3.25	2
1.5	1.5	3.5	3.5	
2.5	2		2.5	3
1.75	1.5	3.75	3	
2.75	2.75		2	2.5
1	2.5	3		3.5
1.75	3.25	2.5	2.5	
2	2.25	2.5		3.25
2	1.5	2.75	3.75	
	1.75	1.75	2.75	3.75
	2.75	1.5	2	3
1.5	2.5	2.25		3.75
	1.75	3.5	3.25	1.5
1.25		3.25	3	2.5
1.75	2.25	2.5	3.5	
2		3	2.25	3.5
	1	3	3	3.25
1.5	1.5	3.75		3.25
1	2.5	3		3.5
1.5	2.75	3.25	2.5	
1.5	2.75	3.25	2.5	
	1	2.5	4	2.5
	1	2.5	4	2.5
1.75	2		2.5	3.75
1.25	1.75	3.5		3.5
1		3	3	3
1.75	1.25	4		3
1.25	2		2.75	4
	1.5	3	4	1.5
2.25	2.25	2.25	3.25	
3	2.25		2.25	2.5
	1.25	2.25	2.75	3.75
2.5	2.5	1.5	3.5	

←絶対

絶対・1	絶対・A	絶対・B	絶対・C	絶対・D
21	20	23	23	0
0	20	20	24	23
20	20	0	23	23
20	20	0	23	21
0	18	22	22	21
17	0	23	20	22
19	0	21	20	21
0	19	22	21	19
20	17	22	22	0
18	18	21	0	24
18	20	23	18	0
17	20	0	21	20
17	17	22	22	0
19	18	0	20	20
19	17	21	20	0
18	20	0	18	19
16	19	18	0	22
15	21	20	19	0
18	19	18	0	19
18	16	21	19	0
0	16	16	20	22
0	18	17	18	20
15	17	19	0	22
0	16	20	20	17
15	0	20	19	19
15	18	18	21	0
16	0	17	17	22
0	12	20	20	20
17	16	19	0	19
13	17	21	0	20
16	19	18	17	0
16	19	18	17	0
0	12	18	23	17
0	11	17	22	19
14	16	0	16	21
12	15	19	0	21
13	0	18	18	17
14	14	20	0	18
12	13	0	19	22
0	14	16	20	15
16	15	14	19	0
17	16	0	15	16
0	10	16	18	20
15	16	14	18	0

表中に見える「0」は計算するために入力されたもので、本来は空欄である。

1音程	1HB	1表工	1詞理解	2A音程	2AHB	2A表工	2A詞理解	2B音程	2BHB	2B表工	2B詞理解	2C音程	2CHB	2C表工	2C詞理解	2D音程	2DHB	2D表工	2D詞理解
4	4	5	5	4	4	4	4	5	5	6	5	5	5	6	3				
3	4	3	3	5	5	3	4	5	5	5	6					4	5	5	6
1	2	2	3	2	1	1	2	3	3	4	1	4	4	3	4				
1	1	1	1	3	3	2	2	4	2	3	3					2	4	4	4

絶対・1	絶対・A	絶対・B	絶対・C	絶対・D	1	A	B	C	D
18	16	21	19	0	2	1.5	2.75	3.75	
13	17	21	0	20	1	2.5	3		3.5

上段の評価者において、絶対評価ではBが21点、Cが19点でBが優位。しかし相対評価ではCが突き抜けて一位であるという評価となり、総合一位が動いている。

下段の評価者において、絶対評価ではBが21点、Cが20点でBが優位。しかし相対評価ではDが音程以外ではすべてBを上回り、こちらも総合一位が動いている。

これは「無理矢理相対評価をさせている」ことから起こるのであり、もっと多くのデータが現れるものと予想したが、この2件のみであった。絶対評価では同じ「5」と判定した項目を、「同じであってもどちらかが優れている、とせよ」と申し渡したために起こったことだと結論できる。

視点を変えると、他の評価者においてはこれが起こらなかったということになり、特に「絶対評価においての採点」は演奏の総合評価との乖離がなかったことの証明と言える。

4. まとめ

本稿の狙いである、絶対評価と相対評価では順位に用いる妥当性が変わるのか、という部分については、十分な考察には至らなかった。順位とは相対的なものであり、相対評価

が優る、という想定であったが、思った以上に評価者（大学生）の絶対評価が《的確》で、相対的な優劣をそのまま6段階評価の評点が反映させていたとも言える。

ただし、絶対評価の場合、評価者によって評価のスタンスが全く異なる。まずは点の甘さ、辛さ。平均して5.5点くらいの点を出す（評価は6と5が半々）評価者もいれば、3.5を下回る平均の点をつける評価者もいる。何よりこの研究に着手して以来の懸案となっている「6段階評価なのに実質的に5段階」「実質的に4段階」という分散の小ささが個々に至って無視できないことが再認識された。1から6まですべて使った評価をしたのは1名のみである。そして、その評点平均は最下位、最も厳しい評価者となっている。

こうした評価者相互で採点基準が違うことは、むしろ絶対評価にとって好ましいことだと考えたいが、頭を抱えたくなるのが分散の小ささ。標準偏差1未満の評価者が36名と3分の2以上の率を占める。あくまで大まかな傾向であるが、標準偏差の小さな（ばらつきの小さい）評価者には高めの点を出すものが多かったことが指摘できる。標準偏差0.7未満の12名のうち、評点平均が20位までに入らなかったのは一名のみであった。

非常に都合の見通しではあったのだが、今回の調査で「甘めの点を出し、ばらつきの小さな評価をする一群」を棄却データとするべき根拠が得られれば、と考えていたのだが、「大まかに」としか論ずることができないということは、確たる根拠は得られていないと言わざるを得ない。

順位算出のためのデータとしては問題がある評価者群ではあるかも知れないが、得られた採点結果を精査することで演奏者に「演奏の成果」を順位以外の側面から感じさせる事の可能性は示せたと考えたい。特に、「相対評価一本」での順位算出は、突き詰めれば人気投票に過ぎない。その根拠がどこにあるかを、絶対評価との併記で可能にできるかも知れない、という目標はこのまま見つめ続けたいと思う。

継続研究でもあり、次回は評価者の数を増やし、その信頼性をあえて落として棄却データについて考える、もしくは、6段階評価が実質的に5段階、4段階となることをなくすため、「持ち点制」評価を試みる、といった展望を抱えつつ、本稿を閉じる。

末筆ながら、録画視聴方式とはいえ、十数人による合唱が可能となったことに光を感じた。感染症による音楽活動への制限が解消されることを強く祈る。