

【原著】

個別学力検査問題の質を測る

— 作題に関する合格者アンケートによる試み —

木村拓也（長崎大学）

本稿の目的は、入試問題の受験者による主観的な評価と、学力検査の得点との関係を探ることで自学における入試問題の作題評価を行う可能性を論じることにある。そのために、合格通知に同封する形で作題に関する合格者アンケート調査を実施した。学生の学力低下や大学全入時代といった社会情勢のみならず、各個別大学での作題の機会の減少やそのノウハウの継承が危うくなっている昨今、自学の入試問題で識別力が高く且つ内容的妥当性を担保するような良問になっているかといった吟味が今後ますます求められてくる。

1. 問題の所在—個別学力検査問題の評価

過年度の入試問題を分析し評価するということは、次年度以降の作題を支援するために重要な情報であることに疑いはない。例えば、大学入試センター研究開発部では、共通第1次試験の初年度から、再下位群L、中の下位群LM、中位群M、中の上位群HM、上位群Mと人数の等しい5群に分割し、その群を横軸、選択肢の選択率を縦軸に取り、設問に対する受験生の反応を「設問解答率分析図」としてグラフ化し、設問の識別力の評価や誤答選択肢の評価を行っている（清水 1983）。また、菊池(1999)では、こうした「設問解答率分析図」やこの分析手法を大問の評価に応用した「大問得点率分析図」において得点分布の違いが分析図そのものに大きな影響を与えないなど、手法の安定性も実証されている。

こうした作題評価については、過去『大学入試フォーラム』上において1～10号（1983-1990）まで、計18科目について評価が掲載されていた経緯がある¹⁾。大学入試センター試験に制度変更があった後は、出題経験者が作題の所感を述べる「出題ノート」が『大学入試フォーラム』に毎年2科目ずつ[12～32号(1991-2009)まで「出題ノート」は34回を数える²⁾掲載されるようになって

いたが、現在は継続的にセンター試験の「設問解答率分析図」が公開されることはない。

だが、現在まで大学入試センター研究開発部で毎年続けられているこうした「設問解答率分析図」による作題評価も特段の時代的要請があるときには、大学入試研究として再び表舞台に登場している。例えば、共通第1次試験の最終年度である1989(平成元)年度の理科四科目に大きな平均点差があったときには、「設問解答率分析図」による作題評価が『大学入試フォーラム』（勝木 1994）や『大学入試研究ジャーナル』（勝木・石井 1995、島貫 1995、一國 1995、山田・池田・越田 1995）紙面上に登場したし、1998(平成10)年度入試から実施された新教科「情報関係基礎」の出題評価に際しては、新教科という理由での作題に関する経験不足や情報という教科自体の進歩の早さによる作題の困難さを少しでも埋めるために参考にした「大学入試センター試験問題作成支援のための統計情報データベース」（吉村他 2005）の1機能として『大学入試センター研究紀要』で紹介されたりした（植野他 2007）。

そうした「設問解答分析図」が大学入試センター試験の作題評価として定着している一方で、個別学力検査の作題評価についてはこ

れまであまり研究報告が多かった訳ではない³⁾。例外的に、東京農工大学の個別学力検査の英語問題について項目反応理論の1種である **Rash Analysis** をおこなった研究(ウィバー、佐藤 1995)や帯広畜産大学での全受験者アンケートがある(小柳 1996・1997)。前者は、「設問解答率分析図」と同様、量的データに基づく評価であるが、後者は、属性項目以外に「受験理由」「科目選択の理由」「問題の難易度」「解答時間の適切さ」「受験科目の多寡の評価」「センターと個別の配点比率の評価」「本学の知識」といった定性的評定を行い作題評価に活用した⁴⁾と報告されている。

実は、作題評価に定性的評定を用いることは、大学入試センター試験問題でも行われてきた(石塚・山田 1987、山田・岩坪 1992)。前者が高校段階の学習段階の達成度評価という指導要録に適した出題になっているかの「内容的妥当性」の評価(石塚・山田 1987:2)⁵⁾であり、後者が作題の事前評価により評定項目から設問正答率を推定する方式の構築(山田・岩坪 1992:37)など、研究目的こそ違えど、4段階の評定カテゴリーを設けて出題経験者や高校教員に定性的評定を行ってもらう方法論では共通している。

ところで、帯広畜産大学の事例では、学科の改組・個別試験の科目変更・選抜方法の方式変更などが僅か数年で起こり受験者集団の急激な変化が見込まれることからこれまでの作題経験が全く通用しないという事情から行われたものであった(小柳 1996:36、1997:45)。本学だけに留まらず、高校生の学力低下によって、その対応を見誤ると途端に個別学力検査の平均点が極端に低くなったり、理科などの教科で科目間の平均点に大きな差が生まれたりする可能性も否めない。また、作題経験者の定年退職などによってその経験が継承されなければ、そうしたリスクはより高まるだけに、今後ますます多くの作題支援情

報が必要とされよう。

本学でもそうした作題支援情報の提供の流れを敷衍し、量的データに基づいた個別学力検査問題において毎年「大問得点率分析図」による作題評価を行いその結果を作題委員会へ返却しているだけに留まらず、定性的評定による作題評価を行うため 2010(平成 22)年度の前期入試から入学手続者に対してアンケート調査を行い、作題委員会や各学部の FD でそのデータを活用することとした。本稿では、その一端を報告する。

2. 調査方法と質問項目の内容

調査対象者は、2010(平成 22)年度前期試験の合格者名 1187 名である。試験中にアンケート調査をすることは試験自体の実施の妨げになるので、事後の分析においてデータの制約にはなるものの調査票の配布は合格通知に同封し、回収は入学手続きの書類と一緒にいった。有効回答数は 888 名であり、回収率は 78.8%であった。本稿で用いる募集単位ごとデータの回収率は、表 1 の通りである。

質問項目は、大まかに述べれば、過去問を解いた年数などの「受験対策状況」「高校での履修科目」「高校での教科科目の得意・不得意」「個別学力検査問題の印象」「個別学力検査問題の傾向」である。小柳(1996・1997)で問われていた「問題の難易度」や「解答時間の適切さ」だけに留まらず、本学独自項目として「問題量が多い」「試験問題の傾向を知っている」「高校の授業だけで十

表1 各学部・学科のアンケート回収率

募集単位	回収率
A	83.5%
B	76.3%
C	84.1%
D	84.4%
E	82.8%
F	77.6%
G	77.8%
H	76.9%

分に対策できる」「特別な受験対策が必要である」「出題傾向が変わった」などの項目を加えている。

3. 受験対策の評価――大学独自科目の評価

小論文や総合問題などの出題を自大学で行っている場合、その出題形式が近隣の競合大学の出題と共通性があるとか、受験科目の互換性が担保されているか、或いは、その難易度如何によって、受験生の受験行動に影響を与える場合が多々ある。

例えば、志願者の出身地を県内・県外比率で比べた場合、一般入試で県内比率が高く、推薦入試などの特別選抜やAO入試での県外比率が高い場合、一般入試で課す教科目やその組み合わせが県外の高校で対応しづらいものである可能性も否めない。こうした受験対策の評価を県内・県外で見えていくことはひいては志願者の受験行動の評価につながっていくであろう。そうした受験行動に関する質問項目について単純集計を行った上で、そうし

た受験行動の評価を多重対応分析（例えば、大隅・ルバール他 1994）にて行ってみた。分析に際して使用したデータは小論文を個別学力検査で実施している募集単位Aのものであり、使用した統計ソフトはJMP8である。

まず、県内／県外という区分で、「志望順位」「出願の際の気持ち」「過去問を解いた年数」「出題傾向を知っているか」について単純集計をもとに大まかな傾向について見ていきたい。「志望順位」については、割合で見ると「第一志望」が県内82.9%、県外60.0%と県内の合格者の方が第一志望で入学する学生が多いことが分かる（図1）。「出願の際の気持ち」については、「合格圏内」と見込んで出願する割合で見ると、県外76.0%の方が県内53.7%よりも高く、一方で「二次逆転」狙いでの出願は、割合として県内19.5%の方が県外4.0%と高い（図2）。県外からの志願し合格に至るものはより安全策を採った上での志願であることがみてとれる。これを裏付けるように、「過去問を解いた年数」を

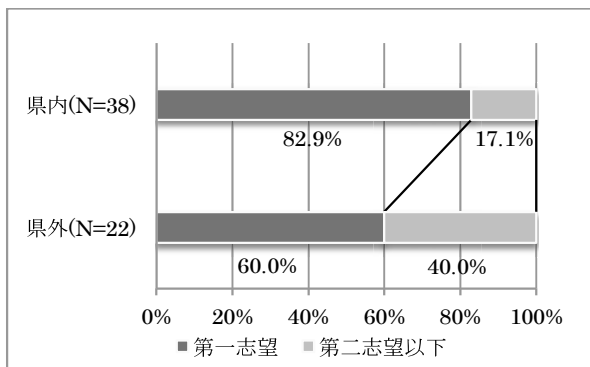


図1 志望順位

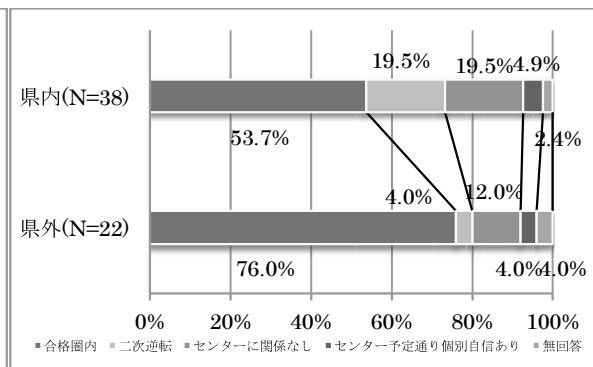


図2 出願の際の気持ち

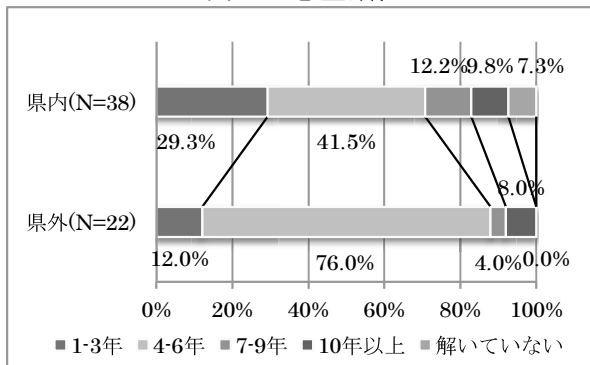


図3 過去問を解いた年数

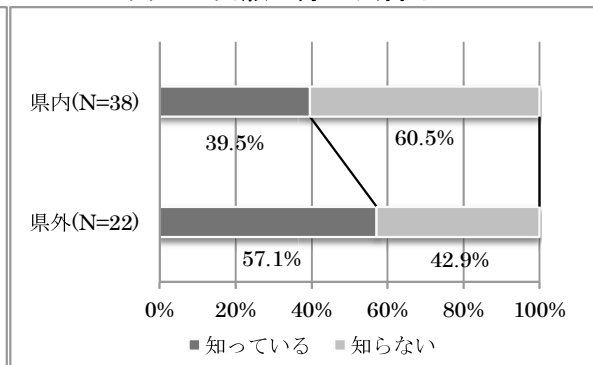


図4 出題傾向を知っている

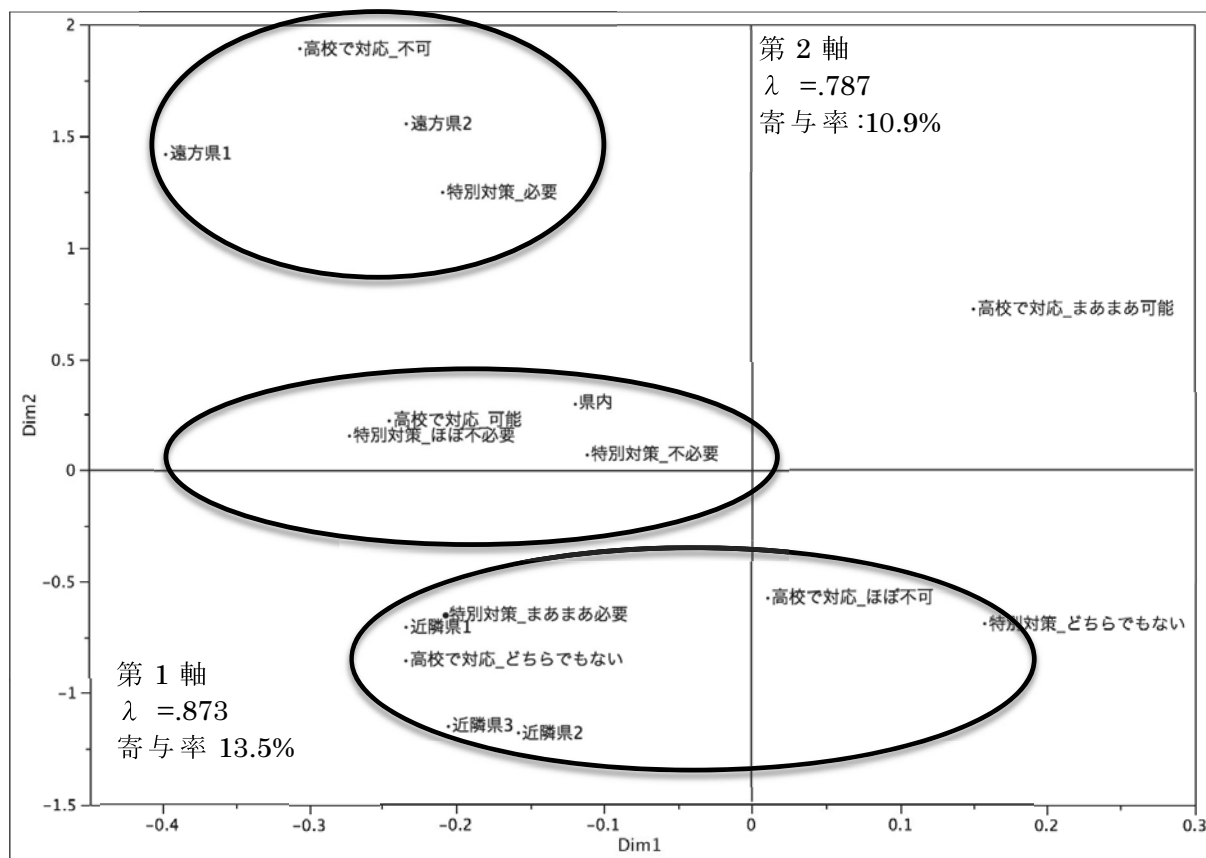


図5 多重対応分析による県別受験行動の評価

見ると、県内の学生で「解いていない」と答えたのが7.3%いたり、「1-3年」とあまり解いていない学生も29.3%いたり、一部で7年以上解いて受験した県内の学生が22.0%占めるにしても、県外からの受験した学生で「4-6年」分解いた学生が76.0%いることを踏まえると、県外の合格者の方が過去問をより多く解いてから受験に臨んでいた(図3)。ちなみに過去問を掲載している教学社の通称『赤本』の掲載年数がたいてい3年単位であることから、「4-6年」分といて受験した県外の合格者は、『赤本』を2冊分の過去問題を解いて試験に臨み合格に至ったことが分かる。また、「出題傾向を知っている」という質問に対しても、「知っている」の割合が県外57.1%県内39.5%と県外の方が高い(図4)。

次に、「高校の授業で対応可能」「特別な受験対策が必要」という質問項目と合格者の

出身県項目を用い、多重対応分析をおこなったものが図5である。大まかに3グループほどに分けることができたが、まず、「遠方県1・2」の場合、「高校で対応が不可」「特別な受験対策が必要」というプロットと同じような配置にある。それと同様の傾向を示しているのが「近隣県1・2・3」の場合であり、例え近隣県であっても「高校の授業での対応」が「ほぼ不可」や「特別な受験対策」が「まあまあ必要」などのプロットと同じような配置にある。これに対して「県内」の志願者のみが「高校の授業での対応」が「可能」や「特別な受験対策」が「ほぼ不必要」「不必要」となり、受験勉強の容易さが伺える。

これらの二つの結果を踏まえると、受験生に対して入試問題の取り組みやすさに県の内外で違いある一方で、県外からの志願者で合格に至るものは高校に頼らず自力で受験問題

に取り組み合格を勝ち取っていることが分かった。このように合格者に限定される制約はあるものの、受験行動の評価にも合格者アンケートが一定の役割を果たすことが伺える。

4. 作題に関する評価 – 教科 X の事例

次に、募集単位 B～H のデータを用いて、教科 X における合格者の入試問題イメージについて見ていきたい。募集単位 A については、個別学力検査において教科 X を課していないので、分析対象からは除外している。ここで検証したいのは、「その教科が得意な人がきちんと得点ができていない試験になっているのか？」ということである。もし、高校生の学力低下にともない、自大学の入試問題が必要以上に志願者に難しく感じさせるものとなってしまった場合、高校教員も含めてそうしたイメージの定着がひいては受験回避の行動を誘引することにも繋がりがかねない。そういった理由でこうした印象評価についてチェックを怠らないことも重要なアドミッションセンターの業務であると考えている。また、五肢選択式の解答と異なり、個別学力試験では、特に、大問形式で記述式の問題であればあるほど、採点基準の設定によっては大問分析が

意味している内容が掴めないことが多い。データを分析したり検討を加えたりするもの全員が答案の採点をしているのではないので、そうした感覚も共有できる方がより望ましいであろう。

早速データを見ていくと、「教科 X が得意である」と答えている割合が多いのが、募集単位 C 58.1%、募集単位 F 74.3%、募集単位 G 74.3%、募集単位 H 75.0% である（図6）。次に、得意／不得意という得意度の 2 カテゴリー、及び、個別試験の制限時間が足りなかった／足りた、の 2 カテゴリーについて、センター試験および個別学力検査の素点について平均値差の検定(Student の t 検定)をおこなった結果が表 2・表 3 である⁶⁾。

これらを見ると、個別学力検査 γ においては、概ね「得意」と答えている合格者の方が「不得意」と答えた合格者よりも点数差において統計的に有意な差があり、合格者に限って言えば、教科 X が得意な人がきちんと得点ができ差がつく試験になっており、制限時間の多寡による個別学力検査の点数に影響差がでておらず、試験時間として妥当であったなどと一定の評価ができるであろう。

表 2 得意 / 不得意によるセンター・個別学力検査の平均値差の検定 (Students の t 検定)

募集単位	センター試験 α					センター試験 β					個別学力検査 γ				
	差分	SD	t 値	df	p 値	差分	SD	t 値	df	p 値	差分	SD	t 値	Df	p 値
B	7.52	1.54	4.89	203	<.0001	8.34	1.64	5.07	203	<.0001	19.41	2.34	8.29	203	<.0001
C	-1.38	3.24	-.42	72	.6712	.21	1.53	.14	72	.8876	12.20	5.53	2.20	72	.0308
D	13.47	4.47	3.01	25	.0059	5.62	3.13	1.80	25	.0843	28.01	11.03	2.54	25	.0177
E	3.25	5.33	.61	21	.5490	8.83	5.19	1.70	21	.1031	15.79	13.00	1.22	21	.2375
F	6.09	1.75	3.47	245	.0006	6.76	1.60	4.23	245	<.0001	19.47	4.44	4.38	245	<.0001
G	-4.26	4.32	-.98	31	.3320	9.38	3.77	2.48	31	.0187	24.63	8.09	3.04	31	.0047
H	-.92	5.22	-.18	36	.8615	12.45	3.68	3.38	36	.0017	12.75	6.26	2.04	36	.0489

表 3 時間が足りなかった / 足りたによるセンター・個別学力検査の平均値差の検定 (Students の t 検定)⁶⁾

募集単位	センター試験 α					センター試験 β					個別学力検査 γ				
	差分	SD	t 値	df	p 値	差分	SD	t 値	df	p 値	差分	SD	t 値	Df	p 値
B	-1.60	1.83	-.87	204	.3857	6.04	1.96	3.07	204	.0024	-.65	3.07	-.21	204	.8321
C	1.15	2.94	.39	72	.6965	2.59	1.36	1.91	72	.0607	1.38	5.19	.26	72	.7918
D	3.16	5.40	.58	25	.5643	0.45	3.46	.13	25	.8977	-16.0	12.5	-1.28	25	.2126
E	-0.10	5.60	-.02	21	.9859	-7.75	5.51	-1.41	21	.1739	-15.6	13.6	-1.15	21	.2631
F	1.79	1.75	1.02	249	.3081	-1.02	1.63	-.63	249	.5317	-2.31	4.50	-.51	249	.6085
G	-4.87	3.97	-1.22	33	.2285	1.60	3.73	.43	33	.6704	0.57	8.32	.07	33	.9459
H	9.23	4.61	2.00	38	.0525	-1.46	4.68	-.31	38	.7571	1.73	6.22	.28	38	.7824

但し、p 値の値は両側検定の値である。また p 値が 5% 未満のものには網掛けをしてある。

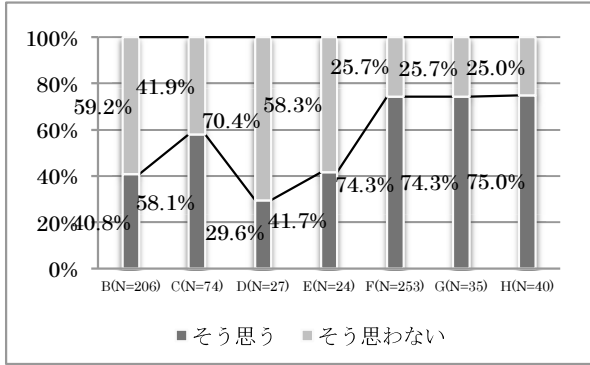


図6 教科Xが得意である

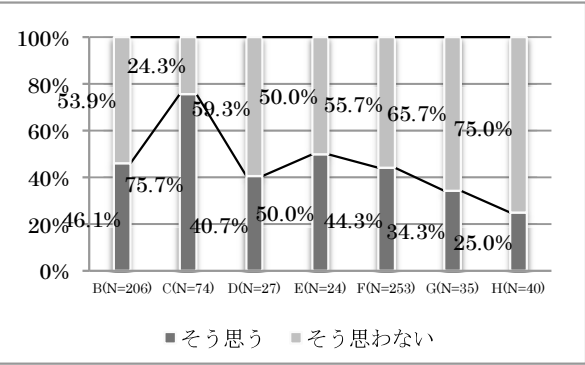


図7 教科Xの問題が難しい

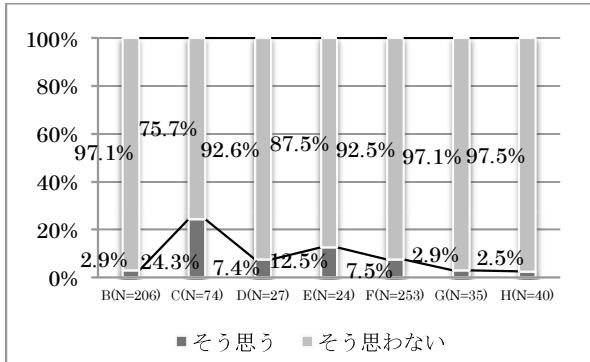


図8 教科Xの問題量が多い

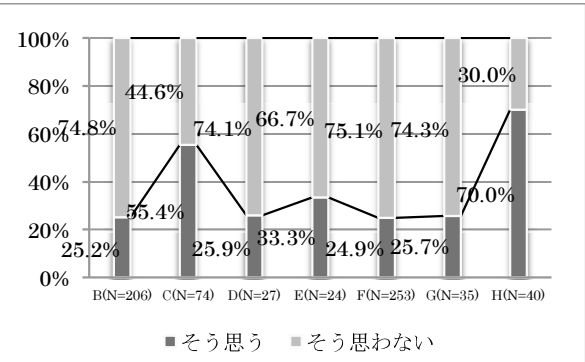


図9 教科Xの解答時間が足りない

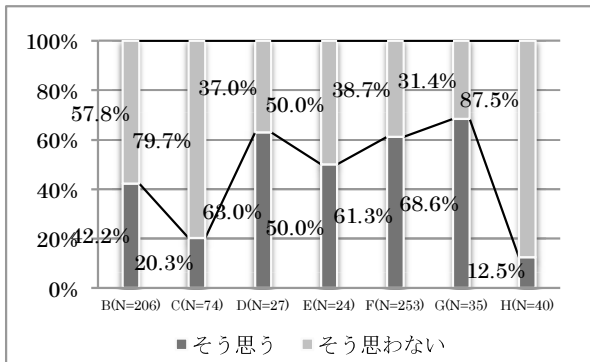


図10 高校の授業で対策が可能

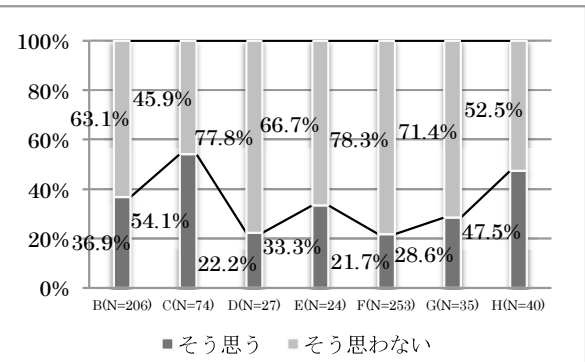


図11 特別な受験対策が必要である

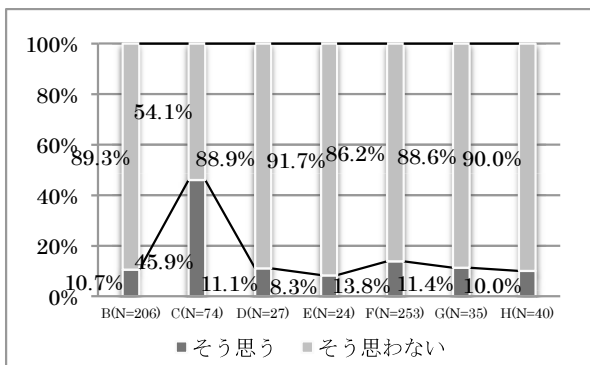


図12 教科Xの出題傾向が変わった

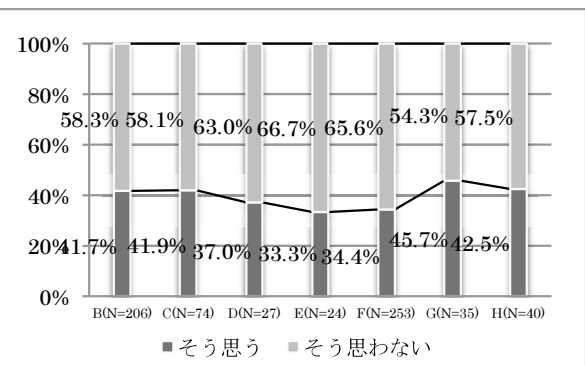


図13 教科Xの出題傾向を知っている

また、表 2 を詳しく見てみると、募集単位ごとに「得意」の意味の違いも見えてくる。例えば、募集単位 C においては、個別学力検査 γ では素点において統計的に有意な差があるものの、センター試験においては統計的に有意な差はない。つまり、個別学力検査 γ の問題で差が付くか否かが募集単位 C における「得意／不得意」を分ける基準になっていると考えられるし、募集単位 G や募集単位 H では、センター試験 α では統計的有意差はでないものの、センター試験 β や個別学力検査 γ では統計的な有意差がでる。合格者データという特性上、こうした結果は、入学後の学習指導にも役立つ情報であると言えよう。

また、図 7 ～ 13 までは、募集単位ごとに教科 X について「問題の難易度」「問題量の多寡」「制限時間の多寡」「高校の授業で対応可能」「特別な受験対策が必要」「出題傾向が変わった」「出題傾向を知っている」を単純集計して割合で表示した結果である。これを見れば、例えば、募集単位 C では問題が難しく半数で時間が足りないと答えたものの、問題量自体が多いわけではなく、出題傾向が変わったと感じている。或いは、募集単位 H などを見れば、教科 X の解答時間が足りないと答える合格者が多く、高校での授業で対策が不可能であり、半数が特別な受験対策が必要であると答えている。このように、作題者の意図がどれだけ受験生に伝わっているかの確認をすることができる。

更に踏み込んで、「その教科が得意な人がきちんと得点ができている試験になっているのか？」ということを問うためにオッズ比で検討してみたい。

図 14 は、「得意／不得意」の 2 値と「難

表 3 2 × 2 のクロス表の例 (図 14 の場合)

	難しくない	難しい
得意	N_{11}	N_{12}
不得意	N_{21}	N_{22}

しくない／易しい」の 2 値で 2 × 2 のクロス表 (例えば、表 3) を作り、対数オッズ比と 95% 水準での信頼区間を求めたものをグラフ化したものである。

同様に、図 15 は、「得意／不得意」の 2 値と「制限時間が足りた／足りない」の 2 値で 2 × 2 のクロス表を作り、対数オッズ比と 95% 水準での信頼区間を求めたものをグラフ化したものである。

例えば、表 3 のとき、オッズ比 θ は次式で表される。

$$\hat{\theta} = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

この標本オッズ比 θ を対数変換した $\log \theta$ の標本分布は大標本のもとで正規分布に近似することが知られており、漸近標準誤差 (ASE) と呼ばれる標準偏差は次式で表される (Agresti 1996:32、太郎丸 2004:72)。

$$ASE(\log \hat{\theta}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

よって、95% 信頼区間は次式で表される。

$$\log \hat{\theta} \pm 1.96 * ASE(\log \hat{\theta})$$

本データはすべてが大標本のサンプルサイズではないが参考までにグラフ内に 95% の信頼区間を表示しておいた。対数変換を行うとオッズ比の 1 が対数オッズ比の 0 にあたる。つまり、1 行目のオッズ N_{11} / N_{12} と 2 行目のオッズが N_{21} / N_{22} 丁度等しいときに対数オッズ比で 0 にあたり、 N_{11} や N_{22} の値が大きいとき対数オッズ比はプラスの値を取り、 N_{12} や N_{21} の値が大きいとき対数オッズ比はマイナスの値となり視覚的に解釈しやすい。

そこで、図各々結果をみれば、図 14 では、募集単位 B ・募集単位 E ・募集単位 G で標本対数オッズ比が大きくプラスの値になっており、教科 X が得意な人が試験を難しくないと

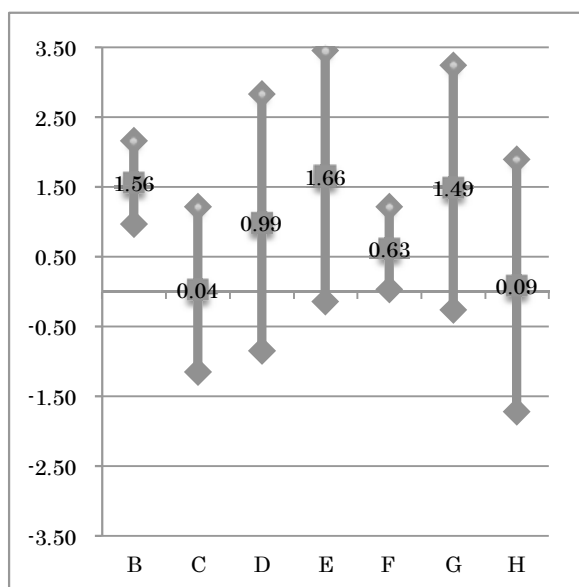


図14 得意度×難易度の対数オッズ比

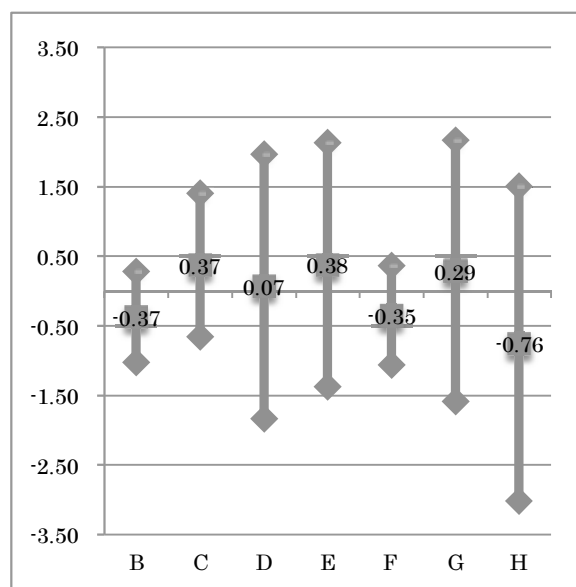


図15 得意度×解答時間の対数オッズ比

感じ、不得意な人が難しいと印象を持ったと言えよう。また、図 15 と比べて全体的に標本対数オッズ比がプラスの値になるなど、この傾向が教科 X の試験全体に見られたことが分かる。また、図 15 では、募集単位 B・募集単位 F・募集単位 H でほんの僅かではあるが標本対数オッズ比がマイナスの値になっており、多少教科 X が得意な合格者で解答時間が足りなかった、と答えた割合が高いことが伺える。このように対数オッズ比を用いれば、合格者が抱く自学の入試問題のイメージが単純集計よりも多くの情報が得られることが示唆された。

5. まとめ—学力低下・大学全入学時代における作題評価の困難とその可能性

林・伊藤・田栗(2008)の調査が明らかにしたように、自学の入試問題における「設問の難易度・識別力の評価」は、国公私立 352 校中、現在実施しているのが僅か 10 校(2.8%)、過去実施していたのも僅か 7 校(2.0%)、これから実施計画があるのが僅か 5 校(1.4%)という非常にマイナーな研究対象であることは否めない(林・伊藤・田栗 2008:152)。だが、

そもそも入学者選抜研究の歴史を紐解けば、入学者選抜方法の妥当性を吟味し世間の批判に耐えうる入学者選抜を行う「リスク管理」を行う専門機関として、「入学者選抜方法研究委員会」として予算が文部省からつけられ各国立大学に組織化されていったのが、そもそものアドミッションセンターの淵源である(木村2008:88)。1982年に刊行された国立大学入学者選抜研究協議会連絡協議会の『大学入試研究の動向』創刊号でも、主な研究テーマ 1 の(1)として「共通 1 次試験(ママ)と第 2 次試験の得点の分布や平均値、標準偏差など」と入学者選抜研究の「基礎研究」として真っ先に挙げられている(入研協 1982:13)。ただ、近年アドミッションセンターが AO 入試の実施とその評価に結びつけられすぎてきたために「総合的かつ多面的な評価」方法の開発とその制度としての検証作業に多くの労苦を注いできた。だが、その労苦にもかかわらず、特段これといった決定打を出せていないのも現実であろう。

その一方で共通第 1 次学力試験の開始当初は盛んに行われてきた作題支援情報についての研究にももう一度スポットを当てても良い

ように思われる。高校生の学力低下や大学全入時代といった大学を取り巻く社会情勢のみならず、大学内の状況として各個別大学での作題の機会の減少やそのノウハウの継承が危うい昨今、テスト理論の基本に立ち戻って、自学の入試問題では識別力が高く且つ内容的妥当性を担保するような良問になっているかといった吟味が求められてくると考える。本稿では、統計手法にそんなに通じていなくとも、ある程度統計的にシンプルな方法で統計ソフトの簡単な操作や Excel さえあればすぐにできる作題評価の可能性について論じてきた。受験対策の評価や作題に関する評価をフィードバックすることで、大学としての入試制度の設計や次年度以降の作題に繋げて貰う体制の担保もアドミッションセンターに課せられた大きな役割の一つであると考えられる。

注

1) 掲載科目の変遷は次表の通りである。

発行年 / 号数	掲載科目
1983 No. 1	化学
1983 No. 2	数学 I 英語 B
1984 No. 3	国語
1984 No. 4	生物 I
1985 No. 5	政治経済 世界史 地理 A 地理 B
1985 No. 6	日本史 物理 I 地学 I
1986 No. 7	国語
1986 No. 8	数学 I 英語 B
1987 No. 9	世界史 生物
1988 No.10	日本史 化学

2) 掲載科目の変遷は次表の通りである。

発行年 / 号数	掲載科目
1991 No.12	世界史 物理
1992 No.14	国語 生物
1993 No.16	日本史 化学
1994 No.17	政治・経済 数学
1995 No.18	地理 理科 I
1996 No.19	フランス語 地学
1997 No.20	生物 英語
1998 No.21	倫理 ドイツ語
1999 No.22	国語 物理
2000 No.23	数学 日本史
2001 No.24	政経 情報
2002 No.25	化学 現代社会
2004 No.27	地学
2005 No.28	政治・経済 化学
2006 No.29	化学 英語
2007 No.30	日本史 地学
2008 No.31	英語
2009 No.32	地理 数学

- 3) 個別学力検査の研究については、共通第 1 次学力試験のときは、共通第 1 次学力試験との関連性から 2 次試験研究として盛んだった時代が存在する。古くは、鈴木・高野(1992)のように 2 次試験がどの程度合否に影響を与えているかという研究であったり、柳井・池田・前川(1988)のように共通第 1 次学力試験との捉え方の対比として 2 次試験の性格を大学生の意識構造から解き明かす研究が存在したりしている。
- 4) その他に評定と素点を組み合わせる手法としては、確信度テスト法(妻藤 2004、張 2007)がある。
- 5) こうした妥当性評価については、統計的指標で表しにくい。石塚・山田(1987)が指摘しているように、数量的な妥当性基準担保の殆どは基準連関妥当性によるものであり、「内容的妥当性」の評価はこうした定性的評価なくしては成立しにくい(石塚・山田 1987:3-4)。

引用文献

- Agresti, A.(1996). *An Introduction to Categorical Data Analysis*, John Wiley & Sons, Inc. =(2003). 『カテゴリーカルデータ解析入門』サイエンティスト社。
- 林篤裕・伊藤圭・田栗正章 (2008). 「大学で実施されている入試研究の実態調査」『大学入試研究ジャーナル』 18 号、147-153.
- 一國雅巳 (1995). 「大学入試センター試験『化学』における問題の難易度」『大学入試研究ジャーナル』 5 号、76-81.
- 石塚智一・山田文康 (1987). 「共通第 1 次学力試験の試験問題の分析—昭和 57 年度『化学 I』を中心として」『大学入試センター研究紀要』 16、1-51.
- 勝木渥 (1994). 「平成元年度共通 1 次学力試験『物理』および平成 2 年度と平成 3 年度の大学入試センター試験『物理』の解答の分析結果からみた問題の評価」『大学入試フォーラム』 14、3-34.
- 勝木渥・石井廣湖 (1995). 「受験生の『物理』

- 理解の傾向と問題点—平成5年度大学入試センター試験の『物理』の解答分析から』『大学入試研究ジャーナル』5号、62-69.
- 菊池賢一(1999).「項目反応理論をもちいた設問回答率分析図の評価」『大学入試センター研究紀要』29、1-8.
- 木村拓也(2008)「アドミッションセンターの系譜学」『日本テスト学会第6回大会発表論文抄録集』88-91.
- 国立大学入学者選抜研究連絡協議会(1982).『大学入試研究の動向』創刊号、1-40.
- 小柳敏郎(1996).「受験者に対するアンケート調査」『大学入試研究ジャーナル』6号、36-42.
- 小柳敏郎(1997).「受験者に対するアンケート調査(2)」『大学入試研究ジャーナル』7号、45-55.
- クリストファー・ウィバー、佐藤容子(2005).「個別学力検査のラッシュ分析」『大学入試研究ジャーナル』15号、147-153.
- 大隅昇・L.ルバール、他(1994).記述的多変量解析法、日科技連.
- 島貫隆(1995).「大学入試センター試験『地学』における出題形式と階層別正答率の関係」『大学入試研究ジャーナル』5号、70-75.
- 清水留三郎(1983).「共通1次試験の実施結果の概要」『大学入試フォーラム』1、32-35.
- 鈴木規夫・高野文夫(1992).「第2次試験の効果の分析」『大学入試研究ジャーナル』2号、14-20.
- 太郎丸博(2005).『人文・社会科学のためのカテゴリーカル・データ解析入門』ナカニシヤ出版.
- 張一平(2007).『確信度テスト法と項目反応理論』東京大学出版会.
- 妻藤真彦(2004).『確信度評定のメカニズムと理論的問題』風間書房.
- 植野正臣・吉村幸・荘島宏二郎・橋本貴充(2007).「大学入試センター試験『情報関係基礎』の統計分析」『大学入試センター研究紀要』36、71-100.
- 山田文康・池田輝政・越田豊(1995).「大学入試センター試験『生物』における階層別正答率と出題形式」『大学入試研究ジャーナル』5号、82-90.
- 山田文康・岩坪秀一(1992).「試験問題の定性的評価に基づく科目難易度の推定—社会・理科についての第1次報告」『大学入試研究ジャーナル』2号、37-41.
- 柳井晴夫・池田央・前川眞一(1988).「共通第1次学力試験と第2次試験に関する大学生の意識構造の分析」『大学入試センター研究紀要』17、145-184.
- 吉村幸・荘島宏二郎・中畝菜穂子・石岡恒憲・橋本貴充実(2005).「大学入試センター試験問題作成のための統計情報データベースの開発」石塚智一(研究代表)『平成14年度～16年度共同研究I報告書 試験問題統計情報の整備に関する研究』大学入試センター研究開発部、21-38.