

段階表示を用いた合否判定方法の試行

林 篤裕 (名古屋工業大学)

大学入試の合否判定における 1 点の重みの意味合いは、特にボーダーライン近傍で予てより議論があった。1 点かそれ未満の差異によって合否が分かれるため、その微小な違いに疑念を持たれてきた。この問題は 2014 年 12 月に発表された中教審答申でも取り上げられ試験成績の段階表示が提案され、これを受けて大学入学共通テストでは従来の素点に加えて段階表示の成績も提供されるようになった。そこで、共通試験と個別試験を合わせて合否を判定する形態を採っている入試(主には国立大学)において、評価資料として素点に代えて段階表示を導入する際の利活用の方法を模索することにした。素点と段階表示の得点の散らばり具合に注目し、その標準偏差の違いが合否判定に及ぼす影響を検討した。

キーワード：得点分布, Stanine, 一般選抜, 配点比率

1 はじめに

2014 年 12 月に公表された中教審答申「新しい時代にふさわしい高大接続の実現に向けた高等学校教育、大学教育、大学入学者選抜の一体的改革について～すべての若者が夢や目標を芽吹かせ、未来に花開かせるために～」では、高校教育と大学教育、そして、両者を接続する大学入試の三位一体の改革が提言された。これを受けて、大学入試では「学力の三要素」を測定して選抜することが求められ、共通第 1 次学力試験、大学入試センター試験と続いてきた共通試験も大学入学共通テストと名称を変えると共に、より思考力・判断力・表現力を問う試験に変わることを目指すことになった。

構想段階では年複数回実施や合教科・科目型試験、総合型試験、また CBT 方式の採用といった方策が盛り込まれていたものの、議論の結果これらについては早々に見送られた。残された方策のうち、ポートフォリオを用いた主体性評価については運営団体が 2019 年 8 月に運用を停止することになり、また、二本柱として注目されてきた「英語民間試験を用いた 4 技能評価」と「国語と数学の記述式問題(各 3 題を想定)」についてもその実施面・評価面での公平性の観点から疑念を払拭することができず、結局実施約 1 年前に迫った 2019 年 11 月と 12 月になって相次いで見送りが決定された。残ったのは、今まで以上に思考力を問う工夫を凝らしたマークシート方式の作題と Stanine と呼ばれる試験成績の段階表示であった。初回となる 2021 年 1 月に実施された大学入学共通テストは予期せぬ事態として COVID-19 感染対応も加わり、本試験が 2 回、追試験が 1 回という過去に例を見ない変則的な実施日程となったものの、幸い関係者のご尽力もありほぼ計画通り無事に実現された。

これまで試験成績には 0 点から満点までの 1 点刻みの素点を用いられてきたが、ボーダーライン近傍では 1 点(場合によってはそれ未満)の差異が合否を分けることも珍しくなく、1 点の意味合いに疑問の呈する考えもあり、このような「1 点刻みの客観性」にとらわれた評価から脱して、多様な評価方法の導入を促す目的から段階表示による成績提供が提言された。上述の中教審答申を受けて、2019 年 6 月に大学入試センターから公表された「令和 3 年度大学入学者選抜に係る大学入学共通テスト問題作成方針(2020 年 1 月に一部変更あり)」には「大学への成績提供等(第 6 節)」として「科目ごとの 9 段階の段階表示」の提供が明記されている。これまで延べで 42 年間にわたって実施されてきた共通試験としての共通第 1 次学力試験と大学入試センター試験では各科目の成績データとしては素点だけであったが、大学入学共通テストからは素点に加えて Stanine と呼ばれる 9 段階の段階表示による段階点が合わせて提供されることになった。

しかし、現状で段階表示を用いた合否判定方法が確立しているとは言えず、各大学に利用の工夫が求められている状況である。事実、2021 年度入試において成績の段階表示を用いた選抜を行うと表明した大学はなかった。そこで、本研究では段階表示を用いた合否判定方法を幾つか提案・試行し、今後の利活用の参考に供してもらおうと考えた。

2 段階表示としての Stanine

Stanine とは、受験者群を点数の低い方から順に、4, 7, 12, 17, 20, 17, 12, 7, 4%の 9 つの左右対称な領域に分割する手法(United States Army Air Forces, 1947; 前川, 2019)で、この一見不規則とも思えるこ

れら割合の元となっているのは、得点分布が平均 μ 、標準偏差 σ の正規分布であった場合に、平均を中心とする $\mu \pm 1.75\sigma$ 、 $\mu \pm 1.25\sigma$ 、 $\mu \pm 0.75\sigma$ 、 $\mu \pm 0.25\sigma$ の 8 箇所まで分布を区切った際の割合にあたる。記憶媒体が高価で貴重な時代に成績を一桁の数字による段階値で表現することを目的に開発された手法であり、得点分布が連続分布であれば上記の割合に分割することができるが、共通試験の得点は整数値を取るため離散分布となり、実際の得点分布データに適用した場合には上記の数値から多少前後する割合で分割されることになる。Stanine で表現された成績は 1 番(低評価)から 9 番(高評価)の階級を示す段階値で通知され、階級値が同じ受験者を同じ評価で扱うことが想定されている。

3 令和 3 年度大学入学共通テストの例

2021 年 1 月に実施された初回の大学入学共通テストからは受験科目の得点(いわゆる素点)に付帯して段階表示として Stanine の階級値も提供されるようになった。具体的には大学入試センターの成績提供データの中の「個人別成績データ」に、各受験者の科目毎

の素点と段階値がセットで格納されている。また、もう一つの提供データである「全受験者成績分布データ」は、(1)平均点等レコード、(2)得点別人数レコード、(3)段階表示レコードの 3 つのパートから構成されており、3 番目のパートに段階表示の境界点数とその階級に含まれる受験者数が格納されている。これまでこの(1)と(2)は提供されていたが、今回新たに(3)が加わったことになる。なお、この境界点数は大学入試センターのホームページにも「令和 3 年度大学入学共通テスト 段階表示換算表」として掲載されている。

試しに、2021 年度(令和 3 年度)の成績データの中から、国語、数学 I・数学 A、英語(リーディング)について見てみる(表 1)と、各階級の下限点数と上限点数に加えて受験者数が判るが、含まれている受験者数の割合は想定割合とは若干ズレが有ることが判る。これは、丁度境界に当たる点数(整数値)に多くの受験者が含まれている場合に、同点者を途中で分割することが出来ないことから、割合にデコボコができるためである。

なお、あまり知られていないようだが、「全受験者

表 1. 段階表示換算表と各階級に含まれる受験者数・割合

段階表示 (Stanine)	国語				数学 I・数学 A				英語(リーディング)				
	想定%	下限	上限	人数	下限	上限	人数	%	下限	上限	人数	%	
1	4.0%	0	53	17860	3.9%	0	22	12523	3.5%	0	23	18978	4.0%
2	7.0%	54	69	30019	6.6%	23	32	25290	7.1%	24	30	31531	6.6%
3	12.0%	70	88	54229	11.9%	33	41	39371	11.0%	31	39	55027	11.6%
4	17.0%	89	109	77226	16.9%	42	51	60633	17.0%	40	51	82586	17.3%
5	20.0%	110	129	91137	19.9%	52	62	73187	20.5%	52	64	91939	19.3%
6	17.0%	130	145	77469	16.9%	63	72	59034	16.6%	65	77	85962	18.1%
7	12.0%	146	159	58451	12.8%	73	82	45049	12.6%	78	87	56914	12.0%
8	7.0%	160	170	30831	6.7%	83	91	27012	7.6%	88	94	34016	7.1%
9	4.0%	171	200	20083	4.4%	92	100	14394	4.0%	95	100	19221	4.0%

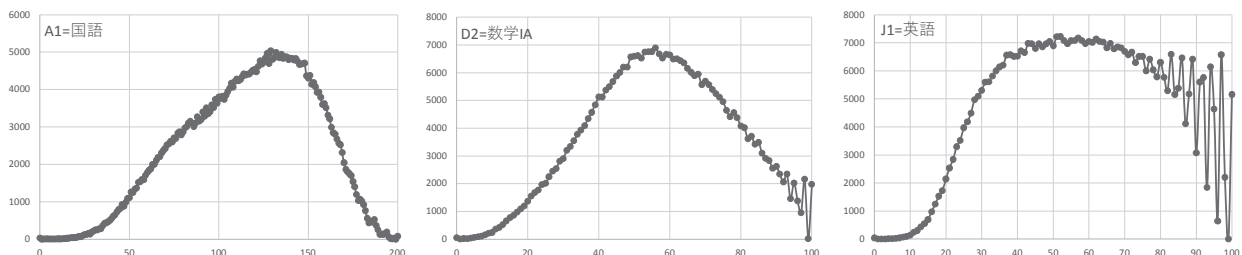


図 1. 各科目の得点分布(左から順に国語、数学 I・数学 A、英語(リーディング))

成績分データ」の 2 番目のパートである得点別人数レコードには 1 点刻みの各得点の受験者数が格納されており、これを用いると各科目ごとの得点分布を知ることができる。試しに上で取り上げた 3 科目の得点分布を図 1 に示す。このデータを用いることにより 9 段階に限定されることなく、利用者が指定する任意の数の段階表示や得点分布の特性に応じた階級の設定も生成可能になる。個別大学の特性に合わせた利用が可能な素材とも言えるので、多方面からの検討を期待している。

4 N 大学における階級分布

前節では大学入学共通テストの全受験者の階級分布を紹介したが、個別大学においては、選抜を行うために大学入試センターから提供される志願者の「個人別成績データ」を用いることによって、当該大学の階級分布を把握することができる。ここでは、秘匿の関係から 9 つの階級をそのまま表示することはせず、階級を 3 つずつにまとめてその受験者割合を表 2 に示す。なお、このデータは N 大学の前期日程または後期日程を志願した、合わせて 4001 名のものであり、英語(リーディング)を受験していない者、および、英語(リスニング)を免除された者をここでは除外してある。

これを見ると、N 大学の志願者は全受験者の分布より高得点側に偏った分布をしていることが判るが、一方で、科目ごとにその偏り度合いが異なることも判る。

表 2. N 大学における各階級ごとの受験者割合

段階表示	想定%	国語			数学Ⅰ・数学A			英語(リーディング)		
		下限	上限	%	下限	上限	%	下限	上限	%
1, 2, 3	23.0%	0	88	3.0%	0	41	2.4%	0	39	3.5%
4, 5, 6	54.0%	89	145	59.9%	42	72	34.9%	40	77	58.7%
7, 8, 9	23.0%	146	200	37.1%	73	100	62.7%	78	100	37.8%

5 段階表示を用いた合否判定

N 大学の一般選抜試験では大学入学共通テストと個別学力試験の合計点を用いて合否判定を行っている。ここではデータの整備の都合で、2021 年度の大学入学共通テストではなく、同様の方法で合否判定を行っていた平成 X 年の大学入試センター試験を例に、当該年の「全受験者成績分布データ」から Stanine を算出し、これを用いて合否判定を行った場合の傾向を捉えることにした。

まず、Stanine の階級を定めるために、各科目の得点分布から想定割合に基づいて 9 群に分割する

表 3. 平成 X 年大学入試センター試験における Stanine の各階級に関する値

段階表示 (Stanine)	国語			数学Ⅰ・数学A			英語(リーディング)			
	想定%	下限	上限	%	下限	上限	%	下限	上限	%
1	4.0%	0	55	3.7%	0	22	3.7%	0	43	3.7%
2	7.0%	56	71	7.1%	23	33	6.9%	44	60	6.9%
3	12.0%	72	89	11.6%	34	43	11.3%	61	85	12.3%
4	17.0%	90	112	17.3%	44	54	18.0%	86	113	16.8%
5	20.0%	113	134	19.4%	55	65	19.9%	114	141	20.3%
6	17.0%	135	152	17.0%	66	75	16.3%	142	161	16.3%
7	12.0%	153	166	12.3%	76	85	12.5%	162	176	12.5%
8	7.0%	167	177	7.2%	86	92	6.8%	177	186	6.9%
9	4.0%	178	200	4.4%	93	100	4.6%	187	200	4.3%

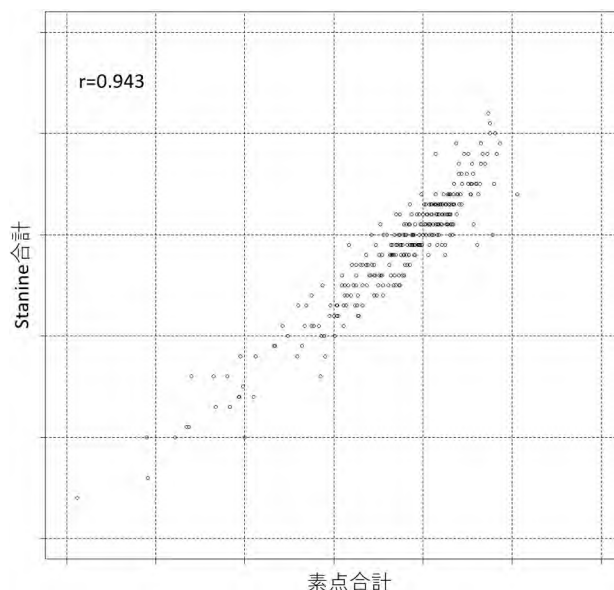


図 2. 素点合計と Stanine 合計の関係

階級の下限值と上限値を確定した。参考までにその中の 3 科目について、関係する数値を表 3 に掲載する。また、N 大学で利用している大学入試センター試験の科目は国語、地歴・公民から 1 科目、数学Ⅰ・数学 A、数学Ⅱ・数学 B、理科 2 科目、外国語(リーディングとリスニング)の 8 コマであり、素点合計(900 点満点)と Stanine 合計(72 点満点)の相関係数は 0.943 であった。このように相関係数が 1 をやや下回っている理由は、同じ大学入試センター試験の成績ではあるものの、段階表示は 9 段階であり、101 段階(国語と英語は 201 段階)の素点よりも粒度が小さいため、その差異から表現情報が落ちているからである。また、両者の関係を示す散布図を図 2 に示すが、Stanine 合計は離散変量であるので、横方向に縞模様状に分布していることが判る。なお、具体的な数値を秘匿するために軸ラベルは省略してある。

表 3 に示した各科目の下限值・上限値に基づい

表 4. N 大学における階級毎の受験者割合

段階表示	想定%	国語			数学 I・数学 A			英語(リーディング)		
		下限	上限	%	下限	上限	%	下限	上限	%
1, 2, 3	23.0%	0	89	2.4%	0	43	1.9%	0	85	2.3%
4, 5, 6	54.0%	90	152	60.4%	44	75	37.1%	86	161	53.1%
7, 8, 9	23.0%	153	200	37.2%	76	100	61.0%	162	200	44.7%

て、N 大学の志願者における大学入試センター試験の各科目の素点を階級値に変換した。上述同様、9つの階級を3つずつにまとめてそれらの受験者割合を表4に示す。表2の時と同様に、N大学の志願者分布は全受験者分布より高得点側に偏っていることが判ると共に、科目毎の偏り方も表2と同様の傾向を示していることから、今回の大学入試センター試験のデータを用いることで大学入学共通テストでの振る舞いを予測するのに問題がないことが判った。

そこで、N大学の入試の特性を把握するために、従来行ってきた大学入試センター試験と個別学力試験を用いた合否判定の状況を確認しておく。N大学には前期日程にも後期日程にも9つの選抜単位があるが、ここでは前期日程の中から比較的志願者

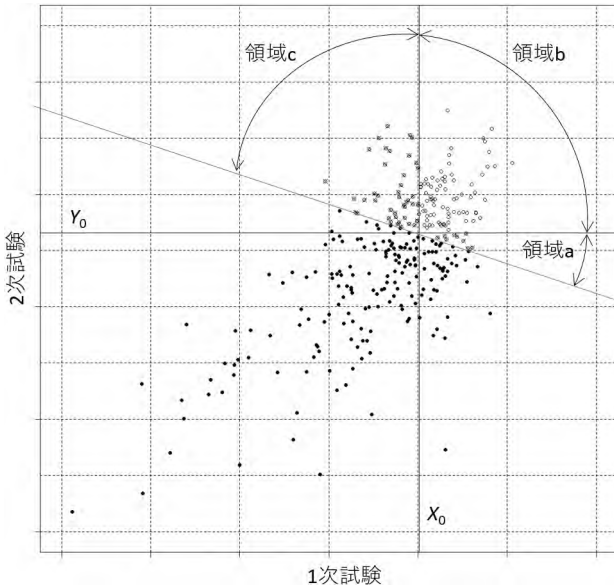


図 3. 選抜単位 A における両試験の得点分布と合格者の特性

表 5. 平成 X 年前期日程の合否入れ替わり率

選抜単位	実質倍率	相関係数	合否入れ替わり率	
			1次試験による	2次試験による
A	2.40	0.675	7.0%	32.6%
B	2.68	0.741	10.0%	33.6%
C	3.42	0.728	8.0%	43.2%

の多い 3 つの選抜単位についてそれらの特性を見てみる。なお、前期日程では大学入試センター試験の得点を 1/2(450 点満点)に圧縮し、個別学力試験の得点(1000 点満点)との合計点(1450 点満点)で合格者を決定している。

紙面の関係からここでは選抜単位 A における、両試験の得点分布と合格者の特性を示した散布図を図 3 に、また、合否入れ替わり率を表 5 に示す。これらを見ると判るように、N 大学では 2 次試験の比重が大きいため、領域 a(1 次試験の成績の優位さを武器に合格した群)の割合が小さく、そのことは合否入れ替わり率からも読み取れる。なお、具体的な数値を秘匿するために図 2 と同様に図 3 の軸ラベルも省略してある。

ここで取り上げた 3 つの選抜単位のそれぞれについて素点と Stanine の階級値の特性を表 6, 7, 8 の左部分に示す。ここでは具体的な数値を明確にしないようにするために、最高点, 最低点は除外し、代わりに統計指標の「範囲」を示してある。また、志願者数や合格者数も示さない。

前期日程では従来は 1 次試験(大学入試センター試験)の素点を 1/2 にし、2 次試験(個別学力試験)の得点との合計点を用いて合否を判定していた。今回、前者の得点に代えて Stanine を用いると、どのように合否の判定が入れ替わるかを観察した。例えば選抜単位 A について、1 次試験の段階表示の範囲は 38.0 点で標準偏差が 6.0 点となっている。これを範囲が 700 点以上で標準偏差が 120 点以上の 2 次試験と単純に合計すると、合否判定には 2 次試験の影響が大きく関わることは容易に想像が付く。そこで、1 次試験の段階表示の合計値を k 倍してから 2 次試験と足し合わせて合否判定を行う方法を考えた。その場合に、従来の素点を用いて合否判定を行った際との合否が入れ替わる発生件数を、合格者数で除して「合否入れ替わり発生率」として示した。

また、k の取り方としては、単純に 2 つの試験を合算しただけの「単純和(k=1)」以外に、1 次試験の素点に代えて Stanine を用いることから、両者の散らばり具合を揃えるように k の値を調整した「等標準偏差」、そして、素点の 100 点満点を 9 段階に変換したことを考慮して「科目毎満点(k=100/9=11.1)」を検討した。なお選抜単位 A を例に「等標準偏差」を説明すると、1 次試験の素点の標準偏差が 71.5 であり、合否判定時には 1/2 に圧縮するため、段階表示の標準偏差を 71.5/2 にす

表 6. 選抜単位 A における各試験得点の特性と合否入れ替わり発生率

選抜=A	1次試験		2次試験		2次のみ		単純和		等標準偏差		科目毎満点	
	素点表示	段階表示	個別学力試験	従来合計点	k=0 合計点	k=1 段階表示	合計点	k=5.95 段階表示	合計点	k=11.10 段階表示	合計点	
満点	900点	72点	1000点	1450点	1000点	72点	1072点	429点	1429点	800点	1800点	
範囲	470.6	38.0	713.0	923.5	713.0	38.0	739.0	226.2	890.2	421.8	1085.8	
平均点	635.4	48.6	491.5	827.6	491.5	48.6	540.0	289.1	780.5	539.1	1030.5	
標準偏差	71.5	6.0	121.4	149.5	121.4	6.0	125.4	35.8	147.4	66.7	172.7	
合否入れ替わり発生率					6.2%	4.7%		3.9%		7.0%		

表 7. 選抜単位 B における各試験得点の特性と合否入れ替わり発生率

選抜=B	1次試験		2次試験		2次のみ		単純和		等標準偏差		科目毎満点	
	素点表示	段階表示	個別学力試験	従来合計点	k=0 合計点	k=1 段階表示	合計点	k=5.44 段階表示	合計点	k=11.10 段階表示	合計点	
満点	900点	72点	1000点	1450点	1000点	72点	1072点	392点	1392点	800点	1800点	
範囲	449.4	46.0	776.0	990.0	776.0	46.0	804.0	251.2	977.5	510.6	1203.0	
平均点	642.1	49.9	493.7	833.0	493.7	49.9	543.6	272.6	766.3	554.1	1047.7	
標準偏差	84.9	7.8	150.5	186.1	150.5	7.8	156.3	42.5	184.2	86.3	222.2	
合否入れ替わり発生率					10.0%	8.2%		5.5%		10.0%		

表 8. 選抜単位 C における各試験得点の特性と合否入れ替わり発生率

選抜=C	1次試験		2次試験		2次のみ		単純和		等標準偏差		科目毎満点	
	素点表示	段階表示	個別学力試験	従来合計点	k=0 合計点	k=1 段階表示	合計点	k=5.82 段階表示	合計点	k=11.10 段階表示	合計点	
満点	900点	72点	1000点	1450点	1000点	72点	1072点	419点	1419点	800点	1800点	
範囲	448.4	45.0	669.0	865.0	669.0	45.0	698.0	260.8	851.3	499.5	1056.1	
平均点	630.8	48.6	451.3	785.0	451.3	48.6	499.9	281.6	732.9	539.3	990.5	
標準偏差	83.8	7.2	137.2	172.2	137.2	7.2	142.6	41.9	170.2	80.3	203.2	
合否入れ替わり発生率					6.8%	6.8%		4.5%		9.1%		

るためには、 $k=71.5/2/6.0=5.95$ とすれば良いことが判る(表中では関係する値を斜体で示してある)。なお、 $k=0$ とすると 1 次試験の影響を全く除去することができ、2 次試験の成績だけで判定することに相当するので、それも参考に掲載した。それぞれ k の値を変更した際の合否入れ替わり発生率を含めた数値を表 6, 7, 8 の右部分に示す。

6 考察

N 大学の場合、1 次試験の Stanine の階級値を拡大させる倍数 k を変化させても合否入れ替わり発生率はそれほど大きくはなく、かつ、変化も乏しかった。これは N 大学では元々 1 次試験の比重が小さく、表 5 から判る通り、従来の判定方法でも 1 次試験による合否入れ替わり率が 1 割までと低いため、素点に代えて Stanine を用いたとしても同様の傾向を示していると考えられる。今回は取り上げなかったが、後期日程では 1 次試験を 1/3 に圧縮して用いているので、今回の傾向がより顕著に表れると想像される。もし、

1 次試験の重みが大きい判定方法を採用している選抜単位であれば違いがより明確になったと思われる。

また、3 つの選抜単位の何れにおいても 2 つの試験の標準偏差を等しく設定した場合(等標準偏差)に、合否入れ替わり発生率が一番小さくなった。これは、等標準偏差になるように変換した後の両試験の得点分布が、従来の素点を用いた場合の分布形状と似た付置になり、結果として合否の判定が似通ったものとなっているからであろうと推察される。

今回は 1 次試験の倍数 k を 4 通り検討したが、例えば散らばり具合として統計指標の「範囲」を用いてこれを揃える等の他の根拠に基づく値で試行した場合の特性や、1 次試験の重みと段階表示の関係性、階級数を 9 以外に変更した際の合否に与える影響や各階級に含まれる受験者数割合、また、段階表示の合計値ではなく各科目ごとに段階表示を用いて選抜を行う方法等、今後もいろいろな利用方法を検討する必要がある。

7 まとめに代えて

今回、我々は大学入学共通テストの開始に伴って提供が始まった段階表示の利用に関して幾つか条件を設定してその利用方法を模索し特性を把握した。

今回調べた範囲では、N 大学においては段階表示を用いても合格者の劇的な入れ替わりが起こるような状況ではないことが判った。しかし、素点を用いた方法と段階表示を用いた方法で、特性の異なる志願者が分離できているのであれば、よりアドミッション・ポリシーに合致した志願者を合格させる方策を採る方が志願者・大学の双方にとって次善の策であると言える。そのためには、分離された両群の特性を明確にする必要があるが、一方で大学側が採り得る受験者の特性を把握する道具建てとしては試験成績や調査書程度であり、入学後の振る舞いを把握することは不合格者に対しては行うことができない。このように、客観的に比較・検討する手段に乏しく、両群それぞれの志願者の特性を明確にすることは容易ではないように思われ、またその特性が定常的なものであるかの検証も必要になる。

とは言え、段階表示の利活用に関する研究はまだ緒に就いたばかりである。大学の入学という「入り口」に重きをおいて選抜を行っている日本においては「1点刻みの客観性」については今後も問われ続けると思われる。そのためには、段階表示の特性を解明することによって、利用場面がよりクリアーになってくると考えられ、段階表示の利用を検討しようとする大学の支援に繋がるので、今後もいろいろな角度から段階表示に関する研究を続ける必要がある。

謝辞

本研究の遂行にあたっては JSPS 科研費 JP16H02051 および JP21H04409 の助成を受けた。

参考文献

- 大学入試センター (2019). 「令和3年度大学入学選抜に係る大学入学共通テスト問題作成方針」,
<https://www.dnc.ac.jp/albums/abm00038189.pdf> (2021年8月30日).
- 大学入試センター (2021). 「令和3年度大学入学共通テスト 段階表示換算表」,
<https://www.dnc.ac.jp/albums/abm00040264.pdf> (2021年8月30日).
- 林 篤裕, 高木 繁 (2021). 「大学入学共通テストにおける段階表示の提供開始に備えて」, 『大学入試研究ジャーナル』 **31**, 239-244.

前川 眞一 (2019). 「得点の段階表示とStanine」 『大学入試センター研究開発部 リサーチノート』 **RN-19-1**.

中央教育審議会 (2014). 「新しい時代にふさわしい高大接続の実現に向けた高等学校教育, 大学教育, 大学入学者選抜の一体的改革について～すべての若者が夢や目標を芽吹かせ, 未来に花開かせるために～」, 中央教育審議会答申 (中教審第177号) .

https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/tou shin/1354191.htm (2021年8月30日) .

United States Army Air Forces (1947). Stanines: selection and classification for air crew duty,

<https://collections.nlm.nih.gov/catalog/nlm:nlmuid-35721550R-bk> (2021年8月30日) .