

論 文

誤答の分析と対策の検討 —テスト・デザインの視点から—

土田 泰子¹

¹一般教育科—英語 (Liberal Arts-English, Nagaoka National College of Technology)

ERROR ANALYSIS AND EXAMINATION FOR IMPROVEMENT IN PAPER TEST DESIGN

Yasuko TSUCHIDA¹

Abstract

While human error analysis and product design based on usability have been studied and performed in this decade, paper test for educational purpose keeps its typical form though it weighs in evaluating learners' achievement. In order to improve test design both in structure and appearance, it is necessary to demonstrate the mechanism that brings errors to learners. This paper aims to reveal why error occurs and how human makes mistakes through considering examinations or paper tests in English class. To inspect this issue, 119 1st graders in Nagaoka National College of Technology were examined using their TOEIC-Bridge score, 1st mid-term and final examination in 2008.

According to the result shown in the bubble chart, it developed that unsuitability of test planning, direction for examiners and form design can bring errors besides learners' faults and shortage of preparation. Not only examiners, the user of a test consists of also researchers, makers, providers, markers, utilizers, and managers. To improve usability of paper test, it is inevitable to consider its design as human interface and reduce or eliminate errors brought by unsuitability and human fault.

Key Words : *educational assessment, error analysis, usability, human interface*

1. はじめに

なぜテストで間違えるのだろうか。間違いには学習上の問題に起因するものと、行動上の問題に起因するものがある。英語のペーパーテストを例にすると、主語と動詞および時制の関連の理解が不十分なために生じる間違いは学習上の問題に起因し、解答欄を間違えたために生じる誤答は行動上の問題に起因する。英語で解答すべき問題を日本語で解答し、さらに解答でスペルミスをするという、複合的な誤答も生じる。テストの結果を評価する上で、数値上の採

点データには表れにくい、誤答の原因となるこれらの要素をいかに汲み上げ、なぜ誤答となったのかを分析することは、効果的な指導を行なう上で不可欠の取り組みである。

ここでは、英語のペーパーテストにおける誤答分析の方法と、そこから得られる調査結果がどのような問題を示唆しているのかを考察し、改善の提案と課題の発見を試みたい。

1. 1 研究の背景

誤り、すなわちヒューマンエラーの分析には様々

な方法があるが、大別すると結果から行なう分析と原因から行なう分析がある。ここではテストでなぜ間違えるのか、どうして間違えたのかを明らかにするために、原因による分析を行なうこととする。小松原明哲はヒューマンエラーの原因を6つに単純化している¹⁾。この分類を元に作成したのが表-1であり、ペーパーテストで生じる誤答をヒューマンエラーの観点から捉えなおすと次のようになる。

表-1 間違いとその原因の分類

分類	内容
(1) 無理な相談	人間能力的不可能
(2) 錯誤	取り違い, 思い違い
(3) 失念	し忘れ
(4) 能力不足	能力, 技量が不足している
(5) 知識不足	すべきことを知らない
(6) 違反	手抜きや怠慢

(作表は筆者による)

(1) 「無理な相談」

出題範囲が未習である、難易度が極めて高い、問題数が多すぎるなど受験者の能力に合わないテストは、問題の意味がわからない、解答時間が足りない等の理由により「無理な相談」による誤答の原因となる。

(2) 「錯誤」

例えば英語のペーパーテストにおいて、英語で答える問題と日本語で答える問題が混在すると、思い込みや思い違い等の理由により「錯誤」による誤答の原因となる。日本人が英単語を学習する際に生じるスペルのローマ字化等の第二言語学習者における母語の干渉による誤答などもこの「錯誤」によるものと分類できる。

(3) 「失念」

「失念」はマークシートへの転記をし忘れる、後で解くつもりで飛ばしていた問題をそのまま解き忘れる等の理由による誤答が例として挙げられる。

(4) 「能力不足」

英文を読む速度が遅いため解答時間が不足する、自然な速度で行なわれる会話を聞き取るリスニング問題に対応できない等は「能力不足」による誤答の原因となる。「無理な相談」との判別は、対処すべき作業が人間の処理能力を超えた要求が行なわれているのか、あるいは訓練を行なうことで対応できるのかによる。

(5) 「知識不足」

単語の知識がない、なぜそうなるのかという原理

に対する知識がないため定型的な問題は解けても応用問題には対応できない等の理由による誤答は「能力不足」に原因がある。教育カリキュラムの中で実施されるペーパーテストに最も多い誤答の理由として挙げられると推測される。

(6) 「違反」

定められたマナーや規則を守らないことから生じるエラーは「違反」に原因があると分類される。ペーパーテストでの事例を挙げると、マークシートのマークの塗り方が薄い、ダブルマークをする等の原因による誤答がある。解答上の制限文字数を超える、定められた語群から解答しない等の誤答も「違反」に分類できる。なお、記号で答えるべき設問に語句で答えるなどは、思い違いに基づく「錯誤」に起因すると分類できる。

1. 2 研究の目的

言語学習、特に第二言語習得論において誤答分析が言語教育に位置付けられたのは1960年代であり、学習者は何をどのように間違えるのかということが長く研究されてきた。特に英語教育の分野では、高等表現における誤りや筆記上の誤りなど、言語習得の過程における誤りに主眼の置かれた誤答分析研究が行われてきた²⁾。しかしここでは研究の対象を英語のペーパーテストに設定し、誤答の類型化を試みたい。その調査と考察の結果から、教科や科目に限定されない誤答分析手法の可能性について検証を進展させたい。すなわち、教育カリキュラム上に位置付けられるペーパーテストにおいて、学習の到達度を測定するツールとしてペーパーテストがどのように機能しているのかを誤答分析から考察し、誤答の本質を明らかにすること、そして誤答をなくすための改善の可能性を探ることが本研究の目的である。

2. 研究の方法

今回分析に用いるのは、本校定期試験のうち、2008年実施一年生英語Ⅰの前期中間試験（6月4日実施）と前期期末試験（7月28日実施）である。調査の対象は一年生のうち、両方の試験に対して後述する「解答分析シート」の提出があった119名である。いずれかの試験を欠いた者、「解答分析シート」の記述が不完全であった者は除外している。人数はのべ数ではなく、2つの定期試験を受け、それぞれの試験に対する「解答分析シート」を提出した、同一の119名を調査の対象としている。

一年生を調査の対象として選んだのは前期中間試験が本校に入学してから始めて受験する定期試験であり、試験の形式や問題設定に慣れていない状態での調査結果と、一年生にとって二度目の試験となる前期期末試験での調査結果とを比較することを目的としていることによる。

分析に用いる定期試験の科目は英語 I で、主に英語の読解力と聴解力を養うことを目的とした指導を週に 2 単位行なっている。分析に先立ち定期試験の客観性を判断する上で、入学当初の 4 月 15 日に実施した TOEIC-Bridge テストとの成績の相関を図-1 にまとめた。

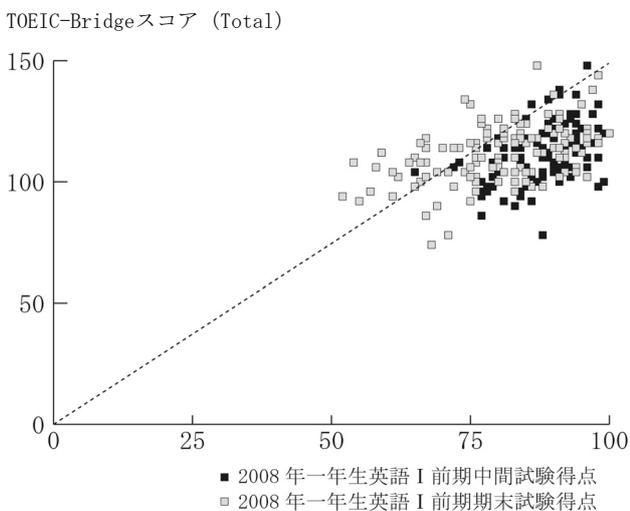


図-1 TOEIC-Bridgeスコアと定期試験得点の散布図

図-1において標本となる受験者は先に述べた119名であり、後述する表-4の回答者に等しい。TOEIC-Bridgeテストの数値はリスニングテストとリーディングテストの合計得点を用いて作表している。いずれの定期試験も100点満点で実施し、調査対象者における前期中間試験の平均点は88.1点、前期期末試験の平均点は80.8点であった。またTOEIC-Bridgeテストは150点満点で、調査対象者における平均点は112.3点であった。得点の分布に差が生じている原因としては、前期中間試験と前期期末試験では実施時期が異なること、また問題作成者が異なることなどが考えられる。

定期試験とTOEIC-Bridgeテストの成績に関しては概ね相関があるものの、図-1においてグラフの対角線よりも下側に数値が集まっていることから、定期試験では高い成績を取る学生においても、英語によるコミュニケーション能力を測定するTOEIC-Bridgeテストでは平均的な得点となっていることがわかる。このことは、いずれのテストもあらかじめ決められ

た知識や技術のレベルに学習者が到達しているかどうかを測定する基準準拠評価の性格が強いものの、客観的な英語運用能力を測定するTOEIC-Bridgeテストとは異なり、定期試験は学習カリキュラムの中に位置付けられ、既習の学習内容に関して理解および習熟の到達度を測定する機能を持つものとして利用されていることに起因すると考えられる³⁾。

2. 1 誤答分析

各定期試験を授業時に返却する際、試験解説と並行した形で学生に「解答分析シート」(図-2)を用いた誤答分析を行なわせた。今回調査に利用した定期試験は前期中間試験、前期期末試験共に7割程度をマークシート方式で実施しているため、採点の段階では各設問についてどの選択肢を選んだのかというデータとその正誤という情報しか得られない。解答が空欄の場合には解答時間が足りずに空欄となったのか、問題が解けずに空欄としたのか、それとも他に理由があるのか、採点者には推測することしかできない。誤答の理由を知るためには、なぜその答えを選んだのかをできるだけ試験から日が経たないうちに、受験者本人に考えさせる必要がある。この解答分析シートは筆者が独自に作成したものである。

1 学年 英語 I 前期期末試験
(2008年 7月 28日 (月) 2 限実施)

組	番号

解答分析

大問	小問	内容	配点	得点	誤答の理由
1	(1)	語彙知識	1		
	(2)	語彙知識	1		
	(3)	語彙知識	1		
	(4)	語彙知識	1		
	(5)	語彙知識	1		
2	(6)	語彙知識	1		

大問	小問	p
7 問(B)	(A)	内容
	①	内容
	②	内容
	③	内容
	④	内容
	⑤	内容

図-2 解答分析シート (部分)

解答分析シートには大問番号、小問番号、内容(語彙知識や聴解など、その設問で何が問われたかを簡潔に示したもの)、配点が予め記載されており、受験者は各自の得点と誤答の理由を小問毎に記入する形式となっている。別紙として記入方法を記載したものを配布し、受験者はそれを参照して誤答の理由を項目番号で記入する。小問には欄外に通し番号が付けられているが、これは総問題数を示す一方で集計の際の指標としても機能する。解答できず空欄であった小問の得点欄には△印を記入することとし、記入後は受験者の誤答状況を一覧することが可能と

なる。また受験者は各自の得手不得手を視覚的に認識し、改善点の発見につなげることができる。このため解答分析シートにはテストの感想と反省を記入する欄が設けられている。受験者が記入作業をしている間に教員が机間巡視し、後述する誤答理由の選択に躊躇する学生に対するサポートを並行することで、学生に対して誤答についての分析的な視点を持たせながら、より適した誤答理由を選択させ、分析の精度を高めることを心がけた。

2. 2 誤答の理由

誤答には様々な理由がある。今回誤答分析を行なうにあたって、理由として例示したのは次の15項目である。

[誤答の理由]

- 1：時間が足りず、問題を最後まで読むことができなかった。
- 2：問題は最後まで読めたが、解答の時間が足りなかった。
- 3：問題は理解できたが、答えが全くわからなかった。
- 4：問題は理解できたが、答え方がわからなかった。
- 5：問題は理解できたが、正しく答えられなかった。
(以下の6～15に当てはまる場合を除く)
- 6：問題は理解できたが、間違った答えを覚えていた。(英文の内容を間違えて理解していた、間違ったスペルや発音を覚えていた、など)
- 7：問題は理解できたが、答え方を間違えた。
(英語で答える問題を日本語で答えた、語群から選ばなかった、など)
- 8：問題を理解できなかった。(問題で何を聞かれているのかわからなかった。)
- 9：リスニング問題の音声聞き取れなかった。
(早い、難しい、など)
- 10：リスニング問題の音量が小さく聞き取れなかった。(席が遠い、騒音があった、など)
- 11：問題が見づらくて理解できなかった。(文字が小さい、印刷が不鮮明、配置が不明瞭、など)
- 12：問題があることに気づけなかった。(裏面に問題があるのに気づけなかった、など)
- 13：解答欄をまちがえた。(マークシートの欄がずれていた、など)
- 14：マークシートでダブルマークをした。(消し方が不十分だった場合を含む)
- 15：正答だったが採点ミスで誤答とされた。
- 0：その他(具体的に記入)

2. 3 ヒューマンエラーとの対応

2. 2の誤答理由は時間、問題理解、リスニング、紙面構成の順となっている。受験者が解答分析シートを記入する上で、なぜ誤答となったのかを考える際に要素となるものを抽出し、適切な選択に導くよう独自に項目化したものである。

誤答理由の項目1および項目2は解答時間に関するものである。表-1におけるヒューマンエラーの分類では(4)「能力不足」にあたるが、選択数が極度に多い場合には(1)「無理な相談」に該当し、テストの設計上のエラーであるといえる。つまり学生の能力不足以外に、テスト作成上の制限時間や出題量設定に問題があったとも考えられる。

項目4から項目8は学力に関するものである。表-1との対応では項目3, 4, 5, 8が(4)「能力不足」または(5)「知識不足」に該当する。項目6と7は(2)の「錯誤」に該当する。

項目9および項目10はリスニング問題に特化した誤答理由である。表-1では(4)「能力不足」(5)「知識不足」に加えて、音声物理的に聞きづらい場合には(1)「無理な相談」に当てはまることも考えられる。テストの設計あるいは実施の際にエラーが生じている可能性がある。

項目11から項目14はテストの書式や印刷に関するものである。項目11は(1)「無理な相談」に該当する、テストのインタフェース上のエラーである。項目12, 13, 14は(2)「錯誤」や(3)「失念」あるいは(6)「違反」に該当する。

また項目15は採点上のミスを検討したものであり、表-1の分類には対応しない。採点ミスは受験者のエラーではなく、テストマネジメント上のエラーである。項目15以外にも表-1の(1)が適用される場合には、同様にマネジメント上のエラーが生じており、テストを作成および実行する上での改善が求められているものである。

これらの理由に当てはまらない場合には項目0として具体的な理由を記入させたが、記入内容や聞き取り調査から上記の項目1から15の中で分類しなおすことができたため、今回の調査では選択数が0となっている。

2. 4 結果の視覚化

調査結果は受験者の正誤状況をテストの小問別に分析するS-P表(Student-Problem Score Table)を拡張し、誤答理由の項目を加えた形での集計を行なった。表-2において通常のS-P表と今回集計に用い

表-2 S-P表 (Student-Problem Score Table) とS-P-R表 (Student-Problem Score-Reason of Error Table) との比較

表-2-1 通常のS-P表

問題 学生	1	2	3	4	5	...	誤答数
○○ ○	1	1	0	0	1	...	2
□ □□	1	0	0	1	0	...	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
正答者数	45	30	17	32	38	...	

表-2-2 誤答理由を加えたS-P-R表

問題 学生	1	2	3	4	5	...	誤答数	
○○ ○			7	4		...	2	
□ □□		5	7		6	...	3	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
誤答の理由	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
	4	1	2	0	5	0	...	5
	5	3	8	10	7	2	...	47
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
計	5	20	33	18	12	...	87	
正答者数	45	30	17	32	38	...		

- ※1 両表はテスト受験者を50人とした作成例である。
- ※2 S-P-R表では正答の場合には空欄とし、誤答の場合には「誤答の理由」として2.2による項目の番号を記入する。
- ※3 S-P-R表では誤答理由別の誤答数集計が可能である。

(作表は筆者による)

得点 (人数 [前期中間, 前期期末])

選択数 (総数)

件数の指標

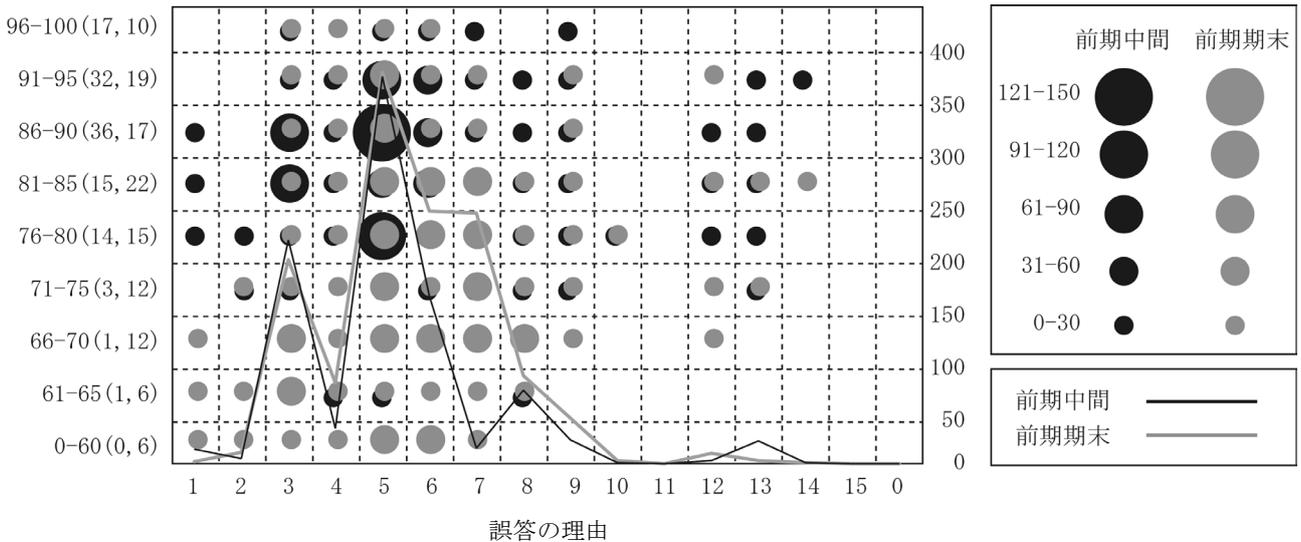


図-3 定期試験における得点と誤答理由の相関図 (Student-Score and Reason of Error Diagram)

た S-P-R 表 (Student-Problem Score-Reason of Error Table) を比較している。

この調査では調査対象者の属する得点域によって誤答理由の傾向にどのような差異がみられるのかを分析するために、誤答理由の選択数を視覚化することを試みている。

視覚化には横軸に誤答理由の項目番号、縦軸に得点域を設定した変則的なバブルチャートを使用し、それぞれのバブルの大きさによって選択数の大小を示す形式である得点と誤答理由の相関図 (以下SRチャート) として図-3の形に集約した。また、今回は

2つの定期試験を比較するため表のプロット域をグリッド化し、それぞれの定期試験でバブルの色を変えたものをずらして重ねることで一つの図に2つの結果を比較できる形で共存させている。図中の折れ線グラフは誤答理由各項目の選択総数を定期試験別に示している。

3. 調査結果

今回の調査では誤答の大まかな傾向と得点間での

差異を検証するために、各定期試験の得点によって学生を5点毎9段階に分け、それぞれの得点域における誤答の理由を項目別に集計した。S-Rチャート(図-3)横軸の1～15および0は、2.2で言及した誤答の理由の番号に対応している。

今回調査に利用した2つの定期試験の問題構成は語彙知識を問う問題、読解問題、語順整序による作文問題、リスニング問題と共通している。総問題数は前期中間試験が65、前期期末試験が66であり、前期期末試験でリスニング問題が1問多くなっている。調査結果のグラフ化にあたり、総問題数の差に関する調整は行っていない。これはこのS-Rチャートによる視覚化が実験的な試みであり、総問題数の差がグラフ上にどのように表れるのかを観察することを意図していることによる。

バブルの大小が示す選択数は、学習者の選択した小問の誤答一問一問に対する誤答理由の項目別総和であり、仮に総問題数65問のテストで全問不正解し、その理由が項目「5」に該当する場合、その学習者の誤答理由「5」の選択数は65になり、得点0で誤答理由「5」の区域に選択数65の大きさのバブルが表示されることになる。誤答理由が多岐に渡る場合にはより小さいバブルがいくつかの誤答理由の区域に表示されることになる。すなわち、バブル分布の度合いは誤答理由の多様性を表し、バブルの大小は誤答理由として選択される頻度を表している。従って、このS-Rチャートからそれぞれの得点域の学習者にどのような誤答の傾向があるのかを分析することが可能となっている。

なお、今回の調査では誤答理由を選択する際に一つの誤答に対して一つの理由項目を選択させている。試験において解答時間が足りない中で取り組んだ結果、解答を急ぐあまり問題を飛ばした、あるいは指示をよく読まず答え方を間違えた、という場合には誤答理由の項目2, 7, 12が当てはまる複合的な事例となるが、今回の調査では受験者が誤答の主たる原因と思われる項目を一つだけ選択する形を採った。複合的な誤答をどのように分析するのかについては、今後の調査および分析時における課題としたい。

3. 1 誤答理由の分布

図-3のS-Rチャートには前期中間試験と前期期末試験の2つの調査結果が重ねて表示されているが、前期中間試験のバブルは前期期末試験のバブルと比較すると分布が局所的である。誤答理由の項目3と5が特に多く、これに対して前期期末試験のバブル分布は広範囲に渡っている。項目11, 15, 0は空域

となっているが、これは両試験において印刷上の不具合や採点上のミスがなかったことを示している。

3. 2 試験における得点と誤答理由の相関

得点の高い受験者は、当然誤答数が少ない。しかしこのバブルチャートでは低い得点域のバブルが大きくなるということはない。これはバブルの大きさが得点域に属する人数の多少も反映しているためである。前期中間試験では得点86-90の層で誤答理由に項目5を選択した人が多い、と見ることができる。チャートで特異なバブルが表れている部分に着目して解答分析シートの集計結果を参照することで、どの設問でどのような間違いが多かったのかを参照し、指導に利用することが可能となっている。

4. 考察

得点に相関する要素の一つとして、受験者が解答に要する時間を上げることができる。英文を読解する速度の速い学生は高い得点域に属することが多く、このことは図-3における誤答理由の項目1や2のバブルの大きさに裏付けられている。91-100点の得点域に属する受験者は両試験に共通して誤答理由に1および2を選択したものはなく、区域は空欄となっている。これに対して相対的に低い得点域すなわち前期中間試験では76-85点の得点域に、前期期末試験では0-65点の得点域に属する受験者が、誤答理由の項目1や2を選択している。

4. 1 大きなバブル

図-3に特徴的なものの1つとして、前期中間試験の得点86-90の層で誤答理由に項目5の区域のバブルが大きい、という点が挙げられる。そこで前期中間試験のどの問題で誤答理由に項目5を選択しているのかを確認すると、選択数146のうち44が次の問題に集中している。

(問) 下線部が次の日本語の意味になるように、()内の語句を並べ替えなさい。ただし、()内では文頭に來るべき語も小文字で示してあります。

「ニュージーランドの人たちは、彼らのキウイに生きていて欲しいのです。」

(kiwis / live / New Zealanders / their / to / want) .

この問題の正答は「New Zealanders want their kiwis to live.」であり、受験者はwantに関する構文want to ～の形式が既習事項として念頭にあったことから誤答したものと推察できる。つまり、得点域が比較的高く、英語に関する学力がある程度高い学習者に多くみられる誤答例であると分析される。

また、図-3で前期中間試験と前期期末試験のバブルを比較すると、前期期末試験のバブルは広範囲に分布している。特に、前期中間試験にはない得点層0-85に渡る区域で項目7にバブルが発生している。そこで前期期末試験における解答分析シートの集計を参照すると、受験者全体の誤答理由としても項目5に次いで項目7が多く選択されており、それらは次の問題に集中している。

(問) ①～⑥までは英語で答え、⑦～⑩までは日本語で答えなさい。ただし、⑦は漢字一文字で答えなさい。

この問題では解答用紙に日本語で作成された表を、英文の内容から判断して適切と思われる語句を記入して完成させるものである。表記が日本語であること、⑦～⑩は日本語で答えるよう指示があることから、問題文で指示されているにも関わらず①～⑥に関しても日本語で解答して誤答となった者が多かった。これは表-1に分類される(2)「錯誤」によるエラーである。テスト時に受験者は緊張や失敗に対する不安等のストレス下に置かれ、このような技量不足や知識不足によるものではないエラーを起こす頻度が通常時よりも高くなる。

誤答理由を選択する際、自己の答案に対する分析的考察を嫌って安易に、比較のおおまかな分類である項目5を選んだ学生も多いと考えることもできるが、この問題は誤答分析という行為や解答分析シートに対する慣れ・不慣れを考慮した考察を行なう必要を示唆している。この慣れ・不慣れは分析的考察能力の習得に置き換えて考える場合には肯定的に評価することが考えられる。二度目の分析となる期末試験での調査ではバブルが中間試験よりも広範囲に分散していることから、調査への慣れが誤答の項目選択にどの程度影響するのかを検証することも今後の課題である。

4. 2 根拠のある誤答

ここで新たに考察の着眼点としたいのは、受験者が「合っている」と考えて解答したものが誤答になる場合である。受験者はどこから正答の確証を得ているのだろうか。

このような誤答の中で、最も致命的なのは問題の読み違いである。この場合の致命的とは、受験者が論理的裏付けを持って解答しているため、そもそもの問題理解の段階での誤りに気づかず、試験時間中に誤りを正すことができる可能性が低いことを意味する。

問題を解くという行動の過程で、ゴールとして設定する解くべき課題を間違えている場合、解き終えた段階でゴールを達成しているため、テスト時間中に解答の確認をしても誤っていることに気づくことができない。認知心理学者であるDonald A. Normanはヒューマンエラーをスリップ(slip)とミステイク(mistake)という2つのカテゴリーに分類した⁴⁾。行為に対して適切なゴールを形成できたのに実行するときにタイミングを間違えたり、思っていたものと違うものを動かしたり、やろうと思っていたことをやらなかったりして失敗する場合等に起こる間違いはスリップに分類される。これに対して間違ったゴールを立ててしまった場合に起こる間違いはミステイクに分類される。Normanはスリップよりもミステイクが大きな事態になることに言及し、見つけ出すことの困難さを指摘している。

4. 3 「勘で答えた」

受験者が行なった解答が、全てなんらかの根拠に基づく場合であるとは限らない。中には勘で答えるものもある。この「勘で答える」というのは、一体どういう状況なのだろうか。

(1) ある程度理解した上での勘

問題の意図を理解し、おおまかな解答の方向性は掴めているものの、選択問題などで1つに絞り込むことができず、確信の持てない状態で解答を行なう場合が、この「ある程度理解した上での勘」に分類される。知識不足に由来し、誤答理由の項目5に該当する。

(2) 根拠のない勘

問題を読んでみたものの全く解答がわからず、選択肢が提示されている場合に起こるのが「根拠のない勘」に基づいて行なった解答による誤答である。記号式の選択問題や語群から選んで解答する問題では、この勘に基づいた解答が正答となる場合もあるので、受験者の学習到達度をより正確に把握するためには、解答に至る過程を正答においても分析する必要がある。

いずれにしても「勘で答えた」という場合には誤答の原因に知識不足があり、上記(1)と(2)の違いは知識の差に由来しており、誤答理由としては

同様に項目5が適用されるものである。このような受験者の行動が、調査結果の中で誤答理由の項目5の選択数を多くする要因の一つとなっている。

4. 4 ケアレスミスとは

「ケアレスミスで満点を逃す」という表現がよくされるが、このケアレスミス、つまり注意不足による間違いは、どのように分類されるのだろうか。

(1) 書き忘れ

英作文を行なう設問で、三人称単数の主語に対する動詞の現在形にsをつけることを怠った場合、これは「書き忘れ」といえるのだろうか。この場合には学習事項を知識として理解はしているが習熟していない事例である可能性があり、学習者に習熟させる、学習内容を定着させることが指導の過程に求められていると考えられる。

これとは別に例えば語順整序問題において並べ替えるべき語句を欠いた誤答の場合には、正確な構文を習得していない知識不足に起因する他に、次の「し忘れ」に分類される、解答欄への転記の際の書き忘れに起因することも考えられる。

(2) し忘れ

書き忘れよりもダメージが大きいのは「し忘れ」である。ページの最後や大問の末尾にあるにある問題を解き忘れる、一旦取り組んでみたが難しかったので後回しにしてそのまま忘れるといった場合もあれば、問題用紙上に解答したものを解答用紙に転記し忘れる場合も考えられる。極端な場合1ページ、あるいは隣接する数ページを「し忘れ」るため、その部分は解答することなく誤答となってしまう。これは正確な学習到達度の測定に支障があり、受験者に対する不利益が大きい。テスト上の指示や冒頭での総問題数提示、ページ番号付与などテスト作成上の配慮によりこのエラーの発生率を下げるようにしたいものである。

問題が難しく解けずに飛ばす、というのは「技量不足」「知識不足」の側面もあり、設定解答時間によっては「無理な相談」に起因することもある。これに「失念」が複合し、また問題番号順に解答しないという「違反」も併発した誤答となる場合も考えられる。これをケアレスミスと単純に捉えていてはエラーは回避されない。

誤答が「失念」に該当する場合であればストレス下であっても平常時の判断力を失わないよう訓練することで改善することができるが、同じような間違い方を繰り返す場合には、原因が「失念」ではなく認識の誤りや思い違いである「錯誤」であると考え

られる。間違いを指摘された時は訂正することができるが、時間の経過と共に自分の思い込みが強さを取り戻し、再度同じ間違いを起こす場合もある。ケアレスミスは小さな間違いであることが多く、減点の度合いも低いことから習慣化すると、改善のためには意識的な働きかけが必要となる。各自が自分の誤答の原因を的確に把握し、一方でテスト実施者は小さな間違いを見逃さず、学習内容の根本的な理解の部分に起因することも念頭におき確認させることが求められる。

4. 5 テストが悪い

これまでの誤答分析の中で、表-1(1)「無理な相談」に関連付けられた項目がある場合には、テスト本体や実施方法など、言い換えればテストのシステムやインタフェース、マネジメントに改善の必要がある。実際のペーパーテストでは事前にユーザーテストを実施することは困難であるが、あるテストで生じた不具合を事例として検証し、別のテストでの改善に利用することは可能である。このような取り組みを継続することで事例を集積し、テスト作成時のガイドラインとなるような指標として集約することもできるのではないだろうか。

5. 結論と今後の課題

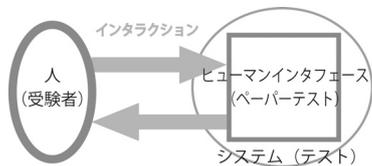
今回の調査では、あるテストに対してどのような誤答があるのかを分析するための手法を提案した。続く段階として求められるのは、誤答の理由に対応した指導方法の改善である。学習上の問題に起因する誤答の分析と改善については先行研究も多く、テストされる内容に応じた教科や科目での専門的な議論が行なわれており、ここでは誤答原因の分析と対策検討とを連携させるための手段を提供した形になる。その際には出題の観点別に分類した形で誤答分析を行なう等の集計上の工夫を行うことで、より詳しい誤答傾向の解析と指導方法の検討に役立てられるのではないだろうか。

しかし受験者の行動上の問題に起因する誤答や、テスト実施者に起因する誤答に関する改善を考える場合には、これまでに議論される機会の少なかったテスト・デザインについて検討する必要がある。テスト・デザインとは視覚上の書式や配置だけでなく、問題設定や実施形式等を総合的に検討したテスト設計を行なうことをいう。そのためにはペーパーテストは人間とのインタラクションという観点でとらえ

なおした場合にどのように位置付けられるのかを考えた。

5. 1 インタフェースとしてのペーパーテスト

ペーパーテストは受験者の学習到達度や習熟度を測定するための「テスト」という機構と「人間」との境界面にあり、人とシステムのインタラクション（相互作用）を司るインタフェースである。黒川隆夫はヒューマンインタフェースを人と道具、機械、システムとのインタラクションに必要なすべての仕組みと定義している⁵⁾。この定義をペーパーテストに適用し、「テスト」と「人間」の間でどのようなインタラクションがあるのかを考えたい。図-4は受験者とテストの間でどのような相互作用があるのかをモデル化したものである。



(作図は黒川の原因を筆者が改変した)

図-4 ヒューマンインタフェースモデル
(受験者とペーパーテストの場合)

受験者がテストとして行なうのは、ペーパーテストの紙面から問題を読み取り、解答を記入するという作業である。これに対してテストは受験者の解答から学習到達度を測定する。結果は得点として受験者にフィードバックされる。一般的な学校教育においてテストのシステムは自動化されている場合は少なく、教員というもう一つの人的要因がこの相互作用に関係することになる。いわば教員も、別の側面からみたペーパーテストのユーザーなのである。

5. 2 ペーパーテストのユーザーとは

学習者の学習効果を測定する際に使用される、紙媒体による一般的な手段であるペーパーテストは、教育の場だけでなく専門技術に関する資格試験など様々な場面で利用されている。しかしペーパーテストを作成する上で、ユーザビリティを踏まえたデザインが行なわれているかどうかは疑問である。テストでの誤答は、受験者ではなくテストのデザインに問題がある場合も考えられる。

日本テスト学会は、テストに関わる者として研究者、開発者（設計・開発する者）、領布者（提供・配布する者）、実施者（実施・採点する者）、利用

者（結果を利用する者）、管理者（結果の情報管理に携わる者）、受験者を挙げている⁶⁾。そしてこれらすべてがペーパーテストのユーザーに該当する。テストの機能を最大限に利用する上で、ユーザビリティという観点から再検討するのは、これまでに行なわれていない試みではないだろうか。

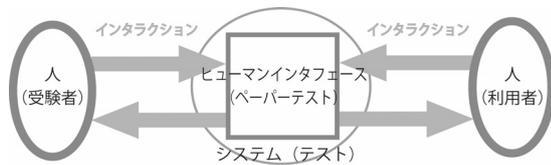
5. 3 ペーパーテストのユーザビリティ

ヒューマンインタフェースとしてとらえたペーパーテストにはユーザビリティを検討する余地がある。ユーザビリティの特性としてJakob Nielsenによれば学習しやすさ、効率性、記憶しやすさ、エラー発生率、主観的満足度の5つが定義される⁷⁾。この特性をペーパーテスト受験者の場合に当てはめると、テストは事前に特別な説明がなくても、受験者が試験開始と同時に解答を始められるよう、テストのやり方に学習の必要がないようにすること、テスト形式を標準化して効率よく解答できるようにすること、しばらく受験しなくても再び受験する際にとどこおりなく取り組むことができること、テスト作成上の工夫によりエラー発生率を低くし、受験者が試験時間内にエラーに気づいて修正できるような配慮がなされていること、問題構成や問題数、出題方法、難易度等の工夫により、学習到達度の測定に妥当で受験者がやりがいや達成感を感じられるようなテスト作りが求められる。今回の調査以前にある高等学校の生徒から「解答欄が狭くて書く気にならなかった」という意見を聞いたことがある。生徒が想定した解答の分量に対して与えられた解答欄が小さかったことに原因があったと考えられるが、これはテスト作成者による配慮不足と結論付けるのではなく、テスト作成者のレイアウト上または採点上の都合に合わせた結果である可能性を考慮する必要がある。すなわち受験者の解答しやすさというユーザビリティよりも、教員の作りやすさ、採点しやすさ、配りやすさといったユーザビリティが優先された結果である。テストの「解きやすさ」を左右するのは問題の難易度だけではなく、見やすさ、読みやすさ、書きやすさ、聞き取りやすさ等、受験者のユーザビリティに関わる要素も大きく関与しているのである。

一方ペーパーテストの作成、実施、採点、結果利用に携わる教員を想定したユーザビリティを考える場合には、テストの作りやすさや採点のしやすさ、結果の利用しやすさ等、「使いやすさ」が要求される。教員の側からのユーザビリティを向上させるための具体的な取り組みとしては、デザインの標準化による編集作業の効率化や設問のデータベース化、

採点や統計処理の効率化のための工夫などが考えられる。

このように「テスト」という一つのインタフェースに対して、異なる方向からのアプローチが生じている場合には、それぞれのユーザビリティに矛盾が起こり、片方には都合のよいシステムがもう片方には使いにくいシステムとなる事態を引き起こす可能性もある。テストは受験者が自分の学力を知る機能と、出題者が受験者の学力を測定するという多面的な機構を持つシステムであり、図-5にモデル化されるように、異なる特性を持ったユーザによるインタラクションがテスト紙という一つのインタフェースを媒介して行なわれている。



(作図は黒川の原因を筆者が改変した)

図-5 ヒューマンインタフェースモデル
(受験者，利用者，ペーパーテストの場合)

テストによっては作成者と採点者、利用者が異なり、それぞれにユーザビリティを考慮する必要がある場合も考えられる。その場合には図-4のモデルにさらに別のユーザが加えられることになる。そこでは扱う「人間」のユーザ特性により求められるインタフェースデザインも異なってくることになる。唯一で最高のインタフェースを追求するのではなく、それぞれのユーザに発生するユーザビリティのバランスをとり、最適なインタフェースとなるようデザインを行なうことが必要となるのである。

5. 4 提案と課題

今回の分析と考察に用いた手法を英語や語学に関する科目だけでなく、広く一般に適用することは可能だろうか。

テストの誤答分析は教科を問わず実施することが可能であり、S-P表および今回用いたS-P-R表の適用も可能である。S-P-R表における「誤答の理由」を教科やテスト内容に応じて変更することで、どのようなテストに対しても応用することができる。今回は15項目にその他1項目を加えた16項目での分析を行なったが、リスニング試験がなければリスニングに関する2項目を削除することができる。項目設定が妥当であるかどうかに関する検証も今後の課題で

ある。

エラーをなくすための方策には予防とセーフティネットがある。予防によりエラー自体が起こる確率を下げ、防ぎきれずに発生したエラーをセーフティネットで救済する。今回の分析に用いた解答分析シートによるセルフチェックは、採点者によるミスを救済するためのセーフティネットとしても機能している。エラーをなくすためには、よくあるエラーを知り、防止のための対策を立てることである。そのための方法として、この研究では誤答理由に注目した誤答分析の手法を提案した。ミスをなくすデザインを考えることは、ヒューマンエラーの予防と削減のための工夫を行なうことに通じる。学校教育の過程に恒常的に存在し、当たり前のようにとらえられているペーパーテストについてカリキュラム上の位置付けだけでなく、受験者と利用者のユーザビリティを考慮した設計を行なうことにも、次の一步への手掛かりがあるのではないだろうか。

参考文献

- 1) 小松原明哲：ヒューマンエラー，丸善株式会社，p. 17，2003.
- 2) 恒田直巳監修，小篠敏明編集：英語の誤答分析，大修館書店，pp. 2-19，1983.
- 3) 基準準拠評価と集団準拠評価については，R. Gagnéらの定義による。
Gagné, Robert M., Wage, Walter W., Golas, Katharine C. and Keller, John M.: *Principles of Instructional Design*, 5th ed., California: Wadsworth, Thomson Learning, Inc, 2005.
R. ガニエ他：インストラクショナルデザインの原理，北大路書房，p. 301，2007.
- 4) Norman, Donald A.: *The Psychology of Everyday Things*, Basic Books Inc., 1988.
D. ノーマン：誰のためのデザイン？；認知科学者のデザイン原論，新曜社，pp. 169-194，1990.
- 5) 黒川隆夫：ヒューマンインタフェース，基礎の基礎，ヒューマンインタフェースシンポジウム論文集，DVD-ROM，2008.
- 6) 日本テスト学会編：テスト・スタンダード，金子書房，p. 20，2007.
- 7) Nielsen, Jakob: *Usability Engineering*, Academic Press, 1994.
J. ニールセン：ユーザビリティエンジニアリング原論，東京電気大学出版局，pp. 21-30，1999.

(2008. 9. 3 受付)