成果評価基準の公正性

～フィギュアスケートにおける主観的評価の検証

公益財団法人日本住宅総合センター　行武　憲史[1]
慶應義塾大学大学院経済学研究科　藤野　玲於奈[2]

要旨

成果主義の導入が進む中，公正で正確な評価制度の重要性は増している．本研究では，人事評価で重要視される主観的評価におけるバイアスの存在について，フィギュアスケートの世界大会のデータを用い検証した．主観的な構成点と客観的な基礎点のそれぞれについて，選手の能力を制御した上で，滑走グループ間における得点差の有無について回帰分析を用いて分析した．その結果，構成点ではグループ間の差が確認される一方，基礎点では確認されなかった．これは，構成点という主観的評価に審査員の選手への期待によって形成されるバイアスが含まれることを示している．

JEL 分類番号：　J710 , Z290, Z280
キーワード：Bias, Competition Policy

[1] 行武憲史：公益財団法人日本住宅総合センター主任研究員，日本大学経済学部非常勤講師，yukutake@hrf.or.jp
[2] 藤野玲於奈：慶應義塾大学大学院経済学研究科修士課程 1 年，fujino.reona@keio.jp

1. はじめに

1.1. 研究の背景と目的

平成 27 年 4 月に閣議決定された労働基準法改正案では，脱時間給」制度を新設するほか，大企業を中心に 90 年代半ばから導入がすすんできた裁量労働制の対象拡大などが盛り込まれている．成果による管理制度を導入する前提として， 公正かつ正確な評価が重要である．しかし，評価の過程にさまざまなバイアスが存在することが，心理学を中心に多くの研究で指摘されている．こうしたバイアスの存在やバイアスへの対処の欠落は，不公平感を招き従業員の勤務態度などを通じて非効率性の問題をもたらす．

スポーツやオーディションといった分野のデータは，客観的なパフォーマンスと審判による主観的な評価の組み合わせが比較的容易に利用できるため，成果の評価の分析に多く用いられている．本研究では，フィギュアスケート競技における，採点の公正性の分析を通じて評価システムのバイアスについて検証する．なかでも，Findlay and Ste-Marie (2004)で示されている期待バイアスに焦点を当て，2005 から 2015 年までの世界選手権とオリンピック過去 3 大会のシングル男女の採点データを用いて分析を行う．

フィギュアスケートの採点は，技術点と構成点，減点に分類される．技術点は，選手が実行した技の達成度を明確な基準の下評価するのに対し，構成点はかつて，芸術点（表現力）といわれたものであり，主観的要素によるところが大きい．本研究では，構成点の評価に着目し評価バイアスの検証を行う．

1.2.先行研究

Morgan and Rotthoff (2014)では，フィギュアスケートなどの採点を伴う競技会におけるバイアスを，期待バイアス，内集団バイアス（国籍バイアスなど），順序バイアス，難易度バイアスという 4 つのバイアスに分類している．本研究では，このうち，審査員の主観が顕著に表れると考えられる期待バイアスに注目し検証を行う．

期待バイアスとは，明確な判断基準がない場合に，評価以前に得られた選手の名声や情報により発生するバイアスである．このバイアスは，Tversky and Kahneman(1974)が示した利用可能性（Availability）ヒューリスティックスと同様のものと考えられる．ヒューリスティックスとは，判断を行う際の情報経路のショートカットといった直感的な判断方法と定義される．利用可能性ヒューリスティックスは，評価に際して，ある選手の演技の得点をその状況が頭に思い浮かびやすい情報（期待）を優先させて判断する状況を指す．

Findlay and Ste-Marie (2004)は，フィギュアスケート競技における期待バイアスの存在を，カナダのケベック州とオンタリオ州における大会のデータを用いて検証している．その結果，事前に選手の名声や情報を知っていることが，その選手に対する過度な肯定的評価を導くという結果が示している．

2. フィギュアスケートについて

2.1. フィギュアスケートの採点方法

現在,フィギュアスケートにおけるシングルのプログラムは,ショートプログラム(以下,SP と記す)とフリースケーティング(以下,FS と記す)で構成されており,先に SP が行われる．SP は，演技時間が 2 分 50 秒以内と短く，ジャンプやスピン等の 7 個の必須要素がある．FS は，演技時間が 4 分 20～40 秒と長く，13 個の要素から構成される．

フィギュアスケートの得点は，①技術点，②構成点，③減点から構成される．これら 3 種類の得点が，SP，FS ごとに集計され，さらに両者の合計がその大会の総得点となる．

技術点とは，選手が実行した各規定要素に与えられる得点の合計であり，基礎点と GOE から成り立っている．基礎点とは実行した技の基礎的な評価のことである．各要素の開始時の動作(ジャンプの踏切りなど)，ジャンプの回転数，演技のレベルといった要素からなる.基礎点は,最多で 3 名の技術審判によって採点され,映像を使って細かいポイントのチェックも行われるため，比較的明確な基準をもつと考えられる．

GOE(Grade of Execution)とは各要素の出来栄えに対する加点や減点のことである．0 を基準とし要素の出来栄えによって，7 段階で評価する．GOE は演技審判によって判定され,加点基準は少し抽象的であるが，減点基準は具体的に定められている．したがって，転倒や回転不足などの場合には，比較的客観的に点数がつけられていると考えられる．

構成点(Program Component Score)は，演技審判が 5 項目（スケートの技術，要素のつなぎ，動作・身のこなし，振り付け・構成，曲の解釈）をそれぞれ 10 点満点で採点する．

構成点の 5 項目は，技術点に比べると基準が不明確であるため，審判員の主観的な判断が入り込みやすいと考えられる．たとえば，曲の解釈という項目は，審判員自体の印象に左右される可能性が高い.

最後に減点は，転倒や落下，演技時間の過不足，ボーカル入り音楽の使用，衣装や小道具の違反，禁止要素の違反といった違反行為等による減点である．

上記のように，基礎点は採点基準が明確であり，審判員の主観的な判断の影響は小さくと考えられ，一方で構成点は採点基準が明確ではないため，審判員の主観的判断が入る可能性がある．GOE は減点部分に関する採点基準は明確であるのに対し加点部分に関する採点基準は明確ではないので，審判員の主観的判断がスコアに与える影響の予測は難しい.

3. 仮説と実証モデル

本研究で想定する期待バイアスは，実際はあまり良くないのに，期待によって高い評価がなされるというバイアスであり，評価にプラスのバイアスをもたらす．フィギュアスケートの世界選手権やオリンピックでは，SP の滑走順は直前の世界ランキングにより，FS の滑走順は SP の順位によりそれぞれ決定され,上位の選手が後のグループへと振り分けら

れる.グループ内における順番は,ランキングや順位に関係なくランダムに割り振られる.このとき，より高い能力がある（と思われる）選手が集まる後半のグループの選手ほど，審判の思い込みにより高く評価される可能性がある．ここでは，この審判の思い込みを期待バイアスとして捉え，グループ間に生じるグループバイアスとして定義する.

世界ランキングやSPの順位でグループ分けされるため,当然のことながらバイアスが存在しなかったとしても，選手の能力そのものによってグループ間の平均得点には差が生じる．しかし，能力の高い選手ほど高い期待によるバイアスの恩恵を受けるとするならば，グループ間の平均得点差は能力以上に拡大すると考えられる．さらに，このグループバイアスには，選手個人に対する期待のバイアスだけでなく，最終グループだからみんな素晴らしいだろうといった，順序バイアス的な要因も含まれると考えられる．したがって，選手の能力をコントロールした後でも，平均得点のグループ差が確認されればバイアスが存在するといえる.

さらに，グループごとに平均的な得点差を発生させる要因としては，選手の本来の実力および，審判による採点のバイアスのほかに，グループ内における演技者のパフォーマンスが相乗（連鎖）的に他の演技者に影響するような効果も考慮する必要がある．前の選手が良い演技をした場合に，よりよい演技が引き出されるようなケースである ． こうした選手間の相乗効果と審判によるバイアスの識別には，基礎点と構成点における採点基準の客観性の違いが役に立つ．相乗効果の発生は，技術的な側面と芸術的な側面の両方を向上させると考えられる．このとき，グループ間の不連続性は，基礎点，GOE，構成点のそれぞれで発生する．一方で，審判による採点バイアスは，主観性の高い構成点でより大きくなると考えられる．採点基準が明確でない構成点にのみ不連続性が確認されれば，そこには明らかに審査員の主観によるグループバイアスの存在することになる.

今回の分析に当たっては，以下の(1)式のような評価関数を考えている.

$$S_i = \alpha + \sum_{j=2}^{4} \beta_j D_{ij} + \gamma N_i + \sum_{k=1}^{K} \delta_k X_{ik} + u_i \tag{1}$$

被説明変数であるS_iは演技iの対象スコアを表し，ここでは基礎点，GOE，構成点を考慮する．D_{ij}はグループダミーを表しており，添え字jがグループ（$j = 2,3,4$）を表す．ここでは，第1グループを基準として考えている．N_iは，国籍バイアスを顧慮したダミー変数で，選手と同じ国籍の審判が審判団に含まれている場合に1をとる．X_{ik}は，個人属性を表しSPの成績，直前のシーズンベストおよびそれらの高次の項を含む．最後に，u_iは誤差項である．また，α，β_i，γ，δ_kは各パラメータを表す．添え字は，iが選手個人の大会ごとの演技，kはk番目のその他の説明変数を表す（$k = 1,\ldots,K$）.

グループダミーについては，第1グループから最終グループまでに対応するダミー変数

で，第 4 グループダミーは，FS の滑走順が 19 番目から 24 番目の選手であれば 1 をとる．ただし，2005・2006 年の世界選手権では第１・２グループは分割されていないため，双方とも第 1 グループダミーに分類している．

グループ分けは SP の順位によって決められるため，選手の能力と高い相関を持つ．グループダミー単独で回帰分析をした場合，後半のグループほどスコアが高くなると考えられる．したがって，審判による期待バイアスの影響をグループダミーの係数によって検証するためには，選手のもともとの能力をコントロールする必要がある．個人の能力の代理変数として，SP の成績と直前のシーズンベストを説明変数として導入する．また，必ずしも能力と得点が線形の関係にあるとは限らないため，これらの高次の項を入れ，選手の能力をコントロールしている．各変数のレベル項と GOE の高次項の選択は，有意水準 10%を基準に有意なものを選択した．

4. 分析結果

推定の結果(表 1)，男子のグループダミーでは基礎点の第 2 グループの係数が 10%水準で有意なことを除いては，基礎点，GOE ではグループダミーが有意ではない．一方で，構成点ではグループダミーが 1%水準で有意に効いている．選手の能力をコントロールしてもなお，第 1 グループと比べて，各グループ間で差が生じている．構成点にだけグループダミーが効いている一方で，基礎点，GOE では明確な有意性はみられないため，相乗効果ではなくグループバイアスがあると判断できる．

女子の場合も，基礎点と GOE ではグループダミーが有意でなく，構成点ではグループダミーが 1%水準で有意に効いている．そのため女子でも，構成点のみグループダミーが確認されるため，グループバイアスがあると考えることができる．

推定式では，基準としている第 1 グループとの有意差のみをみており，他のグループ間のバイアスを確認したわけではない．そこで，各グループ間の差について F 検定(表 2)を行った．その結果，男子の基礎点において第 2 グループと第 4 グループの間に 5%水準で有意な結果が示されたことを除いては，技術点および GOE の他のグループ間の差については，男女ともグループ間の得点差は確認できなかった．一方で，構成点についての差については，すべてのグループ間で少なくとも 10%水準で有意な差が確認された．

以上の結果から，構成点にはグループ間の差がみられ，それ以外のスコアには系統だったグループ間の差がみられなかった．これは，グループダミーにグループ内で生じる相乗効果が含まれておらず，構成点について期待バイアスが存在する可能性を示唆する．

5. 結論

本研究では，評価指標の主観性に焦点をあて成果評価で発生するバイアスについて，フィギュアスケートの成績評価データを用い期待バイアスについて検証を行った。その結果，

審査員の選手への期待によって形成されるグループバイアスについては，構成点についてのみ，その存在が確認された．評価基準がより明確である技術点の基礎および GOE ではこうした傾向はみられず，構成点のようにより主観的な基準に対する採点行動においてバイアスが生じる可能性が高くなる．こうした分析結果は、我が国でも成果主義賃金の導入が進んでいるなか、公正かつ正確な人事評価を考える上で主観的基準の持つバイアスは無視できない問題であることを示すものである。

表 1　推定結果

	Men						Women					
	基礎点		GOE		構成点		基礎点		GOE		構成点	
説明変数	係数	標準誤差	係数	標準誤差	係数	標準誤差	係数	標準誤差	係数	標準誤差	係数	標準誤差
group2_d	**2.57***	**1.32**	**0.02**	**0.65**	**2.39*****	**0.49**	**-0.20**	**0.97**	**-0.22**	**0.52**	**0.87****	**0.39**
group3_d	**1.16**	**1.29**	**0.32**	**0.71**	**3.48*****	**0.53**	**-0.03**	**1.13**	**-0.33**	**0.69**	**1.80*****	**0.48**
group4_d	**-0.66**	**1.99**	**0.04**	**1.06**	**4.32*****	**0.66**	**-0.40**	**1.65**	**0.15**	**0.90**	**3.15*****	**0.63**
国籍ダミー	**-0.66**	**0.82**	**-0.12**	**0.46**	**0.30**	**0.32**	**1.17****	**0.59**	**-0.15**	**0.38**	**0.28**	**0.28**
基礎点(SP)	0.37***	0.14	—	—	—	—	0.51***	0.14		—	—	—
GOE(SP)	0.35*	0.21	0.38***	0.10	—	—	0.25	0.23	0.53***	0.17	0.23**	0.11
GOE^2	—	—	—	—	—	—	0.09**	0.04	-0.004	0.05	0.05***	0.02
GOE^3	—	—	—	—	—	—	-0.01**	0.01	-0.02*	0.01	-0.006**	0.00
GOE^4	—	—	—	—	—	—	—	—	0.002*	0.00	—	—
構成点(SP)	—	—	0.25**	0.11	1.50***	0.09	—	—	0.32***	0.11	1.69***	0.07
SB(FS)	0.18***	0.03	0.05**	0.02	0.04***	0.01	0.20***	0.02	0.04**	0.02	0.03**	0.01
定数項	27.42***	5.41	-15.88	2.06	5.35***	1.47	13.36***	4.43	-12.40***	1.85	3.14**	1.33
R2	0.358		0.376		0.924		0.429		0.451		0.939	
標本サイズ	329						325					

注：それぞれ，***は 1%水準，**は 5%水準，*は 10%水準で有意であることを表す．

標準誤差については，White(1980)による修正を行っている．

表 2　グループ間の係数の差の検定（F 検定）

	Men		Women	
	F値	P値	F値	P値
基礎点				
group2-group3	1.45	0.23	0.04	0.84
group2-group4	4.13**	0.04	0.02	0.88
group3-group4	1.76	0.19	0.11	0.74
GOE				
group2-group3	0.19	0.66	0.03	0.86
group2-group4	0.00	0.98	0.21	0.64
group3-group4	0.13	0.71	0.42	0.52
構成点				
group2-group3	**4.47****	**0.04**	**4.66****	**0.03**
group2-group4	**10.48*****	**0.00**	**18.07*****	**0.00**
group3-group4	**2.97***	**0.09**	**7.72*****	**0.01**

注：それぞれ，***は 1%水準，**は 5%水準，*は 10%水準で有意であることを表す．

参考文献

Findlay, L. C. and D. M. Ste-Marie (2004) “A Reputation Bias in Skating Judging,”

Journal of Sport and Exercise Psychology, Vol.26, pp.154-166.

Morgan, H. N. and K. W. Rotthoff (2014) "The Harder the Task, the Higher the Score: Findings of a Difficulty Bias," *Economic Inquiry*, Vol. 52, No.3, pp.1014-1026.

Tversky, A. and D.Kahneman (1974) "Judgment under Uncertainty : Heuristics and Biases," *Science*, Vol.185, No.4157, pp.1124-1131.

White, H. (1980) "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," Econometrica, Vol.48, No.4, pp.817-838.