

## S-P 表に基づく理解度ポテンシャルの定義とその実証的考察

正 員 三輪 和久<sup>†</sup>      正 員 下村 勉<sup>††</sup>

### Definition of Understanding Potential Score Based on S-P Chart and Its Experimental Discussion

Kazuhiisa MIWA<sup>†</sup> and Tsutomu SHIMOMURA<sup>††</sup>, Members

あらまし 本論文では、テストの評価法として広く利用されている S-P 表に基づいて、新しく「理解度ポテンシャル」なる評価値を定義する。理解度ポテンシャルは、より上位の生徒のより簡単な問題に対する反応ほど大きな値をもつものとして定義される。実践データを用いて本評価値の特性を検討した結果、本評価値を用いることにより、特定の生徒の各問題に対する理解の程度（理解の定着度や理解の深さ）を、定量的に測定することができるようになることが示された。また、「理解度ポテンシャル」と「正答・誤答反応や自信あり・自信なし反応の出現割合」との関係や、「理解度ポテンシャル」と「ポストテストから保持テストへの各反応の遷移パターン」との関係を検討することにより、学習者の客観評価と主観評価に対するより深い知見を得ることができた。更に、既に確立されている学習者の客観評価と主観評価の統合評価法に本評価値を導入することにより、より正確な評価が可能になることを示した。

キーワード 教育評価, S-P 表, 統合評価法, 理解度ポテンシャル

### 1. ま え が き

学習者の理解の状況を、より正確に、より多角的に把握し、評価する方法論を確立することは、教育工学の最も重要なテーマの一つである。近年、教育の現場に情報処理装置としてのコンピュータが導入されるに従って、データ処理に費やされるコストは飛躍的に軽減され、学習者に対するより多くの情報を引き出すことができるさまざまなデータの処理方法が提案され、実際に使用されてきている<sup>(1)</sup>。

S-P 表に基づくテストデータ分析は、処理・分析の方法がわかりやすく簡単であること、各問題に対するそれぞれの生徒の反応の特徴が一目で把握でき、更に全体の傾向を視覚的に理解することができること等のすぐれた特徴をもち、これまでも実際の現場で幅広く使われてきている<sup>(2),(3)</sup>。最近では、回答所要時間を加味したり<sup>(4),(5)</sup>、誤り反応をタイプ別に分類したりすることによって<sup>(6)</sup>、S-P 表を更に有効に利用する試みも加えられてきている。

S-P 表における代表的な評価指数としては、各生徒や各問題ごとの反応の特異性を評価する注意係数 CP や、S-P 表全体の非等質性を評価する差異係数  $D^*$  等が挙げられる。本論文では更に、各生徒の各問題に対する一つ一つの反応に着目し、各反応の示す理解の度合（理解の深さや理解の定着度）を評価する指数として、新しく「理解度ポテンシャル」なる評価値を定義する。理解度ポテンシャルなる評価値の特徴を列挙すると以下のごとくなる。

(1) 学習者の特定の問題に対する理解の程度を従来よりも詳しく定量的に把握することができる。

(2) 学習者の客観評価（正・誤に関する評価）や主観評価（自信あり・なしに関する評価）に関する新たな知見を引き出すことができる。

(3) 特別なデータの測定を必要とせず、形式的評価値として定義されコンピュータを用いて簡単にデータ処理できることから、簡便で実用性の高い評価値であると言える。

(4) 他の評価法と組み合わせて用いることにより、更に精度の高い評価法を実現することができる。

以下では、2.にてこの理解度ポテンシャルを定義し、3.では実践データに基づいて、実際に理解度ポテンシャルの評価値としての特徴を明らかにする。更に4.で

<sup>†</sup> 名古屋大学情報処理教育センター, 名古屋市  
Education Center for Information Processing, Nagoya University, Nagoya-shi, 464-01 Japan

<sup>††</sup> 三重大学教育学部, 津市  
Faculty of Education, Mie University, Tsu-shi, 514 Japan

は、他の評価法の一つである統合評価法に本評価値を導入することによって、より有用な評価法が実現されることを例示する。

## 2. S-P 表と理解度ポテンシャル

S-P 表に基づいて、新しく「理解度ポテンシャル」なる評価値を定義する。

### 2.1 S-P 表

$N$  人の生徒に対して、 $n$  問の問題からなるテストを実施した場合を考えよう。テスト結果は、 $n$ (問)× $N$ (人)の表に表すことができる。表を構成する要素は、2 値評価の場合には 1 (正答) と 0 (誤答) である。

ここで、この表を、横方向( $n$  問の問題が並んでいる)の左から正答率の高い(簡単な)問題の順に、また縦方向( $N$  人の生徒が並んでいる)の上から得点の高い(上位の)生徒の順に、それぞれ並べ換える。更に、並べ換えられた表の中に、表の左から各生徒の得点だけ右にずらして、また表の上から各問題の正答数だけ下にずらして、2 本の区切り線を入れる。前者を S 曲線(student curve)、後者を P 曲線(problem curve)と呼ぶ。

以上のようにしてできた表は、S-P 表(S-P chart)と呼ばれる。図 1 に、生徒数 16 人、問題数 10 問の場合の S-P 表の例を示す。S-P 表は、S 曲線と P 曲線の形と相互の関係を見ることによって、テスト問題や生徒の全体的特徴を直感的に把握することができること、個々の生徒や個々の問題の反応パターンを見ることによって、異質な反応を示す生徒や問題を簡単に発

見することができること等のすぐれた特徴を有している。

### 2.2 理解度ポテンシャルの定義

S-P 表においては、左上の領域には、上位の生徒の簡単な問題に対する反応が、逆に右下の領域には、下位の生徒の難しい問題に対する反応が現れていると考えられる。

ここで、各生徒の各問題に対する理解の程度は、「その生徒がクラス全体に対して占める位置」(上位・下位の程度)と「問題の簡単さ」との関係によって規定できると仮定し(例えば、より上位の生徒のより簡単な問題に対する反応はより深くより定着度の高い理解を伴っているというように考え)、そのような理解度を表す指標として、「理解度ポテンシャル」(以下、単に「ポテンシャル」と略す)なる評価値を以下のように定義する。

すなわち、S-P 表において、上から  $i$  番目の生徒  $S_i$  の左から  $j$  番目の問題  $P_j$  に対する反応  $x(i, j)$  のもつポテンシャル  $v(i, j)$  を、以下のように定義する。ここで、 $x(i, j)$  は二つの値(0 (誤答)と 1 (正答))のうちのいずれかである。

$$v(i, j) = \frac{(x(\cdot, j) - x(i, \cdot))}{N} \quad : 1 \leq x(i, \cdot) \leq n \text{ のとき}$$

$$= \frac{(x(\cdot, j) - N)}{N} \quad : x(i, \cdot) = 0 \text{ のとき}$$

上式において、 $x(\cdot, j)$  は問題  $P_j$  の正答数、 $x(i, \cdot)$  は生徒  $S_i$  の得点であり、具体的には以下で与えられる。

$$x(\cdot, j) = \sum_{i=1}^N x(i, j)$$

$$x(i, \cdot) = \sum_{j=1}^n x(i, j)$$

図 2 は、図 1 の S-P 表に基づいて計算された理解度ポテンシャルの例である。但し、図中の表記は、 $100 \times v(i, j)$  の値が用いてある。

図 1 と図 2 において、反応  $x(10, 2)$  [□印参照] と反応  $x(10, 6)$  [○印参照] は同じ「1」反応であるが、前者は後者よりもより簡単な問題に対する反応であるので理解度ポテンシャルが大きな値となっている。同様に、反応  $x(3, 6)$  [△印参照] と反応  $x(10, 6)$  [○印参照] は、前者は後者よりも上位の生徒による反応であるので理解度ポテンシャルが大きな値となっている。

### 2.3 理解度ポテンシャルの性質

以下には、2.2 で定義された理解度ポテンシャルの性質について考える。

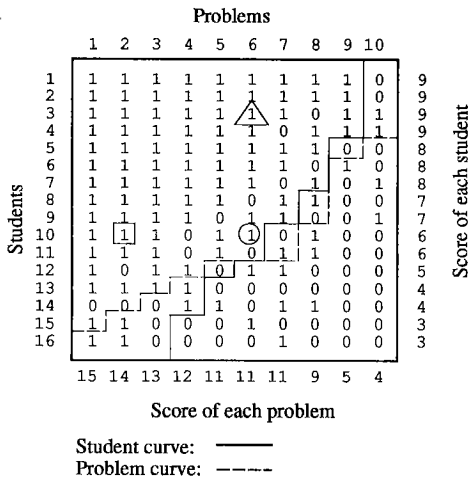


図 1 S-P 表の一例  
Fig. 1 An example of S-P chart.

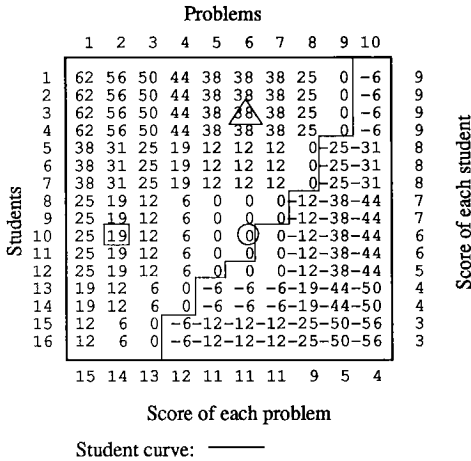


図2 理解度ポテンシャルの分布状況  
Fig.2 The distribution of understanding potential score.

表1 客観・主観評価に基づく5種類の反応

	自信あり	自信なし
正答	R1	R2
誤答	R3	R4
無答	R5	

### 3.1 客観評価と主観評価

理解度ポテンシャルの値の挙動を検討するために、ここでは一般に用いられている「客観評価」に加えて、「主観評価」についても取り上げる。ここでいう「客観評価」とは、教師により行われる「正答」若しくは「誤答」に関する評価のことを示し、主観評価とは、生徒自身により行われる自分の解答に対する「自信あり」若しくは「自信なし」に関する評価のことをいう。実際のテスト場面における主観評価の測定は、通常の解答欄の横に主観評価のための解答欄を用意し、自信ありの場合には○を、自信なしの場合には×を生徒に記入させるといった方法が用いられることが多い。

本論文では、文献(7)に従って、各生徒の各問題に対する反応を、表1の5種類に分類する。

### 3.2 実践データ

中学数学1年の2種類のテスト(テストA:「数と集合」、テストB:「文字式」)を用意した。両テストともに基礎的学力を問うものであり、一問一答式の問題から構成されている。各テストについて、それぞれ4クラス(SS1クラス~SS4クラス、およびM1クラス~M4クラス)の協力を得た。更に、各テストそれぞれについて、授業終了後に行われるポストテスト(Post Test)と、ポストテスト後約1カ月後に行われる保持テスト(Ret Test)の2種類を実施した。保持テストを用いることで、ポストテスト実施時点の各生徒の各問題に対する理解の定着度を確認・検証することができる。なお、ポストテスト実施の1週間後、その結果が採点答案の返却と正答の提示という形で各生徒にフィードバックされている。

表2に、各テスト結果の一覧を示す。表2には、各テストに参加した生徒の数、平均点、S-P表に基づいて得られる差異係数 $D^*$ 、学習者の主観評価に関する評価値である過小評価の基準値 $I_s^*$ 、および過大評価の基準値 $I_o^*$ が示してある。表2は、すべてのテストについて、平均点、差異係数、過小および過大評価基準値ともに標準的であることを示しており、これらのテストの実施条件が一般的であったことを支持している。

## 3. 実践的検討

2.にて定義された理解度ポテンシャルの値の挙動を実践データに基づいて検討し、その定義の有効性を実証的に示す。

表 2 実践データの概略

Test	Problems	Class	Post or Ret	Students	Average	D*	I <sub>u</sub> *	I <sub>o</sub> *
Test A	20	SS1	Post	42	58.9	0.381	0.243	0.216
			Ret	40	59.8	0.339	0.339	0.162
	SS2	Post	43	53.6	0.356	0.196	0.258	
		Ret	43	63.5	0.311	0.193	0.225	
	SS3	Post	43	61.2	0.365	0.296	0.180	
		Ret	42	69.6	0.254	0.261	0.137	
	SS4	Post	45	61.4	0.347	0.265	0.206	
		Ret	44	62.6	0.339	0.348	0.153	
Test B	25	M1	Post	42	52.8	0.290	0.318	0.180
			Ret	37	58.6	0.299	0.257	0.198
	M2	Post	42	49.0	0.347	0.212	0.267	
		Ret	38	59.1	0.351	0.125	0.271	
	M3	Post	42	53.7	0.333	0.211	0.244	
		Ret	40	62.1	0.352	0.118	0.246	
	M4	Post	42	51.5	0.321	0.373	0.151	
		Ret	42	64.8	0.380	0.195	0.196	

以下の検討では、主に SS1 クラスから SS4 クラスに対して行われたテスト A の結果を中心に報告する。M1 クラスから M4 クラスに対して行われたテスト B の結果は補助データとして用いる。なお、後に具体的に述べられるように、両者のテスト結果は、基本的によく一致したものとなった。

### 3.3 ポストテスト結果に基づく検討

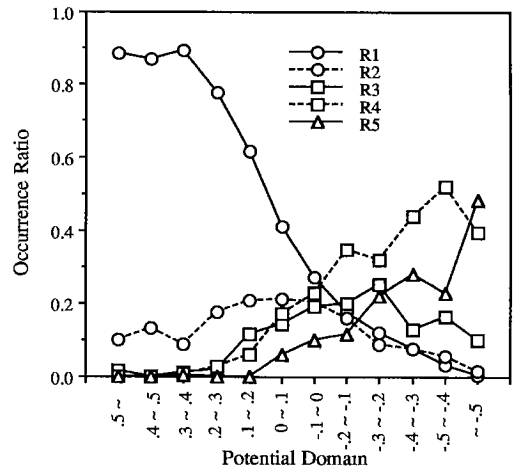
それでは最初に、ポストテストの結果に着目してみよう。

#### (1) 各 $R_i$ 反応の出現割合

特定のポテンシャルの範囲にある S-P 表上の領域を「ポテンシャル領域」と呼び、例えばポテンシャル領域  $0 \leq v(i, j) < 0.1$  のように表す。まず、特定のポテンシャル領域を占める各  $R_i$  反応の出現割合を見てみよう。

図 3 は、S-P 表を 0.1 のポテンシャルの幅で分割し、分割されたそれぞれのポテンシャル領域において、各  $R_i$  反応の出現割合を示したものである。例えば、 $R_1$  反応の出現割合は、“ $R_i$  ( $i=1$ ) 反応の個数 /  $R_i$  ( $i=1, 2, \dots, 5$ ) 反応の個数”によって与えられる。図 3 より、以下の傾向を読み取ることができる。

まず、 $R_1$  反応（正答・自信あり）の出現割合は、ポテンシャルが増大するにつれて単調に増加している（図 3 の—○—参照）。一方、同じ正答反応である  $R_2$  反応（正答・自信なし）の出現割合は、ポテンシャルが負の領域から正の領域に向かうにつれて増加傾向を示すが  $v(i, j)=0.1$  前後で極大値となり、それ以上のポテンシャル領域ではむしろ減少傾向を示している（図 3

図 3  $R_i$  反応の出現割合Fig. 3 Occurrence ratio of reaction  $R_i$ .

の---○---参照)。

次に、 $R_4$  反応（誤答・自信なし）の出現頻度は、ポテンシャルが減少するにつれて単調に増加している（図 3 の---○---参照）。一方、同じ誤答反応である  $R_3$  反応（誤答・自信あり）の出現頻度は、ポテンシャルが正の領域から負の領域に向かうにつれて増加傾向を示すが  $v(i, j)=-0.2$  前後で極大値となり、それ以下のポテンシャル領域ではむしろ減少傾向を示している（図 3 の—○—参照）。 $R_5$  反応（無答）の出現頻度は  $R_4$  反応に同様であった（図 3 の—△—参照）。

ここで、項目反応理論における多岐選択モデルに基づいて、自信あり正答に対応づけられる「確実な理解」

の程度に対する R1~R4 反応の出現分布を理論的に求めると、最も一般的には図 3 に酷似した分布となることが知られている<sup>(6)</sup>。このことは、「理解度ポテンシャル」を、「確実な理解である R1 反応ができる能力」に対応づけることが可能であることを示しており、理解度ポテンシャルの意味付けの一つとして興味深い。

(2) 客観評価・主観評価・一致反応ごとの出現割合

次に、以上の傾向を、客観評価、主観評価、客観評価と主観評価の一致反応の各項目ごとに再検討してみよう。R1 反応と R2 反応の合計は客観評価における「正答」反応に、R1 反応と R3 反応の合計は主観評価における「自信あり」反応に、そして R1 反応と R4 反応の合計は客観評価と主観評価の「一致」反応に対応する。図 4 は、そのようにして、正答反応(correct reaction)、自信あり反応 (confident reaction)、一致反応 (consistent reaction) の出現割合を、各ポテンシャル領域ごとに示したものである。ここで、正答反応の割合は、正答反応 (R1+R2) の個数/(正答反応 (R1+R2) の個数 + 誤答反応 (R3+R4+R5) の個数) で与えられるのに対して、自信あり反応の割合は、自信あり反応 (R1+R3) の個数/(自信あり反応 (R1+R3) の個数 + 自信なし反応 (R2+R4) の個数) によって計算されるべきであると考え、後者に関しては無答 (R5 反応) の個数を分母から除いて計算してある。一致反応の割合の算出についても、後者に同様である。

図 4 より、以下の傾向が読み取れる。

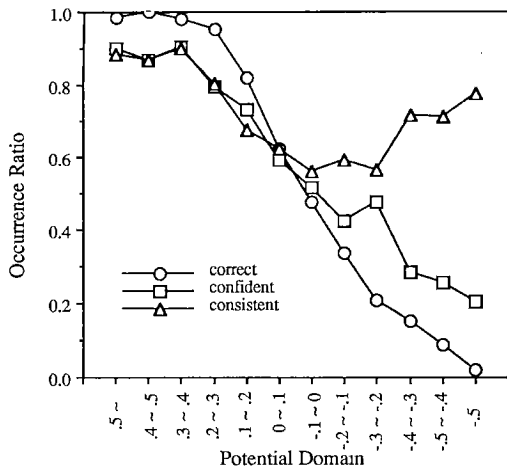


図 4 正答反応・自信あり反応・一致反応の出現割合  
Fig. 4 Occurrence ratio of correct reactions, confident reactions, and consistent reactions.

まず、正答反応の出現割合は、ポテンシャルが増加するにつれて単調に増加している。更にその割合は、 $v(i, j) > 0.2$  の領域ではほぼ 1 に、逆に  $v(i, j) < -0.5$  の領域ではほぼ 0 に達している (図 4 の—○—参照)。

次に、自信あり反応の出現割合も、ポテンシャルが増加するにつれて単調に増加している。但し、 $v(i, j)$  が非常に大きくなっても 1 には達することなく、 $v(i, j)$  が非常に小さくなっても 0 には達しない(図 4 の—□—参照)。

更に、一致反応の出現割合は、ポテンシャル  $v(i, j) = 0$  の前後で極小値を示し、 $v(i, j)$  の絶対値が増すにつれて徐々に増加する (図 4 の—△—参照)。これは、ポテンシャルが大きな領域では、正答反応が大部分を占めると同時に自信あり反応が増加して、R1 反応(正答・自信あり)が優勢になったこと、逆に、ポテンシャルが小さな領域では、誤答反応が大部分を占めると同時に自信なし反応が増加して、R4 反応(誤答・自信なし)が優勢になったことによるものと考えられる。

以上は、客観評価や主観評価におけるポジティブ反応(正答反応や自信あり反応)は、ポテンシャルに対して単調に増加するのに対して、主観評価の正確さ(客観評価と主観評価の一致度)は、ポテンシャルが 0 の近傍を中心にその絶対値が小さい領域においてはその精度が減衰することを意味しており、興味深い知見である。

なお、3.3 に記述された主データとしてのテスト A の結果が抱く傾向は、補助データとしてのテスト B の結果においても、同様に確認されている。

3.4 保持テスト結果に基づく検討

ポストテストにおける  $R_i$  反応から保持テストにおける  $R_j$  反応への遷移を観察することによって、ポストテスト時点における  $R_i$  反応の定着度を観察することができる。以下では、ポストテストの  $R_i$  反応から保持テストの  $R_j$  反応への遷移を、「 $R_{ij}$ 遷移」と呼ぶこととする。

(1) 客観評価の遷移

まずは、客観評価の遷移状況を眺めてみよう。

図 5 は、ポストテストにおける正答反応が保持テストにおいても維持される割合、およびポストテストにおける誤答反応が保持テストにおいて正答反応へ転移する割合を、ポストテストにおける各ポテンシャル領域ごとに示したものである。具体的には、前者は、「 $R_{ij}$  ( $i=1, 2; j=1, 2$ ) 遷移の個数/ $R_{ij}$  ( $i=1, 2; j=1, 2, 3, 4, 5$ ) 遷移の個数」により、後者は、「 $R_{ij}$  ( $i=3, 4, 5; j=$

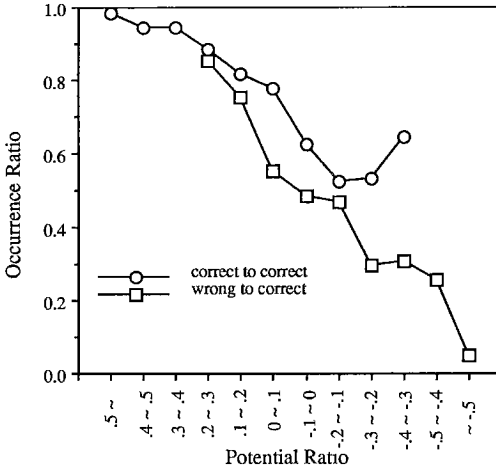


図5 客観評価における正答反応への遷移割合  
Fig.5 Transition ratio to correct reactions in objective estimation.

1, 2) 遷移の個数/ $R_{ij}$  ( $i=3, 4, 5; j=1, 2, 3, 4, 5$ ) 遷移の個数”により与えられる。

図5より、以下の知見が読み取れる。なお、以下の図5から図7では、各ポテンシャル領域における当該の資料数が不足する場合（具体的には、母資料数が20未満になる場合）には、グラフ中への値のプロットを省略してある。

まず、ポストテストの時点での正答反応は、どのポテンシャル領域にあっても50%以上の割合で保持テストにおいても正答が維持されている。特に、ポテンシャルが正の領域においては、ポテンシャルが増大するにつれて保持テストにおいて正答が維持される割合は増加する（図5の○—参照）。

次に、ポストテストの時点での誤答反応も同様に、ポテンシャルが増大するにつれて保持テストでの正答への転移の割合が高くなる。特に、ポテンシャルが正の部分に関しては、50%以上の割合で正答に転移している（図5の□—参照）。

更に、図5は、ポストテスト時点において誤答であった場合も、ポテンシャル  $v(i, j) \geq 0.2$  の領域では80%以上が正答へ転移することや、ポストテスト時点において正答であっても、 $v(i, j) < 0$  の領域では正答維持率が50%近くまで減少すること、すなわち50%近くが誤答へ転移することを示しており、ポストテスト時点における正答/誤答の区別は、保持テストにおいて正答が得られることの決定的条件ではないことを示唆している。加えて、図5において2本の曲線が互いに接

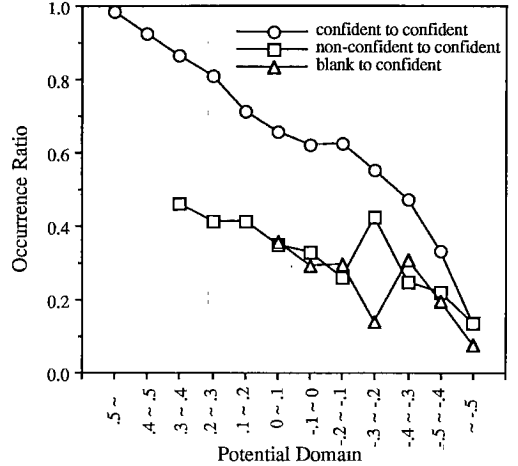


図6 主観評価における自信あり反応への遷移割合  
Fig.6 Transition ratio to confident reactions in subjective estimation.

近しながらポテンシャルの増大に対してともに顕著な増加傾向を示していることは、保持テストで正答を得るための条件としては、ポストテストにおける正誤の区別よりもむしろ、その反応のポテンシャルの値の方がより重要であることを支持するものである。

以上は、理解度ポテンシャルという評価値が、ポストテスト時点において保持テストでの正答への遷移を予測するための有効な指標になり得ることを示している。

## (2) 主観評価の遷移

次に、主観評価の遷移状況を眺めてみよう。

図6は、ポストテストにおける自信あり反応、自信なし反応、および無答反応が、保持テストにおいて自信あり反応に維持/転移する割合を、ポストテストにおける各ポテンシャルごとに示したものである。具体的には、上記の三つの値は、“ $R_{ij}$  ( $i=1, 3; j=1, 3$ ) 遷移の個数/ $R_{ij}$  ( $i=1, 3; j=1, 2, 3, 4, 5$ ) 遷移の個数”、“ $R_{ij}$  ( $i=2, 4; j=1, 3$ ) 遷移の個数/ $R_{ij}$  ( $i=2, 4; j=1, 2, 3, 4, 5$ ) 遷移の個数”、および“ $R_{ij}$  ( $i=5; j=1, 3$ ) 遷移の個数/ $R_{ij}$  ( $i=5; j=1, 2, 3, 4, 5$ ) 遷移の個数”により、それぞれ与えられる。

図6より、以下の傾向を見いだすことができる。

まず、ポストテストの時点での自信あり反応は、かなり広い範囲（ポテンシャル  $v(i, j) > -0.3$  の領域）にわたって、50%以上の割合で自信あり反応が維持されている。また、ポテンシャルが増大するにつれて自信あり反応が維持される割合は単調に増加している（図

6の一〇一参照)。

次に、ポストテストの時点での自信なし反応が保持テストにおいて自信あり反応に転移する割合は、すべてのポテンシャル領域において、50%以下である。また、ポテンシャルが減少するにつれて、保持テストでの自信あり反応への転移の割合はより低くなる(図6の一〇一参照)。

更に、図6の2本の曲線一〇一と一〇一〇がかなりはっきりと分離して存在していることは、保持テストにおいて自信あり反応に維持/転移する割合は、ポストテスト時点における自信あり/なし反応の別にかなり強く依存していることを示すものである。この主観評価の遷移に関する特徴は、客観評価の遷移に関する特徴(ポストテスト時点における正答/誤答の差異よりもその反応のポテンシャルの方がより重要であったこと)と対照的であり、客観評価と主観評価の定着度の違いについての興味深い知見である。

なお、以上の(1)客観評価の遷移および(2)主観評価の遷移における主データとしてのテストAの結果が抱く傾向は、補助データとしてのテストBの結果においても、同様に確認された。

(3) 一致反応の遷移

更に、客観評価と主観評価の一致反応の遷移について見てみよう。

図7は、ポストテストにおける客観評価と主観評価の一致反応、逆反応、および無答反応が、保持テストにおいて一致反応に維持/転移する割合を、ポストテ

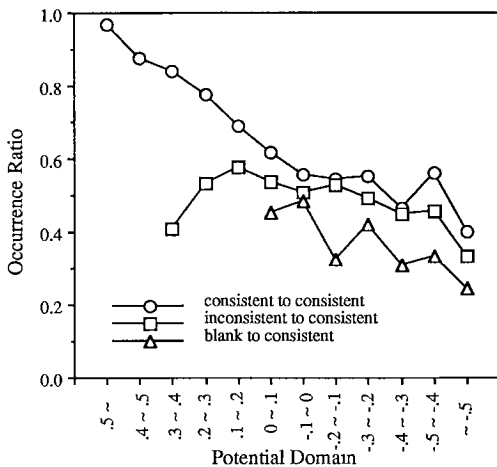


図7 客観・主観評価における一致反応への遷移割合  
Fig.7 Transition ratio to consistent reactions in objective-subjective estimation.

トにおける各ポテンシャルごとに示したものである。具体的には、上記の三つの値は、“ $R_{ij} (i=1, 4; j=1, 4)$  遷移の個数 /  $R_{ij} (i=1, 4; j=1, 2, 3, 4, 5)$  遷移の個数”、“ $R_{ij} (i=2, 3; j=1, 4)$  遷移の個数 /  $R_{ij} (i=2, 3; j=1, 2, 3, 4, 5)$  遷移の個数”、および “ $R_{ij} (i=5; j=1, 4)$  遷移の個数 /  $R_{ij} (i=5; j=1, 2, 3, 4, 5)$  遷移の個数” により、それぞれ与えられる。

図7より、以下の傾向を読み取ることができる(以下、図7の一〇一と一〇一〇参照)。

ポテンシャル  $v(i, j) > 0$  の領域では、ポストテストにおける一致反応が維持される割合は、逆反応が一致反応に転移する割合を大きく上回っている。すなわち、前者の割合がポテンシャルが増大するにつれて単調に増大するのに対して、後者の割合はポテンシャルの増大に対して増加する傾向は認められない。

一方、 $v(i, j) < 0$  の領域では、両者に明確な差は現れず、かつポテンシャルの増減に対する変化もそれほど顕著ではない。

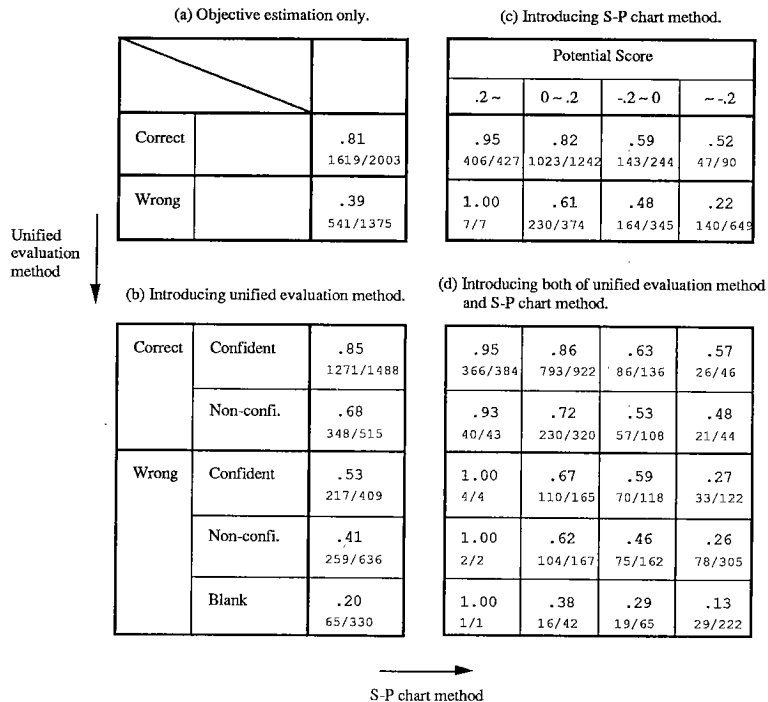
但し、補助データとしてのテストBの結果においては、ポテンシャルが増大するにつれて両者の割合はともに増加する傾向が認められた。よって、客観評価と主観評価の一致反応の遷移に関して安定した知見を得るためには、今後実践データを増やすなりして、更なる検討を重ねる必要がある。

4. S-P表と統合評価法の組合せによる評価

客観評価に加えて主観評価をも用いて評価を行う方法は、これまで下村らにより統合評価法として確立されてきている<sup>(7),(9)(10)</sup>。それによると、同じ正答(若しくは誤答)であっても、自信あり反応と自信なし反応の間にはその理解度の実質的な差異が存在することが確認されている。以下では、理解度ポテンシャルという評価値を導入することにより、更に理解度の差異を明確にすることができることを示す。これは、独立して発達してきたS-P表と統合評価法という二つの評価法を組み合わせることに用いることにより、更に有効な評価法を実現できることを示すものである。

以下、表3に従って具体的に見てゆくこととしよう。表3は四つの表から構成されている。表3(a)は客観評価だけにに基づく評価の状態を、表3(b)は統合評価法の導入によって、表3(c)はS-P表に基づく理解度ポテンシャルなる評価値の導入によって、更に表3(d)は二つの評価法をともに導入することによって

表 3 統合評価法と S-P 表の組合せによる評価



得られる評価の状態を、それぞれ示している。それぞれの表の各項目の上段には、ポストテストにおける各反応が保持テストにおいて正答に維持/転移した割合を示したものである。従って、この割合が大きいほど、正答においては安定した理解を、誤答においては軽度の不理解（もう一息で正答に達することができた等）を表すこととなる。加えて、下段には、各反応の全資料数と正答に維持/転移した資料数も示してある。

さて、統合評価法も S-P 表も利用しない状態（表 3(a)）では、ポストテストにおける各反応は正答と誤答の区別しか存在しない。この状態では、合計 3,378 個の全反応を、2,003 個の正答（正答維持率 81%）と 1,375 個の誤答（正答推移率 39%）に分けることしかできない。以下、前者の正答反応について見てゆく。

その状態から統合評価法を導入することにより（表 3(b)）、ポストテストにおける 2,003 個の正答反応を、1,488 個の自信あり反応（正答維持率 85%）と 515 個の自信なし反応（正答維持率 65%）に分類することが可能になる。更に、S-P 表に基づく理解度ポテンシャルなる評価値を導入することにより（表 3(d)）、1,488 個の正答・自信あり反応を正答維持率がそれぞれ 95%、86%、63%、57% となる四つのグループに、515 個の正

答・自信なし反応を正答維持率が 93%、72%、53%、48% となる四つのグループに、それぞれ分類することが可能になる。

以上は、正答・誤答の客観評価だけしか存在しない状態では分離できなかった異なる質・傾向を抱いた反応を、総合評価法、更には S-P 表に基づく評価法を組み合わせることで用いることによりその差異を顕在化させることに成功し、複数のカテゴリーに分類可能になることを表すものである。これは、複数の評価法を組み合わせることで用いることにより、より有効な評価法を実現できることを具体的に例示するものである。

### 5. む す び

本論文では、S-P 表に基づく理解度ポテンシャルなる評価値を新しく導入し、実践データに基づいて、評価値としての有用性、および客観評価や主観評価との関連性について検討した。そこで得られた主な知見をまとめると、以下のようになる。

(1) 理解度ポテンシャルという評価値を導入することにより、保持テストにおける正答への遷移の割合を予測することが可能になる。このことは、ポストテスト時点での学習者の理解の程度を定量的に把握でき



ることを示すものである。

(2) ポテンシャルが増大するにつれて、客観評価における正答反応、主観評価における自信あり反応の出現割合が増大する。但し、主観評価の正確さ(客観評価との一致度)に関しては、0を中心に絶対値が小さい領域ではその精度が減少し、ポテンシャルの絶対値が増大するにつれてその精度が増加する。

(3) ポストテストから保持テストへの反応の遷移(定着度)を規定する条件としては、客観評価に関しては、ポストテスト時点における反応自体(正答か誤答)よりもポテンシャルの値の方がより重要であるのに対して、主観評価に関しては、その反応自体(自信ありか自信なし)が重要になる。

本論文では更に、以上の知見に基づいて、統合評価法による評価に理解度ポテンシャルなる評価値を導入し、両者を組み合わせることによってより精度の高い評価が実現されることを具体的に示した。

今後の課題としては、以下のものが挙げられる。

(a) 既に明らかにされてきているS-P表の数理的構造に基づき、本評価値の理論的基礎を確立すること。また、他の科目・他の年齢を対象としたテストや等質性のよくないテストの結果を検討するなどして、実践例を増やすこと。これらを通して、理論と実証の両面から、本評価値の一般性をより確かなものにしていく必要がある。

(b) 本論文で得られた知見を実際の教育現場で活用するために、本評価値から得られる情報の教師や学習者へのフィードバックの方法を検討し、今後は実用的なデータ処理システムを開発することなどにも従事してゆきたい。

**謝辞** 本研究の機会を与えて下さった名古屋大学情報処理教育センター長鬼頭幸生教授、および同大学工学部杉江昇教授に感謝致します。また、常々御指導頂いている中部大学工学部織田守矢教授に感謝申し上げます。

## 文 献

- (1) 宇都宮敏男, 坂元 昂監修: “講座教育情報科学3—教育とデータ分析”, 第一法規出版(1988)。
- (2) 佐藤隆博: “S-P表の作成と解釈”, 明治図書(1975)。
- (3) 佐藤隆博: “授業設計と評価のデータ処理技法—ISM教材構造化法とS-P表の活用法”, 明治図書(1980)。
- (4) 永岡慶三, 呉 亜棟: “回答所要時間を加えた統合S-P表”, 信学論(A), **J73-A**, 8, pp. 1423-1430 (1990-08)。
- (5) 永岡慶三, 呉 亜棟: “回答所要時間を加えた統合S-P表における項目の難易度と弁別度について”, 信学論(A), **J75-A**, 2, pp. 414-421 (1992-02)。

- (6) 倉田政彦, 佐藤隆博: “観点別評価のS-P表における項目反応パターンの指数”, 信学論(A), **J71-A**, 4, pp. 1044-1053 (1988-04)。
- (7) 下村 勉, 織田守矢: “統合評価法の実践的検討とその応用”, 信学論(A), **J64-A**, 3, pp. 235-242 (1981)。
- (8) 永岡慶三, 植野真臣: “自信-正誤反応における項目応答理論の提案”, 信学論(A), **J75-A**, 2, pp. 407-413 (1992-02)。
- (9) 下村 勉, 織田守矢: “学習者の自己評価と客観評価との統合評価法”, 信学論(A), **J63-A**, 8, pp. 483-490 (1980-08)。
- (10) Shimomura T. and Oda M.: “Study on Experimental Application of Unified Evaluation Method in Educational Technology”, Trans. IECE of Japan, **E65**, 11, pp. 672-679 (Nov. 1982)。

(平成4年7月3日受付, 5年4月14日再受付)



三輪 和久

昭59名大・工・応物卒。平1同大大学院博士(情報)課程了。工博。平1より、同大情報処理教育センター助手。平3~4米国Carnegie Mellon University, Dept. of Psychology, visiting assistant professor。平5・10月より、同大大学院人間情報学研究所認知情報論講座助教授。認知科学, 人工知能, 教育工学の研究に従事。Cognitive Science Society, 人工知能学会, 認知科学会, 教育工学会各会員。



下村 勉

昭49名大・工・電気卒。昭54同大大学院博士(情報)課程了。工博。昭56愛知県コロニー研究員。昭59三重大学助教授(教育実践研究指導センター), 現在に至る。教育情報工学, 情報教育の研究に従事。共著書「概念形成と評価」, 「学習環境の構築」など。日本教育工学会, 日本科学教育学会, 行動計量学会, 日本ME学会各会員。