

対人評価の正確さに関する研究の展望

柳澤, さおり
九州大学大学院人間環境学研究科 (博士課程)

古川, 久敬
九州大学大学院人間環境学研究科教授

<https://doi.org/10.15017/832>

出版情報 : 九州大学心理学研究. 1, pp.79-93, 2000-03-10. 九州大学大学院人間環境学研究院
バージョン :
権利関係 :

対人評価の正確さに関する研究の展望

柳澤さおり・古川久敬

A REVIEW OF THE STUDIES ABOUT ACCURACY IN PERFORMANCE APPRAISAL

SAORI YANAGIZAWA and HISATAKA FURUKAWA

Performance appraisal must be accurate because it is used the decisions about promotions, salary increase, assignments, and education. This article reviewed previous research on performance appraisal focusing on rating accuracy. First, research on ratee's information processing models and the factors that affect the processing and rating was reviewed. Next three measures of rating accuracy, that is, rating accuracy measure, rating error measure, and psychometric criteria, were discussed. Rating accuracy measure is mainly described by how close an actual rating compares with an expert's rating. Rating error measures indicate to what to extent rating errors, such as halo and leniency, are represented. Psychometric criteria is defined by reliability, such as the rating differences among individuals, and validity, such as construct validity. Rating error measures and psychometric criteria are calculated based on means and standard deviations of actual ratings. It was noted that there are problems with these measures of rating accuracy related to the application of performance appraisal in real situation, and only measure relative rating accuracy. After the problems were described, it was concluded that the rating differences among individuals were worthy of note when considering the application in a real situation. A new information processing model was presented, taking into consideration the rating differences among individuals. Finally, several research issues were presented which are relevant when studying accuracy of performance appraisal in the future.

パフォーマンスの評価は、報酬分配、人事配置、教育などを目的として行われるので、正確である必要がある。本稿では、評価の正確さに焦点を絞り、パフォーマンスの評価に関する先行研究を概観した。まず、評価者の情報処理過程のモデルやその情報処理過程に影響を及ぼす要因についての先行研究を概観した。次に、それらの先行研究で用いられている3つの評価の正確さの測度、すなわち、評価の正確さの測度、評価のエラーの測度、そして、心理測度について示した。評価の正確さの測度は、主に、評価のエキスパートが下した評価に、実際に評価を下した評価者たちの評価がどのくらい近いかによって表される。評価のエラーの測度は、ハローエラーや寛大化エラーなどの評価のエラーがどの程度みられるかによって示される。心理測度は、評価の個人差などの信頼性、および構成概念妥当性などの妥当性によって示される。そして、評価のエラーの測度や心理測度は、実際に下された評価の平均値や標準偏差をもとに算出される。評価の正確さに関する測度は、現実場面の人事考課への適用に関してそれぞれ問題点をもっており、相対的な評価の正確さを測定しているにすぎないことが述べられた。それぞれの問題点が指摘された後、評価の現実場面における適用を考慮する場合には、評価の個人

差を測度として取り上げる価値があることを指摘した。そして、この評価の個人差の生起メカニズムを理解するための従来とは異なる評価者の情報処理過程モデルを提示した。最後に、今後、評価の正確さを検討するために必要とされる研究課題について述べた。

I. はじめに

近年、日本の組織においても、成員が課題や仕事を遂行する際にみせるパフォーマンスの評価が行われ、その評価の結果が報酬分配、人事配置、あるいは教育などに生かされる機会が増えてきた。そのため、対人評価に関する基礎的な研究を行う必要性はますます高まっている。

対人評価は、一般的に、組織において、報酬分配、人員選抜、あるいは人材育成などの目的をもって行われる。そのため、下される評価が正確である必要がある。これを達成するために、組織においては、評価者が正確な評価を行うための指導や評価者訓練が行われている。

本稿では、欧米に比べて日本ではまだそれほど研究がなされていない課題遂行の際にみせる個人のパフォーマンスの評価に関して取り上げる。このパフォーマンスについては、2つのとらえ方がある。一つは、個人の仕事の目標や組織の目標の達成に効果的（非効果的）な“行動（behavior）”をパフォーマンスとしてとらえるものである（Motowidlo & Schmid, 1999; Murphy & Cleveland, 1995）。もう一つは、パフォーマンスを業績といった“行動の結果（result of behavior）”としてとらえるものである（Bernardin & Russel, 1993; Landy & Farr, 1983）。

後者の行動の結果は、個人のパフォーマンスだけでなく、他の要因（例えば、他者の協力、置かれた環境、課題の困難さ）によって影響を受ける。そのため、その個人が優れた行動をとったり、その能力を発揮したとしても必ずしも高業績をおさめるとは限らない。このことを考慮して、本稿では、被評価者のパフォーマンスを、課題や仕事を遂行する上でみせる行動とする。そして、パフォーマンスは、一般的に、自律性、責任感、管理能力などのいくつかの次元から評価されている。

このパフォーマンスの評価に関する研究は、一貫して“正確な評価を行うためにはどうしたらいいのか？”という共通の関心のもとで行われている。

本稿では、まず歴史的流れに沿って、過去に行われてきた評価の正確さに関わる研究について概観する。その後、評価の正確さを測定する際の測度に関わる議論を取り上げる。そして、評価の正確さの測定に関わる問題点を指摘し、従来の研究とは異なる研究の方向性を示す。最後に、評価の正確さを含めた対人評価研究の今後の課題について論じることにする。

II. 評価の正確さに関わる研究の展開

評価の正確さに関わる研究の歴史的な流れは2つに分けられる。それは、1920年から1980年の研究パラダイムと1980年から現在に至るまでの研究パラダイムである。

1920年から1980年までは、評価の尺度や評価の方法を変えることで、評価のエラーをなくすことを念頭において研究がなされてきた。1980年以降は、評価者の情報処理過程を明らかにし、それをもとに正確な評価を行うための要因が検討されている。

以下に対人評価研究における歴史的な流れに沿って、これまでに行われてきた研究について示す。

評価の際に用いられる評価方法、評価尺度

人が課題や仕事を行う時のパフォーマンスに対する評価は、報酬分配という観点から、今から100年以上も前から関心が向けられていた。Wells(1907)は、ヘイローエラー(halo error)と命名こそしていないが、この現象を観察している。本格的な対人評価に関する研究というのは、Thorndike(1920)によって初めて公に発表された評価のヘイローエラーに関する論文から始まるとされている(Murphy & Cleveland, 1991)。

1920年から1980年まで、対人評価に関する研究は、もっぱら寛大化傾向 (*leniency*) やヘイローエラーなどの評価のエラーに焦点が絞られ、評価方法や評価尺度によってこのエラーを減らすことが試みられた。

対人評価の際に用いられる評価方法、あるいは評価尺度として提示されている代表的なものを以下に示す。

まず、仕事を行う際に被評価者がみせるすべてのパフォーマンスではなく、被評価者が行った特定の非常に良かった出来事、あるいは非常に悪かった出来事、いわゆるクリティカル・インシデント (*critical incident*) を評価者が決めて記述する方法が提案された (Flanagan, 1949, 1954; Flanagan & Burns, 1955)。これに類似したものとして、2つ以上の望ましい、あるいは望ましくないパフォーマンスの記述が評価者に提示され、評価者が特定の被評価者に当てはまる記述を選択する強制選択法 (Sisson, 1948) がある。これは、評価の寛大化を防ぐための方法として提示されたがそのような効果はみられなかった。

Smith & Kendall (1963) は、Behaviorally Anchored Rating Scale (BARS) を示した。この評価尺度では、尺度のアンカーが優れているとか劣っているというような形容詞で示されるのではなく、具体的なパフォーマンスによって示されている。そして、評価者は当該の被評価者に当てはまると思う行動のアンカーを選択する。また、これに類似したものとして、評価者が被評価者のパフォーマンスを評価するのではなく、観察した行動を記す行動観察尺度 (Behavioral Observation Scales; BOS) を Kane & Bernardin (1982)、そして Latham & Wexley (1977) らが推奨している。

その他の尺度としては、Blanz & Ghiselli (1972) が提案した被評価者のパフォーマンスのレベル (平均以下、平均的、平均以上など) を決める混合基準評価尺度 (Mixed Standard Rating Scales) などがある。

1920年から1980年までは、以上に示した評価の方法や評価の尺度が開発され、それらと評価のエラーとの関係に関する研究が盛んになさ

れてきた。しかし、Landy & Farr (1980) によるそれらの研究の広範なレビューの結果は、ある方法や尺度が他のそれより優れているということはないことを示している。そして、彼らは評価の正確さを高めるためには、評価方法や評価尺度よりも、評価者の認知過程、すなわち評価者が行う情報処理過程に注目すべきであると主張している。このときから現在に至るまで、対人評価においては評価者の情報処理を考慮に入れて研究が行われている。

評価者の情報処理過程

対人評価にみられる情報処理過程についての主なモデルとして Feldman (1981)、Ilgen & Feldman (1983)、DeNisi, Cafferty, & Meglino (1984)、そして Murphy & Cleveland (1991) らのものがある。

これらのモデルは、評価者は情報の“獲得→符号化→想起”をもとに“評価”を行うことを仮定している点で共通している。また、人間の情報処理能力には限界があるので、観察した全ての情報を処理できないこともこれらのモデルにおいて暗黙の前提となっている。

一方、これらのモデルにはそれぞれ特徴がある。

Feldman, Ilgen & Feldman モデル 1981年に提示された Feldman のモデルでは、情報処理過程を自動化処理と制御的処理とに分類している。評価者は、カテゴリーやプロトタイプに一致するような情報については、意識することなく自動的に処理する。逆に、それらに一致しないような情報については、意識的に注意を払って制御的な情報処理を行う。

Feldman (1981) のモデルに続いて、1983年に Ilgen & Feldman が示したモデルでは、職場の文脈が評価に影響を及ぼす重要な要因であるという考えが取り入れられている。

DeNisi, Cafferty, & Meglino モデル DeNisi, Cafferty, & Meglino (1984) のモデルでは、評価者が行う情報処理過程のうち、情報の獲得の段階に特に注目しており、さらに、情報を想起する際の被評価者のパフォーマンスの原因帰属 (内的帰属、あるいは外的帰属) 過程を重視している。

Murphy & Clevelandモデル Murphy & Cleveland (1991)のモデルでは、評価の際に示される評価の目的を特に重視しており、評価者は評価の目的に応じて情報処理、そして評価を下すとしている。また、彼らは、情報の想起の後に判断(judgement)という段階を想定している。この判断の段階では、評価者は、被評価者のパフォーマンスを観察し、その優劣を率直に評価している。この判断の後に、評価者は被評価者のパフォーマンスの優劣を評価尺度に記入する形式で評価を下す。Murphy & Cleveland (1991)によると、このとき下す評価は、判断と必ずしも一致しない。というのは、実際の組織においては、評価者の思惑などが絡むために、評価者の判断をそのまま評価尺度に記入することはないためである。

評価者の情報処理過程に影響をおよぼす要因

これまで、評価者の情報処理過程に影響をおよぼす様々な要因について取り上げられてきた。これらの要因は、評価者個人の内部の要因と外部の要因とに分けて考えることができる。

内部の要因 評価者内部の要因として、まず評価者の“認知的能力”を挙げることができる。これに関しては、認知的複雑性(Schneier, 1977)、選択的注意能力(Cardy & Kehoe, 1984)そして知性(Smith & Reilly, 1987)などについての研究がなされてきたが、評価の正確さにおよぼす影響は相対的に小さいようである。

評価者がもつ体制化された“知識構造”に関しても多くの要因が検討されてきた。例えば、性別(Maurer & Taylor, 1994)、職業(Kinicki, Hom, Trost & Wade, 1995)、あるいは学歴(佐々木, 1997)などの知識、課題関連の知識(Freeberg, 1969)そして被評価者の以前のパフォーマンスや評価(Kozlowski & Ford, 1991; Sumer & Knight, 1996)についての知識などを挙げることができる。

また、暗黙の人格理論(Krzystofiak, Cardy, and Newman, 1988)や特性理論(Hong, Chiu, Dweck, & Sacks, 1997)など評価者が持つ“人格や特性”についての暗黙の考え方、評価に対する“態度”(柳澤・古川, 1998a, 1998b)、あるいは評価者の“モチベーション”(Mohrman & Lawler,

1983; Salvemini & Reilly, 1993)なども評価者が下す評価の正確さに影響がある。

外部の要因 評価者にとって外部にある要因としては、評価者の情報処理に直接的に影響するものと間接的に影響するものがある。

直接的に影響するものとしては、評価者が評価を行うにあたって与えられる“評価の目標や目的”(Bernardin & Cooke, 1992; 古川, 1994; Murphy, Philbin & Adams, 1989; Williams, DeNisi, Blencoe, & Cafferty, 1985)、評価者訓練のときに与えられる“パフォーマンス次元に関する基準”(Bernardin & Buckley, 1981; McIntyre, Smith, & Hassett; 1984; Pulakos, 1984; Woehr, 1994)などを上げることができる。

一方、間接的に影響するものとしては、グループの内の相互依存性(Liden & Mitchell, 1983)、メンバーの特性が目立つ程度(Brewer, 1979; Linville & Jones, 1980; Quattrone & Jones, 1980)、あるいは評価の尺度(Murphy & Constants, 1987; 1988)などが取り上げられてきた。

外部要因に関して、前者の直接的に評価に影響をおよぼす評価の目的やパフォーマンス次元に関する基準などの要因は、後者の間接的に影響を及ぼす要因に比べて、評価の正確さにおよぼす影響が大きいことをMurphy & Cleveland (1995)は示唆している。

III. 評価の正確さの測定

すでに示した評価者の情報処理過程に影響をおよぼす様々な要因について検討した研究では、評価の正確さを測定するために何らかの測度を用いている。この評価の正確さの測度は、その測定の方法によって大きく二つに分けることができる。一つは、評価者が下した評価の正確さを直接的に測定する方法である。これには、評価の正確さの測度(rating accuracy measures)が該当する。この測度では、評価を下すことを職業とするエキスパートが下す評価を基準として、実際に評価を下した評価者達の評価がその基準にどのくらい近いかによって表されることが多い。

もう一つの方法は、評価者が下した評価の不

正確さを測定することで、評価の正確さを間接的に推定する方法である。これには、評価のエラーの測度 (rater error measures) や心理測度 (psychometric criteria) が該当する。これらの測度では、実際に評価を下した評価者達の平均値を基準として、各評価者たちが下した評価の不正確さが表される。

以下に、評価の正確さの測度、評価のエラーの測度、および心理測度について順次述べる。

評価の正確さの測度

評価の正確さの測度では、最良の状況下でバイアスのない評価者が下す評価を“真の得点 (true score)”とし、これを基準にして、普通の評価者が下した評価との差異を測定することがほとんどである。一般的には、評価を下すことを職業とするエキスパートによる評価がこの基準となる。評価の正確さ測度は、行動に基づく測度 (behavior-based measures) と判定を表す測度 (judgmental measures) とに分けられている (Murphy & Cleveland, 1995)。

行動に基づく測度 行動に基づく測度では、ある被評価者に特定の重要な出来事 (critical incident) が生じたか否かを評価者に回答してもらう。

したがって、この測度では行動 (パフォーマンス) の優劣のレベルを測定することができない。実際の組織では、成員の行動 (パフォーマンス) における優劣のレベルに応じて、報酬分配、人事配置、あるいは教育などを行うことも多い。そのため、限られた範囲で行動に基づく測度が利用されている。

判定を表す測度 判定を表す測度として、最もよく用いられるものが、Cronbach (1955) によって公式として提示された (a) elevation, (b) differential elevation, (c) stereotype accuracy, および (d) differential accuracy である。これらの測度指標を理解する際には、評価者が、複数の被評価者について、それぞれ複数のパフォーマンス次元 (自律性、責任感、管理能力など) を評価するという状況を念頭におく必要がある。

上記の測度指標は、それぞれ評価の異なる側面に注目して正確さの程度を求めている。ま

ず、elevation は、評価者が行った全ての評価の正確さに関わるものである。すなわち、複数の被評価者の全パフォーマンス次元について評価者が行った評価が、全体としてどのくらい正確であるかを示す。この指標によって、各評価者が正確な評価を行ったか否かに関して、総合的にとらえることが可能になる。

次の differential elevation は、複数の被評価者の区別に関わるものである。このときに、パフォーマンスの各次元における評価の正確さを問題としない。この指標によって、各評価者が複数の被評価者の優劣の評価によって、それら被評価者の順位づけを正確にしているかどうかをとらえることができる。

stereotype accuracy は、被評価者が示すパフォーマンスの各次元についての評価の正確さに関わっている。stereotype accuracy では、被評価者間の優劣を正確に区別しているか否かを問題としない。この指標によって、自律性、責任感、管理能力などのパフォーマンス次元のうち、評価者はどの次元をどのくらい正確に評価しているのかをとらえることができる。

そして、differential accuracy は、複数の被評価者が示すパフォーマンスの各次元についての評価の正確さに関するものである。この指標は、differential elevation と stereotype accuracy のいずれの評価の正確さも検討する測度指標で、最も詳細に評価の正確さを検討できる。この differential accuracy については、Borman (1977) や McIntyre, Smith, & Hassett (1984) によって、Cronbach (1955) の公式を変容させた測度が指標提案されている。

評価におけるエラーの測度

評価にみられる主要なエラーとして、(a) 寛大化傾向 (leniency), (b) 中心化傾向 (central tendency), および (c) ヘイローエラー (halo error) があり、これらについては 70 年以上も前から (Kingsbury, 1922) 現在に至るまで検討されて続けてきている。

寛大化傾向 まず、評価における寛大化傾向とは、ある被評価者 (あるいは複数の被評価者) のパフォーマンスを実際よりも寛大に (高く)

評価する傾向である。Saal, Downey, & Lahey (1980)によると、この寛大化傾向は、評価の値が尺度の中央値から離れる程度、評価の散らばりに関する歪度をもとにして測定できる。

中心化傾向 次に、評価における中心化傾向とは、評価段階の中心に評定をつけてしまう傾向であり、ある評価次元における複数の被評価者に対する評価についての標準偏差、評価の散らばりに関する尖度をもとにして測定される(Saal, Downey, & Lahey, 1980)。

ヘイローエラー 最後に、評価にみられるヘイローエラーとは、評価者の全体的な印象に引きずられて、被評価者の特定のパフォーマンス次元の評価を歪める傾向、および評価にあたって評価次元や項目を識別できない傾向である。そして、複数のパフォーマンス次元間の評価の値の相関、複数のパフォーマンス次元の評価の値のなかで第1因子によって説明される評価の分散の割合、あるいは複数のパフォーマンス次元間の標準偏差によって測定される(Saal, Downey, & Lahey, 1980)。

評価における心理測度

評価の正確さを測定する測度として最後に示す心理測度では、評価の信頼性と妥当性を検討することができる。

評価の信頼性 信頼性は、(a)評価項目間の内的一貫性、(b)再検査(test-retest)の一致、そして(c)評価者間の評価の一致に関する程度によってそれぞれ測定する。評価項目間の内的一貫性の程度は、一般的な対人評価の研究ではほとんど取り上げられない。というのは、ある被評価者が行うパフォーマンスが同次元、あるいは同一の構成概念から成り立っていないためである。

ある被評価者について評価を行い、しばらく時間をおいて再び同一の被評価者の評価を行って、それらの評価の一致の程度を検討する再検査法では、評価の測定の長期にわたる安定性が明らかになる。この測定を行うときに、人のパフォーマンスの良否は変化しないということが前提となっている。現実には、モチベーション、あるいは教育などによって、人のパフォーマンス

が変化するために、この測度もあまり用いられることはない。

評価者間の評価の一致は、複数の評価者間で評価の個人差がみられない状態である。その場合には、それらの評価者が下した評価に関して、その正確さについての信頼性が高まる。この測度は、評価の正確さの測度として、それほど多くはないが、用いられてはいる。

評価の妥当性 評価の妥当性は、パフォーマンス次元に関する(a)基準連関妥当性(criterion-related validity)、や(b)構成概念妥当性(construct validity)によって測定される。

評価における基準連関妥当性は、パフォーマンスの評価を、業績など客観的な仕事の結果に関する基準と比較することによって得られる(Heneman, 1986; Roach & Wherry, 1970)。しかし、パフォーマンスが業績の高さに与える影響は限定的であるため(Landy & Farr, 1983)、この妥当性の測度はあまり取り上げられることはない。

評価における構成概念妥当性は、あるパフォーマンスの評価が、そのパフォーマンスを正確にとらえていることを示すことで得られる。この妥当性を示すためには、評価の対象となるそれぞれのパフォーマンスがどのような要素から構成されているのかが詳細に検討される必要がある(Murphy & Davidshofer, 1988)。しかし、構成要素やパフォーマンスの構造を表す適切な指標について検討した研究はないために、評価における構成概念妥当性を示すことは今のところ困難であると思われる。

上記のように、心理測度では、評価者間の評価の一致の程度以外の測度は、評価の正確さを測定する場合にあまり適切でない、あるいは困難と思われる。

評価のエラーの測度と正確さの測度との関係

近年行われている評価に関する研究では、評価のエラーの測度や評価の正確さの測度を用いられることが多いようである。実際、Balzer & Sulsky (1990)は、1976年に使われた測度数と1986年に使われた測度数とを比較し、評価の正確さの測度や評価のエラーの測度を用いること

は、心理測度に比べると、増加傾向にあることを示している。この動向に合わせて、評価の正確さの測度と評価のエラーの測度の関係について検討が行われている。

評価における正確さの測度では、エキスパートの評価を基準にして、評価者が下した評価の正確さを直接的に測定している。一方、評価のエラーの測度では、評価者が下した評価の不正確さを測定することで、評価の正確さを間接的に測定している。両者の測度では、直接的、あるいは間接的に評価の正確さを測定しているので、それらの相関は当然高いと思われる。しかし、実は、それぞれの測度によって測定された評価の正確さの関係は低いことが示されている (Becker & Cardy, 1986; Bernardin & Pence, 1980; Borman, 1977; Murphy & Balzer, 1989)。そして、Murphy & Balzer (1989)によって行われたメタ分析の結果は、それらの測度間の相関の低さを明らかにしている。このことは、2つの測度間でそれぞれ異なる評価の正確さを測定していることを示す。

このようなエラーの測度と評価の正確さの測度の関係は、評価にみられるヘイロー効果について特に顕著である (Becker & Cardy, 1986; Cooper, 1981; Murphy & Balzer, 1986)。すなわち、ヘイローエラーの測度でエラーを示している程度が高いほど、正確さの測度で正確であることを示している程度が高いという関係がみられる。これは“ヘイローと正確さのパラドックス (halo-accuracy paradox)”として知られている現象である。

以上のように、ヘイローをはじめとして、他のエラーに関しても、評価におけるエラーの測度によって測定された評価の正確さと評価の正確さの測度によって測定された評価の正確さは、必ずしも一致しないようである。これは、評価におけるエラーの測度では、実際に評価を下した評価者達の平均値を基準として測定がなされるのに対して、評価の正確さの測度では、評価を下すことを職業とするエキスパートが下す評価を基準として測定がなされるため、それらの基準の違いが測度間の相関の低さにつながっ

ている可能性が考えられる。

V. 評価の正確さの測定に関わる問題点

すでに述べたように、評価の正確さを測定には、評価の正確さの測度 (rating accuracy measures)、評価のエラーの測度 (rater error measures)、そして心理測度 (psychometric criteria) がこれまでの評価研究で主に用いられてきた。これらの測度についての問題点をそれぞれ示すことにする。

評価の正確さの測度

近年、Cronbach (1955) によって考案された elevation, differential elevation, stereotype accuracy, そして differential accuracy といった評価の正確さの測度を用いて研究が行われることが多くなっている (Murphy & Cleveland, 1995)。

この測度は、“最良の状況下で、バイアスのない評価のエキスパートが下す評価は正確である”という前提のもとで求められる。しかし、エキスパートによる評価が果たして本当に正確であるのかどうかを検討できない。そのために、この前提のもとで測度を求めること自体、問題があるであろう。

例え、エキスパートによる評価が正確であったとしても、次のような問題点がある。それは、しばしば指摘されることだが、Cronbach (1955) による評価の正確さの測度は実験室場面では有効であるが、現実場面で行われる人事考課などに一般化できないことである。というのは、評価を下すエキスパートは、バイアスが生じないような状況のもとで評価を行わなければならないが、実際の組織においては、利害、人間関係、あるいは被評価者に対する好悪感情などが絡んで、そのような状況を作り出すことは不可能であるからである。

これらの理由から、エキスパートによる評価を正確さの絶対的な基準とすること自体にも問題があるし、現実の組織における適用も難しいと考えられる。

この測度に対して、エラーの測度と心理測度は、実際に評価を下した評価者達の平均値や標準偏差を基準にして測定される。それらの値

は、絶対的な正確さの基準ではなく、仮の基準としてとらえられている。つまり、基準の正確さの限界を考慮に入れた測度として考えることができるであろう。また、エラーの測度と心理測度は、現実の組織においても用いることができる。しかし、それらの測度についてもまた問題点がある。次に、この問題点について述べることにする。

評価のエラーの測度

評価のエラーの測度は、公式によって求められた値をもとに、ある評価に対してエラーがみられる（みられない）と判断するが、その判断が正しいという保証はない。例えば、中心化のエラーを測定する公式を用いて、ある評価者が下した評価に中心化の傾向があると判断されたとする。しかし、その評価は本当に中心化傾向があったかどうかは確かめられない。実はその評価こそが正確であった可能性もある。このような問題点について、近年、ヘイローエラーについて取り上げた研究が増えている（Cooper, 1981; Kozlowski & Kirsch, 1987; Murphy, Jako, & Anhalt, 1993）。

ヘイローエラーは、評価者の全体的な印象に引きずられて、被評価者の特定のパフォーマンス次元の評価を歪める傾向、および評価にあたってパフォーマンスの各次元を識別できない傾向である。このヘイローエラーを検討する際に、Cooper(1981)が命名した“擬似的なヘイロー(illusory halo)”と“真のヘイロー(true halo)”というそれぞれの状態があることを考慮に入れる必要がある。

異なるパフォーマンスの次元で、かつそれらの間の類似性が低い（相関が低い）とき、それは“擬似的なヘイロー”として定義される。この状態で、評価者がそれらのパフォーマンスの次元に類似した評価を下した場合、その評価にヘイローエラーが生じているととらえることができる。

一方、異なるパフォーマンスの次元であるが、それらの間の類似性が高い（相関が高い）とき、それは“真のヘイロー”と定義される。この状態で、評価者がそれらのパフォーマンスの

次元に類似した評価を下したとしても、その評価にヘイローエラーが生じているとはいえないことになる。

心理測度

心理測度については、最も利用可能性の高いと思われる評価者間の評価の一致、すなわち評価の個人差について言及することにする。この測度も、エラーの測度と同様に、研究に用いる際に注意を払う必要がある。

すでに述べたように、評価の個人差は評価の信頼性と関わっている。複数の評価者が、同一の被評価者について下した評価に個人差がなければ、それらの評価の信頼性は確かに高くなる。しかし、信頼性が高まったからといって、正確な評価であるという妥当性は必ずしも高くない。複数の評価者が寛大化など同じバイアス傾向をもっている場合、それらの評価者が下す評価は個人差が見られないとしても、その評価は正確であるとは言えない。複数の評価者が同じバイアス傾向を持たないために、frame-of-reference 訓練（Bernardin & Buckley, 1981; McIntyre, Smith, & Hassett, 1984; Pulakos, 1984; Woehr, 1994）など、外的にパフォーマンスの基準を教え込む評価者訓練によってバイアス傾向を修正する必要があるであろう。

評価の個人差は被評価者が感じる公正感や満足感にも関係していることは注目に値する。仮に、複数の評価者が同一の被評価者に対して評価を下す場面を想定してみよう。被評価者が自分に下された評価に納得し、公正感を感じるには、手続き的公正、および結果の公正だけでなく、評価者間の評価の一致が重要な鍵となる。それぞれの評価者によって、下される評価がバラバラであるならば、それらの評価には信頼性がなく、どれが正確な評価であるのかが分からず、不満が残るであろう。一方、皆が類似した評価を下していた場合には、被評価者はそれらの評価に納得しやすいことが予想される。もし、被評価者が下された評価を受け入れなければ、組織における評価システムがうまく機能しなくなってしまう。そのため、被評価者の公正感や満足感は特に重要であろう。

実は、これまで、評価の個人差を積極的に取り上げた研究はほとんどみあたらない。しかし、評価の個人差は、評価の信頼性、そして被評価者の公正感や満足感にも関わっているため、この評価の個人差にもっと注目する必要があると思われる。

また、エキスパートによる評価ではなく、実際に評価を下した評価者達の評価の個人差をもとに“真の得点(true score)”を求め、それを基準にした評価の正確さの測度をあらわす方法も考案されている(Murphy & Cleveland, 1995)。

そして、最近では、現実の企業組織においても、面接場面だけでなく、360度評価といった、複数の評価者が1人の被評価者を評価する機会が増えている。そのため、評価の個人差を求めることが可能になりつつある。

このような状況を考慮すると、評価の個人差は、評価の正確さを示す測度として取り上げるに値するであろう。

次に、この評価の個人差について研究を行ううえで必要とされることについて述べることにする。

VI. 評価の個人差の研究で必要とされること

これまで提示されてきた、Feldman(1981), Ilgen & Feldman(1983), DeNisi, Cafferty, & Meglino(1984), そしてMurphy & Cleveland(1991)など評価者の情報処理についての代表的なモデルは、評価の個人差を念頭においたものではないので、このモデルをもとにして評価の個人差が生じるメカニズムを検討することは難しい。

本稿では、評価の個人差の生起メカニズムを理解するために新たな情報処理のモデルを提示する。そして、そのモデルをもとに評価の個人差の研究に必要な検討事項について述べる。

評価の個人差と情報処理過程

評価の個人差は、情報処理の内容、すなわち、被評価者のパフォーマンスのどこに注目するか、またそれをどのように解釈、あるいは意味づけするかといったことが、それぞれの評価者によって異なることから生じる。このような情

報処理の違いは、人間の情報処理能力の違いと関係している。

周知のごとく、人間の情報処理能力には限界があると考えることが妥当であろう。したがって、評価者は被評価者についての情報処理に直面して、被評価者の全ての情報に着目し、またその着目した情報を全て処理し、活用することは難しい。しかしながら、一定のまとまった評価を下さなければならない。これを可能にするため、評価者は何かしらの枠組みに準拠して、被評価者についての限定した情報に着目し、その情報を解釈したり意味づけたりして記憶し、必要に応じて想起し、それをもとに最終的に評価を下すと考えられる。この情報処理の際に評価者が拠り所とする枠組みを“準拠枠”としてとらえることにする。

この準拠枠が、評価者が行う情報処理の内容、そして評価の個人差を生み出す大きな要因として考えられる。すなわち、複数の評価者が異なる準拠枠を拠り所とすることによって、被評価者のパフォーマンスのどこに注目するか、またそれをどのように解釈、あるいは意味づけするかといったことが、それぞれの評価者によって異なり、最終的に評価の個人差が生じると推測できる。

内的準拠枠と外的準拠枠

準拠枠としては、個人が自ら獲得し、保持している“内的準拠枠”，反対に、外的に与えられたり、操作されるような“外的準拠枠”を想定できる。

前者の内的準拠枠としては、性別(Maurer & Taylor, 1994)や職業(Kinicki, Hom, Trost & Wade, 1995), 学歴(佐々木, 1997)など人の特性についての固定的な見方、あるいは課題関連の知識(Freeberg, 1969)そして被評価者の以前のパフォーマンスや評価の知識(Kozlowski & Ford, 1991; Sumer & Knight, 1996)などが当てはまる。

一方、後者の外的準拠枠としては、先述の評価の目的や基準などが考えられる。この内的準拠枠と外的準拠枠を考慮した評価者の情報処理過程を図1に示す。図1に示すように、評価者は自ら持つ内的準拠枠、あるいは外的に与えら

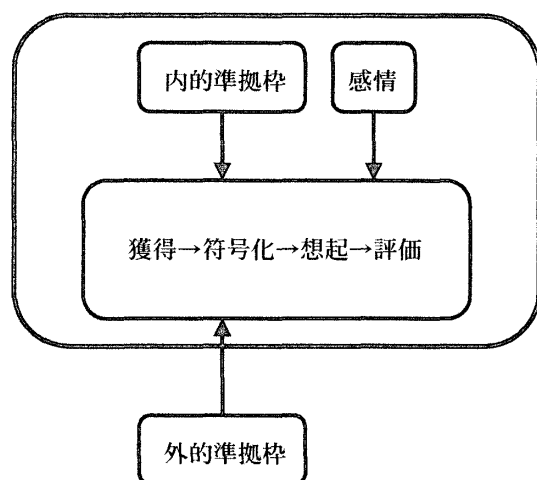


図1 評価者が行う情報処理過程のモデル図

れた内的準拠枠を拠り所として、被評価者の情報を獲得し、これを符号化し、それを想起して評価を下すという一連の情報処理を行う。また、この過程には、評価者をもつ被評価者への感情も影響を与えている。

これまでの評価に関する研究では、この内的準拠枠と外的準拠枠について明瞭には意識されることはなかった。しかし、実際の組織で行われている評価場面に研究知見を適用するためには、どのような評価においても一貫した影響をおよぼしているはずの内的準拠枠、そして組織に一般的に与えられる評価の基準、あるいは評価の目標などの外的準拠枠を同時に考慮する必要があるであろう。

最近の対人評価に関する研究を振り返ると、外的準拠枠、特に評価者訓練で利用されるパフォーマンス次元に関する基準を評価者に教え込み、そのことによって、それを拠り所とさせて正確な評価への効果を検討することについて研究されている (Bernardin & Buckley, 1981; McIntyre, Smith, & Hassett; 1984; Pulakos, 1984; Woehr, 1994)。

一方、内的準拠枠については、これまであまり注目されていなかった。従って、内的準拠枠にはどのような種類があるのかについてはほとんど明らかになっていない。しかし、柳澤・古川(1999)の探索的な研究によって、“評価者の過去経験”、“仕事上の規範についての考え”、あ

るいは“課題状況の困難さについての考え”などの内的準拠枠が存在し、情報処理、および評価の際の拠り所となっていることが示唆されている。

内的準拠枠と外的準拠枠との関係に関しては柳澤・古川(未発表資料)による研究結果から、次のことが考えられる。評価者が情報処理を行うときに、基本的には内的準拠枠を活性化させて、それをもとにして被評価者に関する情報に着目し、意味づけや解釈を行い評価を下す。しかし、外的準拠枠が与えられた場合、評価者は与えられた外的準拠枠を拠り所として情報処理を行う。

内的準拠枠は評価者が育ってきた環境や過去の評価経験に応じてその内容が異なると思われる。そのため、この内的準拠枠のみを拠り所とした場合、明瞭な外的準拠枠を与えられたときと比べて、同一の他者のパフォーマンスについて、異なる評価を下す現象、すなわち評価の個人差がより明確にみられる。

以上の研究結果を考慮すると、評価の個人差を検討するために、内的準拠枠の種類、およびそれが評価におよぼす影響を分析し、さらに外的準拠枠との関係をより詳細に調べる必要があることがわかる。

この内的準拠枠と外的準拠枠に加えて、忘れてならないのが、評価者の“感情”の影響である。従来の Feldman(1981), Ilgen & Feldman(1983), DeNisi, Cafferty, & Meglino(1984), そして Murphy & Cleveland(1991)のモデルでは、評価者の感情の要因を含めていない、いわゆる cool modelであった。しかし、評価者が好意感情を持つほど寛大な評価を行うという研究結果 (Imada & Hakel, 1977; Robbins & DeNisi, 1994; Tsui & Barry, 1986; 柳澤・古川, 1998a)を見過ごすことはできない。また、現実の組織では、上司などの評価者と部下などの被評価者と関係は、感情を伴ったものである。これらのことを考慮すると、DeNisi(1996)や Murphy & Cleveland(1995)やなどが指摘するように、評価者の感情を含めたモデル(hot model)を考案する必要があるであろう。

VII. 総 括

人が課題や仕事を遂行する際にみせるパフォーマンスについての評価は、通常、組織の中で行われる。そして、その評価は、被評価者や組織に大きな影響をもたらす。例えば、報酬分配に関する評価は、被評価者の仕事に対するモチベーションに影響する。人員選抜のための評価は優秀な人材を生かすことに関わる。そして、人材育成のための評価は人の弱みを改善するための教育や強みを伸ばすことに用いられる。そのようにして、効果的に評価を活用することによって、組織のパフォーマンスを上げることが可能になる。究極的には、組織における評価は、このパフォーマンスを上げるために行われるのである。評価が被評価者や組織にこのような影響をもたらすことを考慮すると、評価者が正確な評価を行うことは非常に重要なことであることがわかる。

本稿では、まず評価者の情報処理過程のモデルやその情報処理過程に影響をおぼす要因についての先行研究を概観した。そして、それらの研究で用いられている評価の正確さの測度について説明を行い、それらの測度の問題点を指摘した。そして、これまであまり取り上げられることのなかった評価の個人差に注目することの重要性を述べ、評価の個人差の生起メカニズムを説明するための情報処理モデルを提示した。

従来の評価に関する研究と同様に、本稿でも“正確な評価を行うためにはどうしたらよいか?”という問いに対する完全な答えを出すには至っていない。今後、評価の正確さの測度、および評価者の情報処理過程のモデルやその情報処理過程に影響をおぼす要因についてさらなる検討が必要であろう。

最後に、評価の個人差と同様に、これまで見過ごされていたが、対人評価について研究を行う際に、重要であると思われる事柄を示す。

まず、研究の対象者の“評価経験”についての考慮である。過去の研究では、評価者の評価経験の違いによって情報処理、そして評価にどのような影響がみられるのかを検討した研究は

見当たらない。しかしながら、柳澤・古川(1998a, 1998b, 1999)の研究結果が示すように、評価経験の違いが情報処理、および評価に大きな影響をもつことを考えると、研究の対象者の評価経験は見過ごすことのできない要因と思われる。

次に、すでに示した評価者の“感情”の要因と同様に、実験室よりも現実の組織においてより大きな影響をおよぼす要因に関する研究も重要である。これに関して、すでに、部下の取り入り行動が評価におよぼす影響(Deluga & Perry, 1994; Gerstein, Ginter, & Graziano, 1985; Martin, 1987; Wayne & Kacmar, 1991)が検討されている。また、現実の組織では評価者と被評価者が良好な人間関係を続けなければならないこと(Murphy & Cleveland, 1995)や評価者の報酬がその評価者の所属する職場集団のパフォーマンスと関係していること(Ilgen, Mitchell, & Frederickson, 1981)などの理由から、評価の寛大化が生じると思われる。これら、そしてこれら以外の現実の組織においてより大きな影響をおよぼす要因に関してさらに調べる必要があるであろう。

文 献

- Balzer, W.K. & Sulsky, L.M. (1990) Performance appraisal effectiveness and productivity. In K. Murphy & F. Saal (Eds.), *Psychology in organizations: Integrating science and practice*. Hillsdale, NJ: Lawrence Erlbaum.
- Becker, B. E., Cardy, R. L. (1986) Influence of halo error on appraisal effectiveness: A conceptual and empirical reconsideration, *Journal of Applied Psychology*, 71,662-671.
- Bernardin, H. J., & Buckley, M.R. (1981) Strategies in rater training, *Academy of Management Review*, 6,205-212.
- Bernardin, H. J., & Cooke, D. K. (1992) Effects of purpose of appraisal on discriminability and accuracy of ratings, *Psychological Reports*, 70, 1211-1215.
- Bernardin, H.J., & Pence, E. C. (1980) Effects of

- rater error training: Creating new response sets and decreasing accuracy, *Journal of Applied Psychology*, **65**,60-66.
- Bernardin, H. J., & Russel, E. (1993) *Human resource management*. New York: McGraw-Hill.
- Blanz, F., & Ghiselli, E. E. (1972) The mixed standard scale: A new rating system, *Personnel Psychology*, **25**,185-199.
- Borman, W. C. (1977) Consistency of rating accuracy and rating errors in the judgement of human performance, *Organizational Behavior and Human Performance*, **20**,238-252.
- Brewer, M. B. (1979) In-group bias in the minimal intergroup situations: A cognitive-motivational analysis, *Psychological Bulletin*, **86**,307-324.
- Cardy, R. L., & Kehoe, J. F. (1984) Rater selective attention ability and appraisal effectiveness: The effect of a cognitive style on the accuracy of differentiation among ratees, *Journal of Applied Psychology*, **69**,589-594.
- Cooper, W. (1981) Ubiquitous halo, *Psychological Bulletin*, **90**,218-244.
- Cronbach, L. J. (1955) Processes affecting scores on "understanding of others" and "assumed similarity," *Psychological Bulletin*, **52**,177-193.
- Deluga, R. J. & Perry, J. T. (1994) The role of subordinate performance and ingratiation in leader-member exchanges, *Group and Organization Management*, **19**,67-86.
- DeNisi, A.S. (1996) Cognitive approach to performance appraisal. London and New York: Routledge
- DeNisi, A.S., Cafferty, T.P., & Meglino, B.M. (1984) A cognitive in view of the performance appraisal process: A model and research propositions, *Organizational Behavior and Human Decision Processes*, **33**,360-396.
- Feldman, J.M. (1981) Beyond attribution theory : Cognitive processes in performance appraisal, *Journal of Applied Psychology*, **66**,127-148.
- Flanagan, J. C. (1949) Critical requirements: A new approach to evaluation. *Personnel Psychology*, **2**,419-425.
- Flanagan, J. C. (1954) The critical incident technique, *Psychological Bulletin*, **51**,327-358.
- Flanagan, J. C., & Burns, R. K. (1955) The employee performance record: A new appraisal and development tool. *Harvard Business Review*, September / October, 95-102.
- Freeberg, N. F. (1969) Relevance of rater-ratee acquaintance in the validity and reliability of ratings, *Journal of Applied Psychology*, **53** 518-524.
- 古川久敬 (1994) 集団による情報処理過程 — 問題解決に及ぼすフレイミングの効果 産業・組織心理学会第10回大会発表論文集 Pp71-73.
- Gerstein, L. H., Ginter, E. J., & Graziano, W. G. (1985) Self-monitoring, impression management, and interpersonal evaluations. *Journal of Social Psychology*, **125**,379-389.
- Heneman, R. L. (1986) The relationship between supervisory ratings and results-oriented measures of performance: A meta-analysis, *Personnel Psychology*, **39**,811-826.
- Hong, Ying-yi., Chiu, Chi-yue., Dweck, C. S., & Sacks, Russell. (1997) Implicit theories and evaluative processes in person cognition, *Journal of Experimental Social Psychology*, **33**, 296-323.
- Ilgen, D. R., Mitchell, T. R., & Frederickson, J. W. (1981) Poor performers: Supervisors' and subordinates' responses, *Organizational Behavior and Human Performance*, **27**,386-410.
- Ilgen, D.R., & Feldman, J.M. (1983) Performance appraisal: A process approach. In B.M. Staw (Ed.), *Research in organization behavior* (Vol.5). Greenwich, C.T.: JAI Press.
- Imada, A. S. & Hakel, M. D. (1977) The effect of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews, *Journal of Applied Psychology*, **62**,295-300.

- Kane, J. S., & Bernardin, H. J. (1982) Behavioral observation scales and the evaluation of performance appraisal effectiveness, *Personnel Psychology*, **35**,635-642.
- Kingsbury, F. A. (1922) Psychological tests for executives, *Personnel*, **9**,121-133.
- Kinicki, A.J., Hom, P. W., Trost, M.R., & Wade, K.J. (1995) Effects on category prototypes on performance-rating accuracy, *Journal of Applied Psychology*, **80**,354-370.
- Kozlowski, S. W., & Ford, J. K. (1991) Rater information acquisition processes: Tracing the effects of prior knowledge, performance level, search constraint, and memory demand, *Organizational Behavior and Human Decision Processes*, **49**,282-301.
- Kozlowski, S. W., & Kirsch, M. P. (1987) The systematic distortion hypothesis, halo, and accuracy: An individual-level analysis, *Journal of Applied Psychology*, **72**,252-261.
- Krzystofiak, F., Cardy, R., & Newman, J. (1988) Implicit personality and performance appraisal: The influence of trait inferences on evaluations of behavior, *Journal of Applied Psychology*, **73**,515-521.
- Landy, F. J., & Farr, J. (1980) Performance rating, *Psychological Bulletin*, **87**,72-102.
- Landy, F. J., & Farr, J. (1983) The measurement of work performance: *Methods, theory, and applications*. New York: Academic Press.
- Latham, G. P., & Wexley, K. N. (1977) Behavioral observation, *Personnel Psychology*, **30**,255-268.
- Liden, R.C., & Mitchell, T. R. (1983) The effects of group interdependence on supervisor performance evaluations. *Personnel Psychology*, **36**,289-300.
- Linville, P. W., & Jones, E. J. (1980) Polarized appraisals of out-group members, *Journal of Personality and Social Psychology*, **38**,689-703.
- Martin, D.C. (1987) Factors influencing pay decisions: Balancing managerial vulnerabilities, *Human Relations*, **40**,417-430.
- Maurer, T.J., & Taylor, M.A. (1994) Is sex by itself enough? An exploration of gender bias issues in performance appraisal, *Organizational Behavior and Human Decision Processes*, **60**, 231-251.
- McIntyre, R. M., Smith, D., & Hassett, C.E. (1984) Accuracy of performance ratings as affected by rater training and Perceived purpose of rating, *Journal of Applied Psychology*, **69**,147-156.
- Mohrman, A. M., & Lawler, E. E. (1983) Motivation and performance appraisal behavior. In F. Landy, S. Zedeck, & J. Cleveland (Eds.) *Performance measurement and theory*, Hillsdale, NJ: Lawrence Erlbaum. Pp.173-189.
- Motowidlo, S.J., & Schmid, M. J. (1999) Performance assessment in unique jobs. In D. Ilegen, & E. Pulakos (Eds.) *The changing nature of performance*. San Francisco: Josey-Bass. Pp.59-83.
- Murphy, K. R., & Balzer, W. K. (1986) Systematic distortions in memory-based behavior ratings and performance evaluations: Consequences for rating accuracy, *Journal of Applied Psychology*, **71**,39-44.
- Murphy, K. R., & Balzer, W. K. (1989) Rater errors and rating accuracy, *Journal of Applied Psychology*, **74**,619-624.
- Murphy, K. R., & Cleveland, J. N. (1991) Performance appraisal: An organizational perspective. Boston: Allyn & Bacon.
- Murphy, K. R., & Cleveland, J. N. (1995) *Understanding performance appraisal* London: Sage.
- Murphy, K. R., & Costants, J. I. (1987) Behavioral anchors as a source of bias in rating, *Journal of Applied Psychology*, **72**,523-579.
- Murphy, K. R., & Costants, J. I. (1988) Psychological issues in scale format research: Behavioral anchors as a source of bias in rating. In Cardy, S. Peiffer, & J. Newman (Eds.), *Ad-*

- vances in information processing in organizations* (Vol.3). Greenwich, CT: JAI Press.
- Murphy, K. R., & Davidshofer, C. O. (1988) *Psychological testing: Principles and applications*. Englewood Cliffs, NJ: Prentice Hall.
- Murphy, K.R., Jako, R. A., & Anhalt, R. L. (1993) Nature and consequences of halo error: A critical analysis, *Journal of Applied Psychology*, 78,218-225.
- Murphy, K.R., Philbin, T.A., & Adams, S.R. (1989) Effect of purpose of observation on accuracy of immediate and delayed performance ratings, *Organizational Behavior and Human Decision Processes*, 43,336-354.
- Pulakos, E.D. (1986) The development of training programs to increase accuracy in different rating tasks, *Organizational Behavior and Human Decision Processes*, 38,76-91.
- Quattrone, G. A., & Jones, E. E. (1980) The perception of variability within in-groups and out-groups: Implications for the law of small numbers, *Journal of Personality and Social Psychology*, 38,141-152.
- Robbins, T. L., & DeNisi, A. S. (1994) A closer look at interpersonal affect as a distinct influence on cognitive processing in personal evaluation, *Journal of Applied Psychology*, 79,341-353.
- Roach, D.E. & Wherry, R. J. (1970) Performance dimensions of multi-line insurance agents, *Personnel Psychology*, 23,239-250.
- Saal, F. E., Downey, R. G., & Lahey, M. A. (1980) Rating the rating: Assessing the quality of rating data, *Psychological Bulletin*, 88,413-428.
- Salvemini, N. J., & Reilly, R. R. (1993) The influence of rater motivation on assimilation effects and accuracy in performance ratings, *Organizational Behavior and Human Decision Processes*, 55,41-60.
- 佐々木美加 (1997) 学歴が人物評価に与える影響 産業・組織心理学研究, 10,145-154.
- Schneier, C. E. (1977) Operational utility and psychometric characteristics of behavioral expectation scales, *Journal of Applied Psychology*, 62,541-548.
- Sisson, E. D. (1948) Forced choice: The new Army rating, *Personnel Psychology*, 1,365-381.
- Smith, K. H., & Kendall, L. M. (1963) Retranslation of expectation: An approach to the construction of unambiguous anchors for rating scale, *Journal of Applied Psychology*, 47,149-155.
- Smither, J. W., & Reilly, R. R. (1987) True intercorrelation among job components, time delay in rating, and rater intelligence as determinants of accuracy in performance ratings, *Organizational Behavior and Human Decision Processes*, 40,369-391.
- Sumer, H. C., & Knight, P. A. (1996) Assimilation and contrast effects in performance ratings: Effects of rating the previous performance on rating subsequent performance, *Journal of Applied Psychology*, 81,436-442.
- Thorndike, E. L. (1920) A constant error in psychological ratings. *Journal of Applied Psychology*, 4,25-29.
- Tsui, A. S., & Barry, B. (1986) Interpersonal affect and rating errors, *Academy of Management Journal*, 29,586-599.
- Wayne, S. J., & Kacmar, K. M. (1991) The effects of impression management on the performance appraisal process, *Organizational Behavior and Human Performance*, 28,356-378.
- Wells, F. (1907) A statistical study of literary merit, *Archives of Psychology*, 1(7)
- Williams, K. J., DeNisi, A. S., Blencoe, A. G. & Cafferty, T. P. (1985) The role of appraisal purpose: Effects of purpose on information acquisition and utilization, *Organizational Behavior and Human Decision Processes*, 35,314-339.
- Woehr, D.J. (1994) Understanding frame-of-reference training: The impact of training on the recall of performance information, *Journal of Applied Psychology*, 79,525-534.

柳澤さおり・古川久敬（1998a）「評価」に関する態度が評価に及ぼす効果についての研究，産業・組織心理学会第14回大会，Pp.172-174.

柳澤さおり・古川久敬（1998b）評価の寛大さにおよぼす「評価に関する態度」の媒介効果，九州大学教育学部紀要，43,Pp.293-300.

柳澤さおり・古川久敬（1999）評価者が持つ“内的準拠枠”の探索的研究，産業・組織心理学会第15回，Pp.136-139.