

アセスメント・センターの概要と心理測定上の課題

HAYASHI, Yoichiro / 林, 洋一郎

(出版者 / Publisher)

法政大学キャリアデザイン学部

(雑誌名 / Journal or Publication Title)

Bulletin of the Faculty of Lifelong Learning and Career Studies / 法政大学キャリアデザイン学部紀要

(巻 / Volume)

8

(開始ページ / Start Page)

211

(終了ページ / End Page)

228

(発行年 / Year)

2011-03

(URL)

<https://doi.org/10.15002/00007391>

〈研究ノート〉

アセスメント・センターの概要と 心理測定上の課題

法政大学キャリアデザイン学部准教授 林 洋一郎

過去と未来の業績の関係

実証的な心理学の分野において、“Past behavior predicts future behavior” という有名な言葉がある (Ouellette, & Wood, 1998)。心理学の目標が、個人の行動の記述、予測、制御にあるとするなら、当人の過去の経験や行動が将来の行動予測に役に立つという仮定である。

例えば、企業や組織が新しい人を採用する際、応募者の履歴 (bio data)⁽¹⁾を参考として採用の可否を決定することは十分にありえるだろう (Mael, 1991)。我々も、初見の相手がどんな人物であるかを考える際、当人のこれまでの経験、過去の業績などを手がかりとしながらその人物を理解しようとする。過去から未来を予測するという発想は、学術的にも日常的にも一般的なものである。

しかしながら、過去の体験や業績があまり参考にならない状況や過去の情報が入手できない状況もある。例えば、「名選手、必ずしも名監督ならず」という格言が野球界にあると聞く。この格言は、プレイヤーとしての経験や業績が指導者としての力量を保証するとは限らないことを含意している。プレイヤーとしての業績 (行動) が指導者としてのコーチング技術や実績 (行動) を必ずしも予測しない可能性である⁽²⁾。換言すれば、プレイヤーとしてのスキルや経験とコーチで求められるスキルや経験は違う。よってプレイヤーとしての過去の経験や業績が、コーチとしての未来のパフォーマンスを保証しないという示唆である。これとは別に、新卒採用は、基本的に就業経験のない学生応募者か

212 法政大学キャリアデザイン学部紀要第8号

ら選抜を行う。これは、職業人として優秀さを直接的に示す情報が十分に手に入らない状況である⁽³⁾。

採用活動の効果性は、入社後に活躍し業績をあげる人材を受け入れることが目標である。よって将来の業績や行動を直接的に役立つ情報や経験がない場合、それらを間接的に知る必要がある。そのために種々のアセスメント技法が開発されてきた。例えば、日本の採用場面においては、ペーパー・ペンシル型テスト (paper pencil test) —これには、認知能力テスト (言語性検査)、パーソナリティ検査などが含まれる、や面接 (interview) もよく使われている。ペーパー・ペンシル型テストの例としては、リクルートが開発、販売しているSPI 2をあげることができる。これら以外にも、誠実性検査 (integrity test) などが考えられる。

以上の技法は、将来の行動や職務で求められる資質などを直接的に観察して捕捉するのではなく、認知能力やパーソナリティなどに焦点を当てて、一般的な視点からいわば間接的に推測する手段である。これに対して、将来、求められる職務や課題解決の状況などを模擬的に創出し (simulate)、そうした状況における応募者や被評価者 (assesee) の振る舞いやパフォーマンスを直接観察するアセスメント技法がある。この技法そのものは演習 (simulation) と呼ばれるが、演習を中心に複数の標準化された測定技法を組み合わせ、テスト・バッテリーとして標準化されたものがアセスメント・センター (Assessment Center = AC) である。

ACは、“Past behavior predicts future behavior” が成立しない状況において、未来に求められる状況を模擬的に作り出し、その場面における行動やパフォーマンスを手がかりにして未来の行動を予測しようという戦略である。よって、過去の行動や経験が入手できない場合でも、模擬状況の行動によって将来の行動や業績を予測できるという利点がある。本論文は、ACに関する海外の研究を中心にその概要を紹介することが目的である。

アセスメント・センターの定義と概要

ACは、複数の標準化された測定道具からなるアセスメントプロセスであると定義できる。様々なアセスメント技法が組み合わされた一種のテスト・バツ

テリーである。複数の評価者 (assessors) が、一人の候補者 (candidate) や被評価者 (assesse) の行動について、主に演習を用いて評定する。個々の演習に臨む候補者の行動を評定するにあたって、事前に職務に関連した種々の次元 (dimension) あるいはコンピテンシー (competency) を設定する。問題解決、オーラル・コミュニケーション・スキルといった次元である⁽⁴⁾。演習とは、現実の職場場面に近似した模擬環境を作り出し、候補者に擬似環境で種々の課題に従事させるというアセスメント技法である。

複数の評価者が、候補者の行動やパフォーマンスを事前に設定した次元や観点に従って多角的に評価するという特徴も持つ。複数の評価者から得られる個々の評価スコアは統計的な手法によって統合されて、総合得点のような値が算出される。

ところでACを実施する目的は、次の3種類に区分される。1つ目は、昇進・昇格など、選抜 (selection purpose) を目的とするものである。2つ目は、診断的なアセスメント・センターと呼ばれるもので、個人の技量や能力を診断することを目的 (diagnostic purpose) とするものである。ACから得られる評価に基づき、各個人に必要な教育・研修の要素を把握する。さらに適材適所の配置を判断する材料に資することを目的とする。3つ目は、管理者に必要な職務遂行スキルを開発したり、発達させたりすることを目的 (developmental purpose) するものである。発達の・開発的なACとである。

当初のACは、候補者の選抜を目的として開発、実行された。現在は人材育成の観点からACを活用しようという考えが広まりつつある。選抜と診断を目的とするACが、伝統的なアセスメント・センターといえるだろう。一方で、開発的なACは、Developmental Assessment Centerと呼び、DACと略記される場合もある (Thornton & Rupp, 2006)。

なお、アセスメント・センターとは、日本語における医療センターやショッピング・センターのように箱物の施設や建物があるという訳ではない。様々なアセスメント技法を組み合わせたテスト・バッテリーを意味する。産業・組織における心理測定技法の総合パッケージと考えてよい。よって、Assessment Center Methodと呼ばれる場合もある。

表1に典型的なACの概略を整理した。数日間をかけて実施する研修合宿に

表1：ACの典型的な概要と流れ

準備：評価者訓練 (AC実施前)	準拠枠訓練などを用いて評価者訓練を行う。複数の評価者 (multiple assessors) が、一人の候補者を評価する。よって評定者間のばらつきをコントロールする必要がある。評定者訓練は、評価者の視点や軸を一貫させ、評定者間信頼性 (Inter Rater Reliability) を高めることが期待される。
ACの実施 (1日～3日の期間)	複数の評価者が、評価者と被評価者の割合は、1：2が一般的である。「リーダーレス・グループ・ディスカッション」、「ケーススタディ」、「ロール・プレイ」、「プレゼンテーション」、「イン・バスケット」など。
複数の評価の調整と すり合わせ (AC実施の当夜)	複数の評価者は、各被評価者のパフォーマンスについて判断し、討議を行う。最終的に5件法によるリッカート尺度を使用して各評価者の総合的なパフォーマンスについて評価する。
結果の フィードバック	口頭および文書で被評価者に結果を報告する。DACの場合は、その後の人材開発に役立てるので詳細なフィードバックとなる。

近い状況といえるかもしれない。

ACの10の基本要件 (10 essential elements for AC)

International Task Force on Assessment Center Guidelines (2009) は、ACを構成するために必要となる基本10要素を提示している。

1) 職務分析／コンピテンシー・モデリング (job analysis / competency modeling)：ある職務で成功するために求められる次元 (dimension) やコンピテンシー (competency) を、職務分析によって特定する必要がある。職務分析によって、ACの中で評価すべき内容を導く。

2) 行動の分類 (Behavioral classification)：候補者が示す行動は、意味のある適切なカテゴリーに沿って分類され、評価されるべきである。カテゴリーは、具体的には、行動次元 (behavioral dimensions)、属性 (attributes)、特性 (characteristics、適性 (aptitudes)、資質 (qualities)、技能 (skills)、能力 (abilities)、コンピテンシーや知識である。

3) **適切なアセスメント技法**：適切なアセスメント技法を選択することによって、職務分析で同定された次元とその次元と結びつきの強い行動やパフォーマンス（候補者の）を引き出す工夫をする。両者の関連は、行動×アセスメント技法のマトリックスによって表現される。

4) **マルチプルなアセスメント (multiple assessment)**：複数のアセスメント技法が用いられなければならない。これらの技法は、ペーパー・ペンシル・テスト、面接、演習、などである。自己評定だけでなく360度評価のような他者からの評価も組み入れるとよい。

5) **演習 (simulation)**：実際の職務場面を模した状況を創出する。そして、候補者がその場で提供される様々な刺激にどのような行動反応を示すかに注目する技法である。演習にはリーダーレス・グループ・ディスカッション (Leaderless Group Discussion=LGD)、イン・バスケット (In-basket)、ロール・プレイ (role play)、プレゼンテーション、ファクト・ファインディング (fact finding)、ビジネス・ゲームなどが含まれる。近年、演習を創出するにあたって、様々なメディアやテクノロジーが用いられている。コンピュータを用いたヴァーチャルな状況、ビデオ、インターネットを用いたものなどがある。

6) **複数の評価者 (multiple assessors)**：候補者を評価する人物が、必ず複数であること。これによって評価のバイアスや偏りを防ぐことができる。評価者には、人口統計学 (e.g., 人種、国籍、年齢、性別など) と経験 (e.g., 組織での地位、部門、管理職かどうか、心理学者など) という点から多様な人材をそろえることが大事である。

7) **評価者訓練 (Assessor training)**：評価者は、評価の次元や視点などについて習熟していなければならない。こうした手続きは、評価者間のばらつきを抑制し、一貫した評価を可能にする。評価者たちは、複数人から得られる評価スコアに関して、議論を通して、統合していく必要がある。複数スコアのすり合わせに方法についても学ぶことが望ましい。最近、準拠枠訓練 (frame of reference training) の有効性が認識されている。FOR とは、複数の評定者が一人のある候補者を評価する場合、共通の軸やフレームに従って評価判断を行うことが強く望まれる。評定者によって評価の視点や軸がばらつくと、信頼性の高い評価を行うことが難しくなる。FOR とは、複数評定者が参照する枠組

216 法政大学キャリアデザイン学部紀要第8号

み（フレーム）を一致させ、評定者の違いによる測定誤差を除去しようという試みである。もちろん完全に取除くことは不可能であるが、候補者パフォーマンスの良し悪しの判断について、ある程度の一貫した評価を下すことを目指している（Woehr, 1994; Dierdorff, Surface, & Brown, 2010）。

8）行動（業績）の記録と得点化：評価者は、候補者が示す行動を正確に観察し記録する必要がある。ここでは、行動基準評定尺度法（Behaviorally Anchored Rating Scales = BARS）、行動観察尺度法（Behavioral Observation Scales = BOS）、行動観察尺度法（Behavioral Observation Scales = BOS）などが用いられる。

9）データの統合（data integration）：ACは、複数のアセスメント技法から得られた得点を集計し、総合的なAC評価を導出する必要がある。その際、次元×技法のマトリックスを参考にする。はじめに技法別に各次元の得点を算出し、技法をまたいで次元ごとに次元の総合点を算出する。最終的に、総合的な単一評価のようなスコアを算出する。しかし全く異なる次元を単一のスコアに収束させることは、統計的に必ずしも許容されない。データ統合のテクニックに関しては、さらなる洗練と研究が必要であるといわれている（Thornton & Rupp, 2006）。

代表的な演習の技法

オーラル・プレゼンテーション（oral presentation）：候補者は、ある特定の課題に関する情報が与えられ、それを精査することを求められる。その上で取引先や社内の上位者など、重要な他者に対してプレゼンテーションするという状況を設定する。この状況で、発表資料を作成し、それを使ってプレゼンテーションを行うように求められる。準備時間は、10分から1時間くらいであるが、与えられる課題や情報の複雑さの程度に応じて変動する。プレゼンテーションの時間は、おおよそ30分くらいである。

ロール・プレイ（role-play）：予想されるポジションの役割を模擬的に演じるという演習である。特に、そのポジションに就いた場合と一緒に仕事を行う人物とのやりとりや相互作用に注目する演習である。管理職というポジションについてロール・プレイが実施される場合、候補者は管理者として部下、上

司、同僚、顧客などとの交互作用を行う状況が創出される。模擬状況の部下や上司は、実験協力者が役割を演ずる。

リーダーレス・グループ・ディスカッション (Leaderless Group Discussion =LGD)：複数の候補者（だいたい4～8人）が決められた複数の課題を解決するように求められる。時間はおおむね1時間くらい。候補者は、課題解決に向けての議論を行い、さらにグループとして書面で解決策を示すように求められる。役割が付与されない形式と支店長など、特定の役割が与えられる形式の2種類ある。チームワークやリーダーシップ・スキルなどの行動次元を測定することに役立つ。

イン・バスケット (In-basket)：個人単位で実施される演習である。典型的なイン・バスケット演習は、急病で職を離れた管理者の仕事、急遽、引き継ぎというような状況を候補者に提示する。候補者は、はじめに会社の組織構造や理念について説明を受ける。次いで、前任者が残した未処理の書類箱が示される⁶⁾。そこには、様々なメモ、手紙、要望書、応募者書類など、喫緊に処理が必要な案件が残されている。現在は、電子情報の形式でこうした情報が与えられることもある。さらに候補者は、ニューヨークで開かれる取締役会で、引き継ぎ案件について報告をしなければならないというような状況が提示される。よって候補者は、1時間から2時間くらいで案件を迅速に処理しなければならない。制限時間が過ぎた後、案件をどのように処理する予定であるかを整理し、記述することが求められる。この記述回答が評価材料となる。イン・バスケットで再現されている状況は、候補者が、将来、実際に就任する状況に類似したものである。

オーラル・ファクト・ファインディング (Oral Fact Finding)：候補者は、はじめに問題となっている案件や事柄そして背景に関する記述を読んで情報を得る。さらに情報提供者を演じる役割演技者に適切な質問を投げかけ、必要となる情報を得る（事実を発見=fact finding）という演習である。この演習の核は、役割演技者からいかに価値のある情報を引き出すかという点である。最終的に、得られた情報に基づいて解決策を提案し、説明することが求められる。

その他の演習：演習とはいえないが、能力検査、パーソナリティ検査、略歴作成などペーパー・ペンシル型の技法や面接などの技法が用いられる。ビジネ

ス・ゲームが演習に組み入れられることもある。

ACの歴史的発展

ここではACの歴史的発展の経緯について説明する。産業・組織心理学の理論や技法の多くが軍隊や兵士の研究から発展したように、ACも当初は兵士の適性やスキルを検査するための手段として開発された。

ACは、1930年代初めに、ドイツの心理学者が軍の仕官登用にあって開発したプログラムに原型があるといわれている。このプログラムは、リーダーシップ潜在特性を測定するために開発された。複数の評価者が、候補者の行動を様々な次元から評価し、いろいろな演習を組み合わせることで個人の複雑な行動を理解するという方向性はこの時点で打ち出されていた。その後、同様の方法を用いたアセスメント技法が、イギリス軍、カナダ軍、オーストラリア軍に広がっていった。

次に、第二次大戦中に、米国の戦略事務局（Office of Strategic Services = OSS）がスパイの作業員を選抜するために開発したプログラムがある。これは、ACが洗練された技法へと発展する契機となった⁶⁾。なお、戦略事務局は、1947年にアメリカ中央情報局（Central Intelligence = CIA）に改組され、現在に至っている。

OSSのプログラムは、種々のアセスメント技法を組み合わせた3日間から構成されるプログラムであった。なお、ワーク・モチベーションの欲求理論で有名な、Henry Murrayもプログラムの開発に関与している。

行動次元やコンピテンシーというより、パーソナリティ（一種の潜在的な内的要因）を包括的に捉えようとした。例えば、木材を使って木造の家を作るという模擬状況における候補者の対応に注目する状況テストがある（situational test）。ここでタイプの異なるふたりの助手と一緒に仕事することが求められる。助手の一人は、受身的で怠け者であり、もう一人は自己主張が強く、勝手に物事を進めようとする。彼らは、実は評価者でもあるが、候補者の働きを様々な点から妨害する。候補者が、こうした困難な状況に対してどのように対処するかを評価することによってリーダーシップ能力や感情の安定性を測ろうとした。

1950年代中頃に、民間転用が本格的に始まった。AT&T（アメリカ電話電信会社）のプログラムは、ACをビジネス場面に始めて応用した事例として良く知られている。これはOSSモデルを管理者の選抜や候補者の発掘に適用した。OSSとの大きな違いは、演習を通して観察できる職務関連の行動次元やコンピテンシーを重視した点である。個人の資質の中でも、目に見える、観察可能な顕在的側面を重視している。イン・バスケット、リーダーレス・グループ・ディスカッション、ロール・プレイといった現在のACでも頻繁に用いられる演習を含むものであった。さらに、電信電話業界に従事する若手管理者の選抜でなく、ACを通じた人材育成という側面も有していた。ここに現在のACの様式がほぼ完成した。

現在、ACは職務に関連する様々な知識、技能、能力、その他の特性（Knowledge, Skills, Abilities and Other characteristics = KSAOs）を評価するための総合パッケージ（テストバッテリー）と考えられている。さらに発達・育成目的のACが優位になりつつある。International Task Force on Assessment Center Guideline(2000, 2009)がACの標準形を明示し、倫理的配慮についての指針も提起している。

ACの課題と活用

最後に、ACの課題や問題点について説明したい。はじめにACの信頼性と妥当性について概観する。特に、構成概念妥当性（construct validity）の問題について論じる。次に、ACの我が国における応用についてコスト・ベネフィットの問題も指摘しながら論ずる。

ACと信頼性と妥当性

ACの場合、複数の評価者が、候補者の行動や業績を各次元に即して評価する。よって評定者間信頼性（Inter Rater Reliability = IRR）が問題となる。IRRは、複数の評定者が下す判断がどれくらい一致（類似）しているかどうかを示す。既存の研究では、一定の値が得られている。おおよそ.60から.90の値が報告されている（Lievens, 2002; Thornton & Rupp, 2006）。ただしこれは次元の種類や評価者訓練の精度によって変動すると考えられている。信頼性を高め

220 法政大学キャリアデザイン学部紀要第8号

るためには、準拠枠訓練 (Frame of Reference Training) の有効性が強調されている。

予測妥当性 (predictive validity) に関しては、表2のようになっている。他の予測因 (predictors) (面接や認知能力など) に比べて妥当性係数が特に低いことはないということが良く分かる。表2における研究以外にも妥当性係数が報告されているが、.25-.39という結果も得られている (Arthur et al, 2003)。

ところで心理測定論から指摘されるACの課題は、構成概念妥当性 (construct validity) である。構成概念妥当性については様々な考え方があがるが、

表2：アセスメント技法と職務業績 (job performance) の関連性 (Levy, 2009 より)

予測因	妥当性係数の推定値	推定値が掲載されている研究	研究タイプ
一般認知能力 (General cognitive ability)	.53	Hunter & Hunter, 1984	Meta-analysis
	.62	Salgado et al, 2003	Meta-analysis
	.48	Bertua et al, 2005	Meta-analysis
個別認知能力 (Specific cognitive ability)	.40 - .50	Gatewood et al, 2007	Narrative review
	.35 - .56	Salgado et al, 2003	Meta-analysis
	.35 - .50	Bertua et al, 2005	Meta-analysis
精神運動能力 (Psychomotor ability)	.53	McHenry et al, 1990	Meta-analysis
	.40	Hunter & Hunter, 1984	Meta-analysis
正直性テスト (Integrity tests)	.34 - .47	Ones et al, 1993	Meta-analysis
			Meta-analysis
状況判断テスト (Situational Judgment tests)	.34	McDaniel et al, 2001	Meta-analysis
経歴・履歴書 (BioData)	.52	Vinchur et al, 1998	Meta-analysis
			(sales)
			Meta-analysis
面接 (Interview)	.37	McDaniel et al, 1994	Meta-analysis
			Meta-analysis
構造化面接 (Structured)	.44	McDaniel et al, 1994	Meta-analysis
非構造化面接 (Unstructured)	.33	McDaniel et al, 1994	Meta-analysis

本論文は、Campbell & Fiske (1959) が提唱した多特性多方法行列 (Multi Trait - Multi Method Matrix = MTMM 行列) に基づき評価する方法に従う。MTMM 行列とは、ある特性を複数の方法によって測定する状況から得られる相関行列である。AC はまさにこの状況に一致する。つまり多特性 = 行動次元やコンピテンシーであり、多方法 = 演習の種類である。

Campbell & Fiske (1959) は、同一次元で異なる方法 (mono dimension hetero method) の間の相関は、異なる次元で同一の方法 (hetero dimension mono method) による相関よりも高くなるべきであると主張した。これが証明された場合に、構成概念妥当性ありと判断される (表3参照)。つまり構成概念妥当性の要件は、特性間の相関の強さは、同一方法に起因する相関の強さを上回るということである。つまり方法による相関は測定誤差であるので、これはコントロールされるべき相関である。

AC における例として、問題解決能力と職務への積極性をリーダーレス・グループ・ディスカッションとオーラル・プレゼンテーションで測定した状況を考

表3：アセスメント・センターにおける多特性多方法行列 (MultiTrait-MultiMethod Matrix, MTMM 行列)

	演習1 (LGD)			演習2 (In-basket)		
	A	B	C	A	B	C
演習 1						
次元 A	(.74)					
次元 B	.22	(.81)				
次元 C	.17	.21	(.48)			
演習 2						
次元 A	.66	.15	.11	(.38)		
次元 B	.02	.72	.17	.19	(.41)	
次元 C	.06	.09	.69	.29	.36	(.27)

222 法政大学キャリアデザイン学部紀要第8号

えてみたい。この場合、リーダーレス・グループ・ディスカッションで測定した問題解決能力とオーラル・プレゼンテーションで測定した問題解決能力の間の相関（同一特性で異なる方法）は、リーダーレス・グループ・ディスカッションにおける問題解決能力と職務への積極性の間の相関（異特性で同一の方法）よりも高い値を示すことが期待される。

しかしこれまでの研究は、これとは逆の結果を示すことが多い。つまり同次元異演習の相関係数のほうが、異次元同演習という相関係数よりも高い値を示す傾向が報告されている（Sackett & Dreher, 1982; Klimoski & Brickner, 1987; Lievens, 2002; 2009）。これは、“Construct-Validity puzzle”と呼ばれている（Sackett & Dreher, 1982; Lievens, 2009）。この結果は、ACは構成概念妥当性に欠ける不適切なアセスメント技法であるということを示唆する。

この問題については現在も様々な角度から議論が行われている。パーソナリティの人間 - 状況論争（Michel, 1968）や特性賦活理論（Trait Activation Theory=TAT; Tett & Guterman, 2000）といった理論に従えば、同一演習間の相関が高まることは不自然な結果ではない。これらの理論は、パーソナリティとは通時的にあるいは状況を超えて首尾一貫（coherent）したのではなく、状況によって変化する性質であると強調する。よって状況に影響されて（引っ張られて）、候補者が演習ごとに振る舞いを変化させても、それは自然なことであるという考え方である。

そこで、ある行動次元やコンピテンシーを測定するとしても、その行動次元をうまく吸い取ることのできる演習とそうでない演習があると思われる。つまり焦点をむける次元に適した演習を選ぶべきであるという提起がなされている。例えば、Lievens, Chasteen, Day, & Christiansen (2006)、は、性格5大因子を取り上げ、各因子の測定に適した演習は異なると報告している。この点については、稿を改めて詳しく論じたい。

ACの日本への適用

ACは、上に説明したように、ペーパー・ペンシル型のテストとは異なりより現実性（fidelity）の高いテストであり、また観察可能な行動次元や業績に注目するという点で有効性の高い技法であると思われる。

一方で、ACを実施することは組織に大きなコストして跳ね返ってくる。この点が、現実にビジネス場面に応用する場合の大きな障害となると考えられる。特に、日本にACを本格的に導入する場合、コスト・ベネフィットの観点から割に合わないという批判にさらされやすいと考えられる。

しかし、日本の新卒採用の場面を見ると、ACの演習を部分的に導入している。例えば、オーラル・プレゼンテーション、リーダーレス・グループ・ディスカッション、状況インタビューなどは我が国の新卒採用場面でも使用されている。こうしたアセスメント技法を、ACの観点から、より信頼性・妥当性の高いツールに改善する工夫が求められる。採用企業の多くは、複数の評価者が特定の候補者や応募者の評定を行うと考えられる。複数評価の信頼性について、評定者間信頼性 (Inter Rater Reliability) などによって検証する必要があると思われる。

採用活動において信頼性や妥当性を高めることは確かに手間が掛かる。しかし、フェアな採用活動につながる。フェアであるという社会的評判は長期的に考えれば組織にとってプラスになると期待される。

そもそも新卒採用とは、“Past behavior predicts future behavior” という考えが役に立たない場面である。よって、ACが活用される余地も多いのでないだろうか。我が国の採用活動は、新卒採用のウェイトが極めて高く、米国などに比べれば雇用調整も難しいと考えられている。そこで筆者は、新卒採用場面において、演習技法を部分的にも導入してはどうかという提案をしたい。多少のコストを投入しても、組織にフィットする人材を見つけることは長期的な利得につながると思われる。新卒者に対するACの活用に関しては、内外を問わず構造化されたプログラムが開発されているとはいえない。今後は、就業経験の無い候補者や応募者の選抜にACをどのように活用すべきかについて、実践と理論の双方の観点から探求することが大切であろう。

[注]

- (1) 日本における履歴書とはやや異なる。パーソナリティ、スキルなどの種々の個人特性を含むようなより包括的なドキュメントである。
- (2) 「名選手必ずしも名監督ならず」という格言は、科学的に証明されてい

224 法政大学キャリアデザイン学部紀要第8号

る訳ではない。客観的な事実かどうかについて詳細な検討が必要となろう。2010年に就任一年目でロッテを日本一に導いた西村徳文監督、中日の落合博満監督、ヤクルトや楽天の監督を歴任した野村克也氏など名選手でありながら監督としても実績をあげている。筆者は、“Past behavior predicts future behavior”を説明するための例としてこの格言を取り上げた。この格言を妥当なものとして認めている訳ではない。

- (3) それゆえ米国を中心に、妥当な選抜技法を特定するために、メタ分析を用いた予測的妥当性研究が数多く提起されている。表2を参照。
- (4) Arthur, Anthony, McNelly, & Edens (2003) は、AC で使用される代表的なディメンション（コンピテンシー）として配慮・他者への意識（consideration/awareness of others）、コミュニケーション（communication）、意欲・動因（drive）、他者に対する影響行使（influencing others）、組織化と計画（organizing and planning）、問題解決（problem solving）である。BIG 6とも考えられる一般的な（general）な次元。各次元は、さらに細分化された下位次元を従えている。
- (5) In-basket は未決の書類が格納されている箱を意味する。
- (6) OSS がスパイや課報員の適性を知るために AC を活用したという事実は、AC の特質をよく反映している。スパイとは非常に個別性・特異性の高い“職務”であると思われる。よって候補者のスパイ適性を知るために、過去の経験や実績はあまり参考にならないと考えられる。そこで、実際の課報活動を模した状況を作り出し、演習を通じてスパイに適した人材を発掘したと考えられる。

[引用文献]

- Arthur, W., Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125–154.
- Bertua, C., Anderson, N. and Salgado, J.F. (2005). The Predictive Validity of Cognitive Ability Tests: A U.K. meta-analysis. *Journal of Occupational and Organizational Psychology, 78*, 387–409.
- Campbell, D. T., & Fiske, D. W (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

- Dierdorff, E. C., Surface, E. A. & Brown, K. G. (2010). Frame-of-reference training effectiveness: Effects of goal orientation and self-efficacy on affective, cognitive, skill-based, and transfer outcomes. *Journal of Applied Psychology*, *95*, 1181–1191.
- Gatewood, R. D., Field, H. S., & Barrick, M. (2007). *Human Resource Selection* (6th ed.). Fort Worth, TX: Harcourt College.
- Hermelin, E., Lievens, F. and Robertson, I.T. (2007) The Validity of Assessment Centres for the Prediction of Supervisory Performance Ratings: A meta-analysis. *International Journal of Selection and Assessment*, *15*, 405–411.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin*, *96*, 72–98.
- International Task Force on Assessment Center Guidelines. (2000). Guidelines and ethical considerations for assessment center operations. *Public Personnel Management*, *29*, 315–331.
- International Task Force on Assessment Center Guidelines. (2009). Guidelines and Ethical Considerations for Assessment Center Operations. *International Journal of Selection and Assessment*, *17*, 243–253.
- Klimoski R, Brickner M. (1987). Why do assessment centers work? The puzzle of assessment center validity, *Personnel Psychology*, *40*, 243–260.
- Levy, P. E. (2006). *Industrial/organizational psychology: Understanding the workplace, 3rd edition*. New York: Worth Publishers.
- Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of assessment centers: An examination of assessor and assessee effects. *Journal of Applied Psychology*, *87*, 675–686.
- Lievens, F. (2009). Assessment centres: A tale about dimensions, exercises, and dancing bears. *European Journal of Work and Organizational Psychology*, *18*, 102–121
- Lievens, F., Chasteen, C. S., Day, E. A., & Christiansen, N. D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. *Journal of Applied Psychology*, *91*, 247–258.
- Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata

- items. *Personnel Psychology*, 44, 763–927.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L. & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79, 599–616.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 80, 730–740.
- McHerny JJ, Hough LM, Toquam JL, Hanson MA. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, 43, 335–354.
- Mischel, W. (1968). *Personality assessment*. New York: Wiley. 託摩武俊 (監訳) 1992 パーソナリティの理論－状況主義的アプローチ 誠信書房.
- Ones DS, Viswesvaran C, Schmidt FL. (1993). Comprehensive meta-analysis of integrity test validation: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78, 679–703.
- Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin*, 124, 54–74.
- Sackett P. R, Dreher G. E. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology*, 67, 401–410.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & deFruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, 56, 573–606.
- Tett, R. P., & Gutterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, 34, 397–423.
- Thornton, G. C., III, & Rupp, D. E. (2006). *Assessment centers in human resource management*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Tsacoumis, S. (2007). Assessment centers. In D. L. Whetzel and G. R. Wheaton (Eds.), *Applied measurement: Industrial psychology in human resources management*. Mahwah, NJ: Lawrence Erlbaum Publishers.

Vinchur, A. J., Schippmann, J. S., Switzer, F. S., III, & Roth, P. L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology, 83*, 586–597.

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.

ABSTRACT

A review of assessment center and its measurement issues

Youichiro HAYASHI

This article reviews concepts and structures of assessment center and discusses how it can be applicable to the Japanese organization. First of all, I explain the basic tenets of assessment center from essential elements, types of simulations and history. Here, it is emphasized that the simulation methods in assessment center are expected to be reliable and valid tool for future behaviors among candidates, especially in case that the past behaviors among them are not available or predictive of the future performance. Secondly, the issue of construct validity is raised with an eye to the MTMM (MultiTrait- MultiMethod) matrix and then a couple of ways to handle this problem are discussed. Finally, I suggest that assessment center methods, involving simulation exercises, are to be useful selection tools among college students who generally have no occupational experiences.