

採用選考における作文評価

—達成動機測定の試み—

人事測定研究所 鷺坂由紀子・二村英幸・山岸建太郎

A Narrative Essay Test in Personnel Selection: A Trial of Measuring the Achievement Motive

Yukiko SAGISAKA, Hideyuki NIMURA, & Kentaro YAMAGISHI,
(Human Resource Research Institute)

This study provides the information of reliability and implications of narrative essay tests for measuring achievement motive as a part of employee selection processes. To develop the key achievement motive explicit and concrete rating criterion, the TAT scoring method was applied to data ($n=100$) of essay tests gathered from seven raters. Three raters were chosen from entry level workers and the other four were professional writers of verbal testing items, forming two contrast groups. The reliability of the achievement motive ratings was calculated for each group by the interrater reliability approach, resulting that there was no significant difference between the groups. Coefficients of each group's achievement motive, general mental ability and personality traits were calculated, suggesting the possibility that the achievement motive ratings thus derived from essay tests implies individuality.

1. 背景と問題

近年、IT技術の進展を背景として企業の採用選考状況が大きく変化している。特に募集や初期選考段階においてはWebを活用した採用方法が飛躍的に広がりつつある。例えば、Web上でエントリーシートと呼ばれるフォーマットに属性情報と同時に小論文や作文を書かせ、その内容によって第一次選考を行う企業も数多く見受けられるようになった。例えば、「10年後の自分」や「あなたの個性を表現してください」といったテーマで、500字から1000字程度で記述させるものが多い。しかしながら小論文、作文によって得られる個人差の尺度に関する研究は、エッセイの機械的な評定についてのBurstein, Kukich, Wolff, Lu, Chodorow, Braden-Harder, and Harris (1998), 入学試験における石井(1981)といった研究がわずかに見られるにとどまり、

本研究は、経営行動科学編集委員会の推選を受け、第3回年次大会において報告された内容に加筆したものである。

第1筆者電子メール連絡先：

yukiko_sagisaka@hrr.co.jp

人事測定研究所 URL: <http://www.hrr.co.jp>

採用選考の実際場面における有効性に関する研究実績は国内外において皆無である。

このようなエントリーシートに代表される作文評価は、客観的なアセスメント手法としてどの程度の信頼性や妥当性を有しているのであろうか。本研究の予備的な研究として、村井・二村(1999)は、実際に採用選考で用いられた作文を対象として、2つの評定者群を設けて両群の評定の信頼性や評定された人物特性を模索する研究を本研究に先立って行った(以下、予備研究と呼ぶ)。一方の評定者群は言語能力の評価を職務とする4名であり、「総合評価」「文章力」「内容の魅力」「自主性・挑戦性」をはじめとする7つの評定の観点を設定し、各観点ごとの評価基準にもとづき5段階で評定させた。事前に評定訓練を通して評価基準のすり合わせを行った上で、全員に同じ100編の作文を評定させた(以下、専門群訓練後評定とする)。他の評定者群は、作文評定の経験がほとんどない企業の管理職者8名で、評定訓練は行わずに「次の選考段階へ上げたいかどうか」を4段階で評定させた(以下、非専門群採否評定とする)。

その結果、第1に、作文評価の信頼性については、専門群訓練後評定の「総合評価」において4名の評定者間の相関で中央値.59と一定の水準が確保された。一方、非専門群採否評定における評定者間の相関は級内相関係

数で.20を下回る低い水準になった。このことから、事前に評価基準を定めない場合、評定者間の相関によって推定される信頼性はきわめて低いことが確認され、安易な作文評定に警鐘を鳴らす結果となった。しかし、同時に、評定能力の高い評定者が評定を行なうこと、事前に評価基準を設定し、評定訓練などを実施して評定者間でそれを共有することや、複数名による評定を行うことでその信頼性を高められる可能性も示唆された。これは、大学入試での石井の小論文評価に関する研究において、小論文入試における信頼性・妥当性・公平性を確保するために、評定者への評定に先立っての解説の適切さとその徹底、評定者の小論文入試に対する考え方と評定経験の蓄積が重要であるとの指摘と同様の結果である。

以上のように、作文評価を実際の採用選考に適用するためには評定者の評定能力や評定経験などの評定者の特性に左右されずに信頼性を確保することができる評価基準が求められている。

第2に評定内容というテーマは、各評定者が作文を通してどのような人物特性の個人差を評価しているのかというものである。予備研究の専門群訓練後評定における「総合評価」(評定者4名の平均値)がどのような人物特性の評価に重点が置かれているかを検討するために、総合検査SPIの能力検査の2尺度(言語能力・非言語能力)と性格検査の17尺度との相関分析を行った。性格検査は17尺度をグループ主軸法で因子分析を行い5大性格因子(Big5)の内容にあわせて構成した5つの合成尺度(外向性・情緒不安定性・親和性・誠実性・開放性)を用いた。その結果、「言語能力」「非言語能力」との相関はそれぞれ.25,.20であったが、性格の5因子とはそれぞれ.17,-.17,.29,.31,.09が見られた。この結果から、評定者が主に文章表現力や文章構成力などの言語的能力を評価することを仮説として設定した「総合評価」であったが、むしろ「誠実性」など性格特性に焦点を当てるほうが、有効であることが示唆されたわけである。採用選考において企業人能力として評価すべき性格要素でかつ作文の素材によって評価しやすい個人差を検討すると、達成動機を有力な側面と考えることができる。

達成動機とは、何らかの価値を見出した目標を成し遂げようとする動機のことであり、心理学においても1950年代より人間の行動を考える上での中核的概念の一つとされ、その測定法についても数多く研究が行われてきた。McClelland, Atkinson, Clark & Lowell (1953)では、課題統覚テスト(TAT)による測定法が考案されている。それは、TAT図版についての想像的物語を作らせ、その物語の内容について詳細に定められた採点基準に従って採点をする方法である。これは、自分を自由

に表現した作文を通して達成動機を測定できる可能性を示唆している。また、達成動機に関わる性格特性が職務遂行能力と関連が深いことは、質問紙法による性格検査の妥当性研究によても指摘されている(Hough 1992, 二村・今城・内藤2000)。以上から、達成動機は作文のように自由に自己表現された素材によって評定が可能で、またその個人差は企業人能力の重要な要素である可能性があると考えることができる。

採用選考にとって、信頼性を確保しやすくかつ有効な個人差に焦点をあてる作文評定法が求められるが、それは言語的能力ではなく達成動機に焦点をあてることにより、有効な個人差尺度を得ることができる可能性があると考えられる。

2. 目的

本研究では達成動機に関連する側面の評価に焦点をあて、信頼性の高い作文評価基準を模索し、その評定される個人差の内容について検討する。

第1に信頼性の高い評価基準というテーマについては、達成動機に関連する個人差の評価に焦点をあて、エントリーシートに代表される作文評価において、作文の評定能力や評定経験といった評定者の特性にあまり左右されず一定の水準の信頼性を確保できる評価基準の模索を試みる。

第2に評定内容の確認というテーマについては、評定結果と予備研究の専門群訓練後評定との関係を分析すると同時に、予備研究と同様に性格検査との関係を分析し、作文を通して達成動機に関連する側面の評価の可能性を考察する。

3. 方法

作文と採点者

本研究では、サービス業A社の1999年度大学新卒者の採用選考において会社説明会時に課された作文からランダムに抽出された100編の作文を用いるが、これらは予備研究において用いられたものと同じである。作文のタイトルは「学生時代に力を入れたこと」で、20分間という時間制限の下にA4サイズの用紙1枚に書かれたものである。字数制限はないが、平均約600字であった。なお、現実の選考では作文評定の結果は選考の資料としては用いられず、面接の参考資料としてのみ適用されている。

2つの採点者群を設け、その1つは、採点経験のほとんどない、一般事務を職務とする者(非専門群と呼ぶ)3名である(本研究では実験者側で詳細に評価基準を設定したため、本研究について記述する場合は「評定」を

「採点」、「評定者」を「採点者」とする)。

もう一方の採点者群は、長年言語能力問題の作成に携わっているスペシャリスト(専門群と呼ぶ)4名である。本研究の結果と予備研究の専門群訓練後評定の結果との比較を行なう上で、専門群として作文の評定能力や評定経験についての条件をそろえる必要があったため、この4名は予備研究の専門群訓練後評定の評定者と全く同じメンバーで構成した。

両群ともに各採点者に100編すべてを採点させた。なお、両群ともに、文脈効果を相殺するために採点者ごとに作文の呈示順序を変えた。

評価基準の設定

前述のように、予備研究で「学生時代に力を入れたこと」というタイトルで書かせた作文を通して達成動機と関連の深い「誠実性」を評価できる可能性が示唆されていること、また McClelland, et al. においても作文を通して達成動機の評価の可能性が示されていることなどから、本研究では採点の観点を達成動機のみに絞ることとした。

次に、TAT図版についての想像的物語から、達成動機を測定するための採点基準が明確に定められている McClelland, et al. を参考に、「学生時代に力を入れたこと」というタイトルで作文を記述するプロセスにおいて、達成動機がどのような表現の個人差として表われるかを検討した。そして、次のように採点における着眼点と採点基準を設定した。なお、それぞれの着眼点のスコアを足し合わせた「合計スコア」を総合の評定値として扱うこととした。

採点における着眼点として、「目標の有無」「目標の質」「活動量」「目標達成状況」「学習量」の5つを設定した。そして、上記5つの着眼点についてのスコアの与え方に採点者の主觀がなるべく反映されないように、採点基準を次のように設定した。

「目標の有無」は、対象作文の中に自ら主体的に設定した目標の有無を判断させた。複数種類の目標が記述されていることもあったが、1つでも目標がある場合については、1点とした。目標が記述されていない場合は0点とした。

「目標の質」は、「目標の有無」において目標記述がある場合、その目標がどの程度の範囲の人たちとの関わりの中で達成できるかという軸で次のような3段階で判断させた。インターンシップなど学生生活から一步踏み出し社会との接点を多く持つ目標と判断される場合は3点、サークルやアルバイトなど通常の学生生活におけるコミュニティ範囲内で実現できる目標と判断される場合は2点、そして、資格取得など人との関わりが少なくほ

ぼ独力で達成できる目標と判断される場合は1点とした。複数の目標記述がある場合は、その中でもっともスコアの高いものを対象とした。

「活動量」については、目標達成に向けてとった方法や手段の種類の記述を数えて、その数をスコアとした。複数の目標がある場合は、それぞれの目標を達成するためにとった方法や手段の種類の合計数をスコアとした。なお、「目標の有無」において目標記述がないと判断された場合は、学生時代に取り組んできた活動の種類を数えて、その数をスコアとした。

「目標達成状況」については、最終的にその目標を達成したかどうかに関する記述を評価させた。「目標の有無」において目標記述があると判断され、かつ、それを達成したという記述が見られた場合は2点、「目標の有無」において目標記述はあったものの、達成したという記述がない場合は1点、そして、「目標の有無」において目標記述がないと判断したものについては0点とした。なお、「目標の有無」において目標記述が複数あった場合は、そのうちのどれか1つでも達成していれば2点とした。そして、複数あった目標記述のうち、いずれも達成していない場合は1点とした。

最後の「学習量」については、学生時代に力を入れたことを通して結果として学んだことに関する記述の種類を数えて、その数をスコアとした。また、一連の活動やそこから学んだことを踏まえて、今後の目標について具体的に記述されている場合もスコアに含めた。

採点に先立って、両群ともに約2時間をかけて採点者訓練を行った。まず、ある1編を例として提示し、採点における5つの着眼点ごとに採点基準を説明した。その後、2編について、実際に採点をさせ、実験者側で用意した採点模範回答のフィードバックをしながら、採点基準の確認を行った。予備研究の専門群訓練後評定における評定訓練では約3時間をかけて討論を通じた評価基準のすり合わせを実施したのに対し、本研究での採点訓練は訓練ではなく、実験者側で設定した評価基準の理解を目的としたものであった。

4. 結果

4. 1. 信頼性係数の算出

5つの着眼点の信頼性

各群内での全採点者のスコアの整合性を見るために各群別に信頼性係数(Cronbachの α 係数)を求めた(表1)。非専門群と専門群では採点者数が異なるため、専門群(4名)の α 係数は非専門群(3名)の α 係数と比較できるように3名の場合に相当する値に補正してある。表1に非専門群、専門群それぞれにおける α 係数が

表1 採点の信頼性係数

(非専門群3名と専門群4名) 対象作文100編, α 係数

	非専門群	専門群 (補正值)
目標の有無	.60	.80
目標の質	.53	.79
活動量	.73	.76
目標達成状況	.63	.81
学習量	.82	.81

示されている。

非専門群は、「活動量」と「学習量」において α 係数でそれぞれ .73 と .82 、専門群の .76 と .81 という結果と同等で、かなり高い整合性が得られた。

しかし、「目標の有無」「目標の質」「目標達成状況」の3つの着眼点においては、非専門群は α 係数が .53 から .63 と低い水準となったが、専門群では α 係数が .79 から .81 とやや高い水準となり、両群間で差が見られた。

着眼点の間の相関

達成動機の水準を表す「合計スコア」を算出するのに先立って、5つの着眼点の関係を確認するために、非専門群、専門群それぞれにおいて内部相関を求めた(表2)。これは各群別に、5つの着眼点それぞれについて採点者のスコアの平均値を算出し、着眼点の間の相関を求めるものである。

目標記述について採点をする「目標の有無」と「目標の質」の間では、両群とも .92 と .94 ときわめて高い相関

となった。また、「目標の有無」と「目標達成状況」との相関も .96 と .98 ときわめて高い結果となった。これは、「目標の有無」において目標があるとされたものの中で「目標達成状況」において達成していないとされる作文が実際はほとんどなかったことによるものである。その影響で「目標の質」と「目標達成状況」についても .92 と .93 ときわめて高い相関となったものと考えられる。

「学習量」については、非専門群と専門群の間で、「目標の有無」との相関がそれぞれ .27 と -.05 「目標の質」とは .31 と .02 、そして「目標達成状況」とは .34 と .01 と、異なる傾向が見られた。全体的に見て、「学習量」をのぞいては両群でほぼ同様の傾向が確認された。

「合計スコア」について

前述の着眼点の分析で、「目標の質」と「目標達成状況」は非専門群において信頼性がやや低かったこと、および両群において「目標の有無」との相関がきわめて高かったことから、「目標の有無」「活動量」「学習量」の3つの着眼点のスコアを加えて「合計スコア」を算出することとした。「合計スコア」についても、5つの下位の着眼点と同様に採点者間のスコアの α 係数を求めた(表3)。この場合にも専門群の α 係数については先と同

表3 合計スコアの信頼性係数

(非専門群3名と専門群4名) 対象作文100編, α 係数

	非専門群	専門群 (補正值)
合計スコア	.78	.82

表2 着眼点の間の相関係数(採点者のスコアの平均による)

(非専門群3名と専門群4名) 対象作文100編

非専門群					
	目標の有無	目標の質	活動量	目標達成状況	学習量
目標の有無	1.00				
目標の質	.92	1.00			
活動量	.48	.49	1.00		
目標達成状況	.96	.92	.51	1.00	
学習量	.27	.31	.04	.34	1.00

専門群

	目標の有無	目標の質	活動量	目標達成状況	学習量
目標の有無	1.00				
目標の質	.94	1.00			
活動量	.46	.52	1.00		
目標達成状況	.98	.93	.51	1.00	
学習量	-.05	.02	.00	.01	1.00

様の補正を加えた。

非専門群は α 係数.78、専門群についても α 係数（補正值）.82と、両群同等で、やや高い水準の結果が得られた。

4. 2. 採点内容の確認

まず得られたスコアと予備研究における専門群訓練後評定との関係を分析するために両者の相関分析を行い、次に「合計スコア」と総合検査SPIとの相関係数を求めた。

専門群訓練後評定との相関分析

「目標の有無」「活動量」「学習量」「合計スコア」と、予備研究の専門群訓練後評定における「総合評価」との相関係数を求めた（表4）。

非専門群では、「総合評価」と合計スコアとの間に.40とある程度の相関が見られ、また、下位の着眼点である「目標の有無」「活動量」「学習量」とは.21から.37の相関係数を示した。

また専門群では、「合計スコア」では.58とやや高い相関が見られ、また、下位の着眼点では.34から.42の相関係数が得られた。

総合検査 SPI との相関分析

作文評価の「合計スコア」と総合検査SPIで測られる

得点との相関係数を求めた（表5）。表5に示された相関係数の平均は、両群ごとに、全採点者の合計スコアの平均値を用いた。なお、総合検査SPIの性格検査の17尺度は、5大性格因子（Big 5）の内容に合わせて5尺度にまとめた。

非専門群は、「情緒不安定性」において-.24から-.06、平均で-.15、達成動機と関連の深い「誠実性」においては.03から.28、平均で.14という値が示されたが、統計的に有意ではない。一方、「言語能力」とは-.04から.05、平均で-.01と相関は見られなかった。

専門群は、「誠実性」において.21から.31、平均で.29と、7尺度の中でもっとも高い相関が見られた。また「親和性」において.13から.23、平均で.21、「情緒不安定性」においても-.26から-.11、平均で-.20と低い相関が見られた。「言語能力」とは非専門群と同様に、-.09から.14、平均で.01と相関は見られなかった。

5. 考察

5. 1. 信頼性の検討

作文評価を採用選考に適用するためには、評定の信頼性として、同一評定者による複数回評定間における安定性、複数の評定者間の評定の相関、および下位評定項目間の内的整合性の3つの方法によって求められる。

表4 専門群訓練後評定とスコアとの相関係数

（非専門群3名と専門群4名）対象作文100編

	非専門群				専門群			
	目標の有無	活動量	学習量	合計スコア	目標の有無	活動量	学習量	合計スコア
専門群訓練後評定の総合評価	.37	.33	.21	.40	.35	.42	.34	.58

表5 性格・能力検査と合計スコアとの相関係数

（非専門群3名と専門群4名）対象作文100編

採点者	非専門群				専門群			
	A	B	C	平均	D	E	F	G
外向性	.15	.04	.05	.09	.20	.11	.09	.14
情緒不安定性	-.24	-.06	-.09	-.05	-.14	-.17	-.11	-.26
親和性	.16	-.01	.04	.07	.13	.19	.15	.23
誠実性	.28	.03	.06	.14	.26	.21	.21	.31
開放性	.03	.11	.00	.06	.18	.03	.00	.10
言語能力	-.04	-.04	.05	-.01	-.09	.14	.05	-.05
非言語能力	.03	-.07	-.06	-.04	-.06	-.01	-.07	.00

※総合検査SPIの性格検査の17尺度はグループ主軸法により次のような5つの因子にまとめた（村井・二村、1999）。

<外向性>社会的内向性、内向・外向<情緒不安定性>敏感性、自責性、気分性（得点が低いほど安定している）<親和性>持続性、感情・思考<誠実性>達成意欲、自信性、知覚・判断<開放性>内省性、直観・感覚

非専門群は、5つの着眼点単位では α 係数が.53から.82と、低い水準のものから高い水準のものまでばらつきがあり、「合計スコア」では α 係数.78とやや高い水準の信頼性となった。専門群は、着眼点単位では α 係数(補正値)が.76から.81、「合計スコア」でも α 係数(補正値)が.82と、いずれもある程度の水準の信頼性を示した。このように着眼点単位で見ると非専門群と専門群の間で信頼性に差がある結果となったが、「合計スコア」ではほぼ同等で、両群ともに相応の水準の信頼性が確保された。両群の「合計スコア」の信頼性は、採点者訓練の程度や評定内容・手法などが簡便化されたにも関わらず、予備研究の専門群訓練後評定における「総合評価」の評定者間の相関の中央値.59ともほぼ同等の水準である。以上はある1つの観点を測定するために複数の下位の着眼点を設けてそれらの合計をスコアとする工夫が必要であることを示唆すると考えられる。

つぎに着眼点の構造から考察を試みる。5つの着眼点の関係を確認するために非専門群と専門群において着眼点の間の相関を求めたところ、「学習量」について、「目標の有無」とはそれぞれ.27と-.05、「目標の質」とは.31と.02、また「目標達成状況」とは.34と.01と、両群間で異なる傾向が見られ両群における微妙な差異がうかがわれた。しかし、他はほぼ同様の相関となり、着眼点は両群間で全体として異ならない構造となっているものと考えられる。

以上の結果から、事前に採点者訓練を施し評価基準を明確に設定したこと、つまり達成動機に絞り採点における着眼点とその採点基準を詳細に設けたことで、評定経験のほとんどない採点者においても評定能力の高い専門の採点者とほぼ同様の採点作業が行われたものと思われる。

なお、本研究で設定した着眼点には、「活動量」と「学習量」といった達成動機に関連する記述の量を数えさせるものや、「目標の質」のように記述内容のレベルを判断させるものが含まれていた。前者は、非専門群においてそれぞれ α 係数.73と.82、専門群でもそれぞれ α 係数(補正値).76と.81と両群同等で、ある程度の水準の信頼性となった。一方、後者は、専門群で α 係数(補正値).79とやや高い水準が確保されたものの、非専門群では α 係数.53と低い水準となつたことから、記述内容のレベルを判断させる着眼点は、採点基準を明確に設定することが難しく、信頼性を確保する上では、予備研究の専門群訓練後評定を行ったように、評定能力の高い採点者によって、ディスカッション形式などを通して、多くの例を用いてそれら一つづつについて評価レベルを討議していくような採点基準のすりあわせが必要である

と考えられる。

しかし、記述の量を数えさせる「学習量」について着眼点の間の相関を見ると、非専門群と専門群の間で前述のように微妙な差異がうかがわれた。専門群の採点者の採点後の感想によると、学んだことが抽象的に書かれている場合、学習の内容を主観的に推し量っての評価を加えたことが表明された。一方の非専門群は主観的判断をせず、文字通りの採点になっていたものと思われる。このことから専門群は非専門群よりも「学習量」において採点基準にばらつきが生じていたものと考えられる。

5. 2. 採点内容の確認

本研究で得られたスコアが仮説通りに達成動機に関連する側面の個人差を意味しているかどうかを検討するために、まず本研究で得られた各着眼点のスコア、合計スコアと、予備研究の専門群訓練後評定における「総合評価」との相関分析を行い、次に総合検査SPIとの相関分析を行なった。

専門群訓練後評定の「総合評価」との相関分析では、「合計スコア」において非専門群で.40、専門群で.58という両者の間のある程度の相関関係が確認された。

「合計スコア」と総合検査SPIとの相関分析結果によると、非専門群は「情緒不安定性」において-.24から-.06、平均で-.15、達成動機と関連の深い「誠実性」においては.03から.28、平均で.14の値が見られた。専門群は「誠実性」と.21から.31、平均で.29と、合成7尺度の中で最も高い相関係数が得られた。なお「親和性」で.13から.23、平均で.21、「情緒不安定性」とも-.26から-.11、平均で-.20と有意な低い相関が見られた。予備研究において言語能力に関する評価を意図したにもかかわらず性格的要素の個人差とも相関が見られたが、ここでは企業人能力と関連の深い達成動機の抽出を意図し、わずかながらその可能性が示唆されたといえる。さらに情緒不安定性においても他の尺度より高い相関が両群で得られたが、それは企業人能力と無関係ではない要素と考えができるもので、達成動機の個人差のみが抽出されたというよりは、達成動機あるいは企業人能力に関わる情動的側面の個人差が評価され得る可能性が示唆されたものと考えることができる。

一方、「言語能力」との相関は、非専門群では-.04から.05、平均で-.01、専門群は-.09から.14、平均で.01と相関は見られなかった。予備研究の専門群訓練後評定における「総合評価」では「言語能力」と平均で.25、「非言語能力」と.20という相関が見られたが、本研究ではほとんど相関が見られなかった。予備研究においては能力的要素を評価することが、また本研究では達成動機

を評価することが意図されていたことから当然の結果ではあろう。しかし、本来作文には言語的能力の個人差があらわれやすいものと考えられるにもかかわらず達成動機など情動的な個人差が抽出された可能性が示唆されたことに特別の意味が見出せる。すなわち達成動機など、欲求や情動的側面に関わる個人差は、企業人能力との関連性が議論されながらその測定手法は必ずしも具体化され得ていない。作文がこれらの個人差を測定する手法になりうるか否かの判断は今後の研究を待たなければならぬが、本研究によってその可能性が示唆されたものと考へることができよう。

6. 今後の課題

作文評価に関する今後の研究課題には、次のようなことが挙げられる。本研究では、作文評価の内容を検討するために総合検査SPIの性格尺度を5大性格因子（Big 5）にまとめ、そのなかの達成動機に近似する概念の「誠実性」を比較対象としたが、達成動機そのものを測定する確立された尺度との相関分析を行い、その測定可能性を検証する必要があろう。作文の評定値と入社後の職務遂行能力との相関分析、すなわち予測的妥当性の検証も必要であろう。これにより作文から抽出される個人差の情報の含意が明確になると同時に、企業の採用選考などへの適用の有効性を検討することができる。

また、研究法の観点からもいくつかの課題がある。予備研究と本研究ではともに同じ作文サンプルを対象に研究を行なってきたが、同じタイトルで書かれた別の作文標本を対象にした場合も、信頼性や採点内容について同様の結果が得られるかについて確認する必要がある。また、別のタイトルで書かれた作文を通して達成動機に関連する人物特性の評価を行う場合、どのような着眼点と採点基準を設定すればよいか、例えば「協調性」や「独創性」といった他の個人差の側面については、どのようなタイトルを選定し、どのような着眼点や採点基準を設定すれば測定が可能かが研究テーマとなる。

また、本研究では「目標の質」という着眼点を設けて

3段階で採点をさせたが、非専門群では信頼性が低くなり、記述内容の質の評価には評定能力の高さや採点基準のすりあわせが求められることが示された。一方で、本研究では目標についてのみ質に関する採点を行ったが、他の活動や学習についても質についての評価を試行し、質の評価によってどのような個人差情報が加わるのか、その含意を確認する必要があると思われる。

参考文献

- Hough, L.M. 1992 The "big five" personality variables – Construct Confusion: Description versus prediction. *Human Performance*, 5 (1 & 2), 139-155.
- 石井巌 1981 「論文試験とその評価」について. 行動計量学 8 (1), 22-29.
- Burnstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Braden-Harder, L., and Harris, M.D. 1998 Automated scoring using a hybrid feature identification technique. *Proceedings of the Annual Meeting of the Association of Computational Linguistics*, August, 1998. Montreal, Canada.
- McClelland, D.C., Atkinson, J.W., Clark, R.A., & Lowell, E.L. 1953 *The achievement motive*. New York: Appleton-Century-Crofts.
- 村井智恵子, 二村英幸 1999 採用選考における作文評価, 経営行動科学学会第2回年次大会発表論文集, 56-62.
- 二村英幸, 今城志保, 内藤淳 2000 管理者層における性格検査、知的能力検査の妥当性に関するメタ分析的研究—従業員規模、業種、職務内容別のメタ分析—, 経営行動科学学会第3回年次大会発表論文集, 55-64.
- (平成13年2月28日受稿, 平成13年3月30日受理)