教師の予測の精度を解析する数理モデルの開発とその適用
- 見過ごされてきた学力・学習力を検出する実証的方法の提案－

植阪友理[1]・中川正宣[2]
[1]：東京大学教育学研究科学校教育高度化センター　[2]：東京工業大学社会理工学研究科
E-mail: [1]：y_uesaka@p.u-tokyo.ac.jp, [2]：nakagawa@nm.hum.titech.ac.jp

問題と目的

教育心理学や自己調整学習研究では，学習者が身につけるべき学力や学習スキルに関して様々な提案がなされている。しかし，こうした視点は，従来の学校教育の中では十分に取り入れられているわけではない。なぜならば，心理学的研究知見を踏まえて開発されたテスト，数学的学力・学習力診断テストCOMPASS（componential assessment, 市川ら，2009）によって，そうした学力・学習力が十分に身についていない実態が明らかになっているからである（植阪ら，2010)。つまり，これまでの研究知見に基づく学力・学習力は，学校現場では「見過ごされてきた力」である可能性が高い。

しかし，テスト得点の低さを示すだけでは，「見過ごされてきた力」であるということためには不十分であり，このことを実証的に示すのは容易ではない。この問題を解決する視点として，学習者の得点分布に関する教師の予測に着目することが考えられる。多くの教師が学習者の得点分布を正しく予測できていない（もしくは全体的に過大／過小評価する）課題が明らかとなれば，より実証的に「見過されてきた力」であることを示すことができる。しかし，従来のテストに関するモデルは，古典的テストモデルやIRTなど，学習者の学力を定量的に解析するためものが中心であり，教師の予測を解析する数理モデルではない。そこで，本研究では，教師の予測の精度を解析する数理モデルを新たに提案し，数学的学力・学習力診断テストCOMPASSに適用する。

教師の予測の数理モデル

■教師の予測のモデル

はじめに，図1に示したように，ある一つの課題についてあるクラスや学校の子どもがどの程度の得点をとれるのかを，教師が予測するという場面を考える（学習者のテスト得点の分布を直接的に予測させるのは難しいため，0-1点，2-3点・・・のようにいくつかのカテゴリに分け，何%程度の学習者がそこに属するのかを問うこととする)。

課題が1つであり，予測すべき集団も1つである場合には，KL情報量を用いて分布同士の類似度を計算し，教師の予測の精度を相互に比較することができる。しかし，課題が複数ある場合や，基準となる学習者の実態が異なる場合には，適用することができない。そこで，予測と実測の関係を，図1のようにαとβをパラメータとする関数として一般化する。Q(i)はある教師の予測の分布，P(i)は実際の分布を指す（iは課題，jは教師，kはカテゴリを示す)。対数をとったあとの式を見ると明らかなように，このモデルはlnQijk，lnpPikを変数とし，αijとβijをパラメータとする線形モデルと捉えられる。

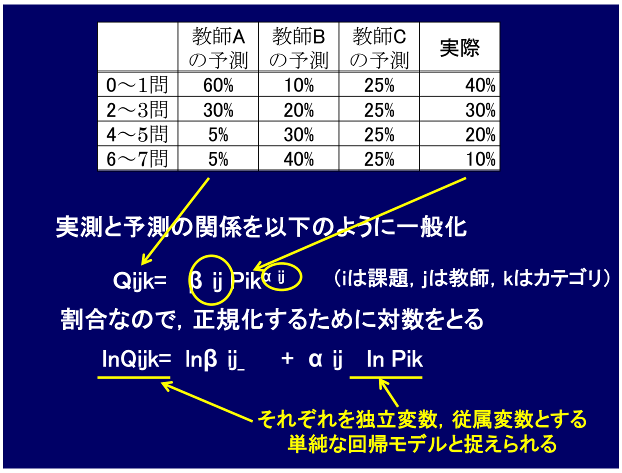

図１

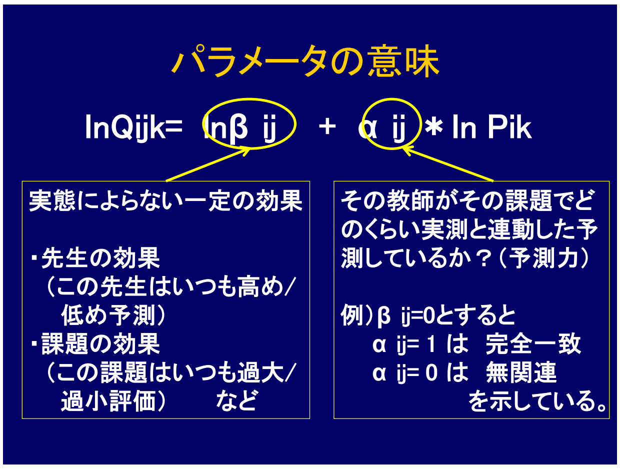

図２

■パラメータの意味

提案したモデルにおけるパラメータがそれぞれ何を示しているかをもう少し考えてみると，図 2 に示すように，α は教師の予測力を示しており，β は実態によらない一定の効果の総和を示している。1 人の教師のある課題において得られる α と β という 2 つのパラメータが，課題による違いや教師による違いによってどのように異なるのかを解析することによって，課題要因による効果や教師要因の効果を，定量的に示すことができる。

どのように解析するか？

■階層線形モデル（HLM）の利用

では，具体的にどのように解析できるのだろうか。線形モデルのパラメータ自体を変数として捉え，定量的に解析するモデルとして階層線形モデル（HLM）がある（図３）。これらのモデルでは，図４のように基本となるモデルで解析し，それぞれの式の誤差項が有意であった場合には，変数を新たに投入したモデルで解析するといった方法で進められる。本研究の分析に当たっては，HLM を用いて行った。

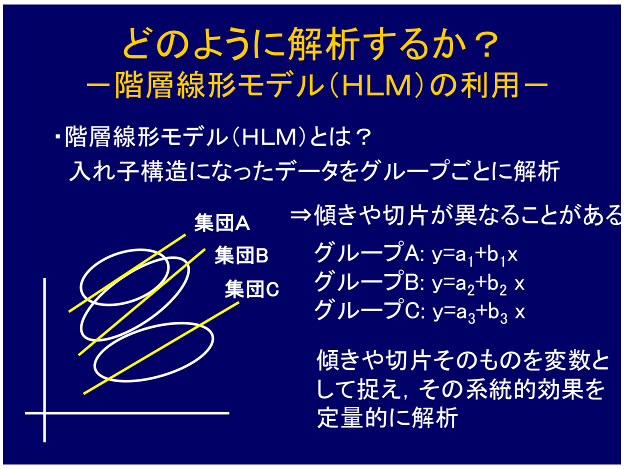

図３

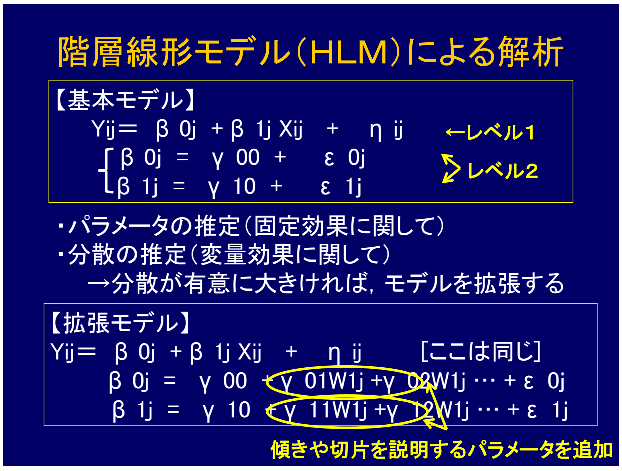

図４

実際のデータへの適用

■対象者と手続き

対象者は公立中学校 2 年生，計 468 名である（4 都道府県から 4 学校）と数学を担当した教師 18 名である。COMPASS（詳細は市川ら，2009 参照）を実施する前に，教師に対して質問紙を配布。教師に生徒の実態を予測するように依頼した。（予測させる際， 1 つの課題について 4−5 カテゴリに分け，何%の学生が所属するのかを質問した）。その後，生徒に COMPASS を実施した。

■解析結果

はじめに基本モデルで解析した結果，図 5 に示すように全体の誤差（η_{ijk}）が有意であった。そこで，モデルを拡張して解析した。最終的には，全体の誤差が最小となるような，図 6 のモデルを採択した。具体的には，切片と傾きに課題要因と教師要因を投入した。課題要因や教師要因の効果を解析する際には，ダミー変数を用いて分析した（1 つの課題，1 人の教師などを基準としてその効果を明らかにした）。

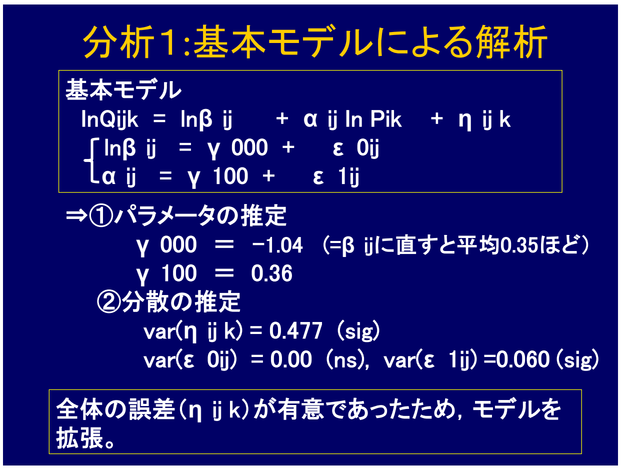

図５

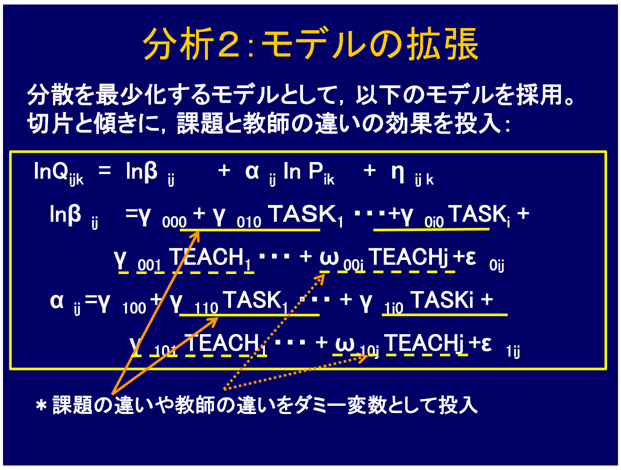

図６

パラメータにおける課題要因や教師要因の効果を解析したところ，図８，図９に示すように，予測の精度が高い課題や低い課題，予測精度高い教師や低い教師などが明らかとなった。さらに，図１１，図１２に示すように，実態によらない課題や教師ごとの効果が大きい（または小さい）課題や教師が明らかとなった。例えば，図表をどの程度自発的に利用しているのかを評価する課題（図表の自発的作成課題・図表の利用課題）では，計算課題の一部や単純な文章題などに比べてうまく予測できていない様子が伺えた。図表に関わる 2 つの課題は，子どものパフォーマンスも低いことが示されている課題である（植阪ら，2010）。この課題における学習者の得点が低いのみならず，教師の予測も正確ではないことから，外的リソースを自発的に利用するといった学力要素は，従来の学校教育では見過ごされてきた力であると考えられた。

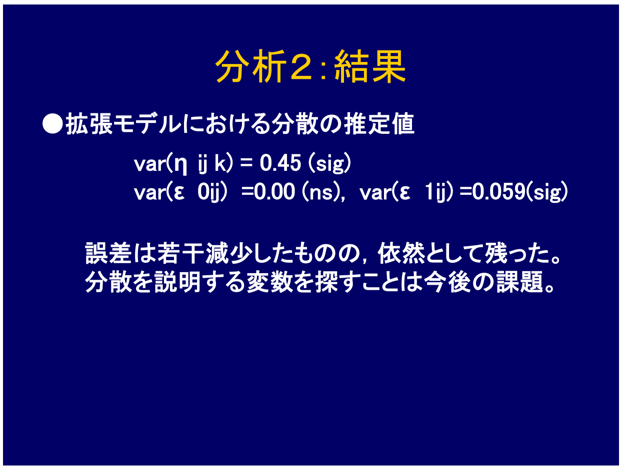

図７

分析2：結果

●傾き（α ij，予測力）における課題効果について：

※ α ijの値が最小であった図表の利用課題を基準に算出

（12,11,1,2,6,5,4,7,8）＜（9,3,10）　※有意水準5%

課題名	推定値	
12．図表の利用	（基準）0.000	γ 100の推定値は0．218
11．図表の自発的作成	0.031	
1．単純計算（四則・小数・分数）	0.047	
2．単純計算（正負・文字式）	0.062	
6．概念判断	0.122	
5．概念説明	0.133	
4．図表の解釈・作成	0.152	
7．論理判断	0.175	
8．単純速算（加減）	0.198	
9．単純速算（乗除）	0.291	
3．基本文章題	0.349	
10．工夫速算	0.451	

自発的な図表の作成や書き込みなどをみる問題（11，12）などでは，計算課題の一部(10,9,8)や単純な文章題（3)などに比べてうまく予測できていない。見過ごされてきた学力と考えられる。

図８

分析2：結果

●傾き（α ij，予測力）における教師効果について：

※ α ijの値が最小であった教師（101）を基準に算出

（101,404,403,104,･･･303)<(405,102,105) ※有意水準5%

教師ID	推定値	
101	（基準）0	γ 100の推定値は0．218
404	0.029	
403	0.038	
104	0.039	
201	0.086	
401	0.105	
303	0.118	
･･･		
205	0.257	
405	0.350	
102	0.367	
105	0.368	

一部の教師は，他の教師比べて予測力が高いことが示された。

図９

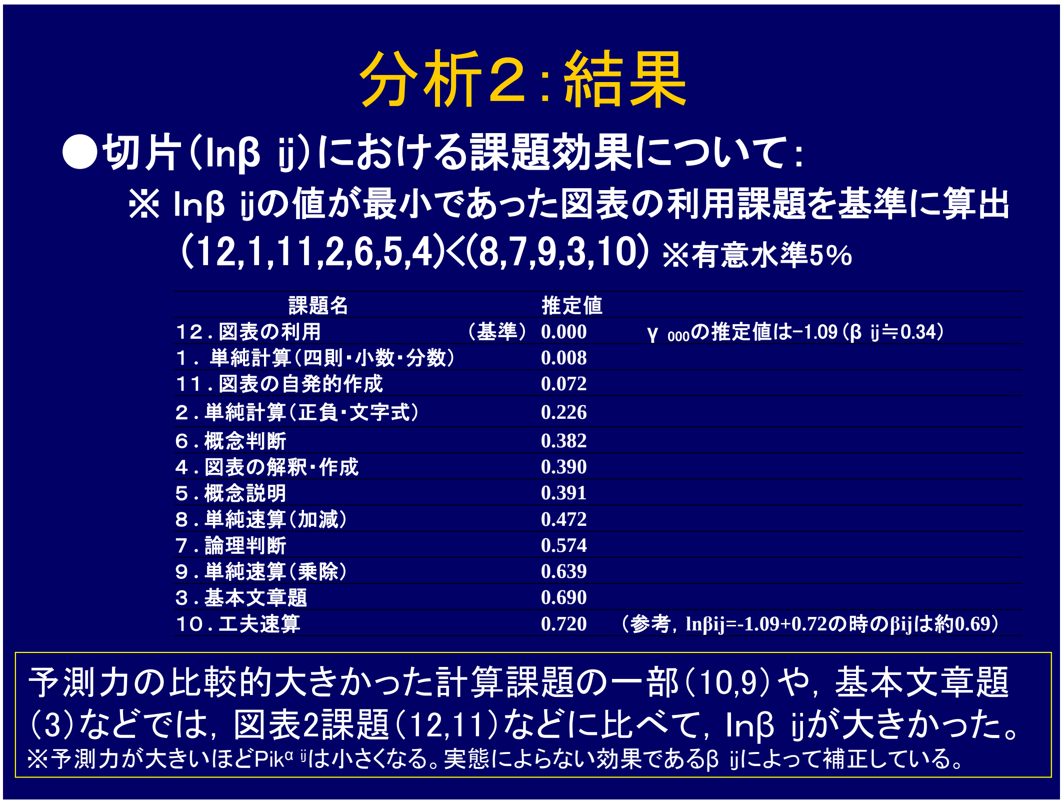
分析2：結果
●切片（lnβ ij）における課題効果について：
※ lnβ ijの値が最小であった図表の利用課題を基準に算出
（12,1,11,2,6,5,4）<（8,7,9,3,10） ※有意水準5%
課題名 推定値
12．図表の利用 （基準） 0.000 γ 000の推定値は−1.09（β ij≒0.34）
1．単純計算（四則・小数・分数） 0.008
11．図表の自発的作成 0.072
2．単純計算（正負・文字式） 0.226
6．概念判断 0.382
4．図表の解釈・作成 0.390
5．概念説明 0.391
8．単純速算（加減） 0.472
7．論理判断 0.574
9．単純速算（乗除） 0.639
3．基本文章題 0.690
10．工夫速算 0.720 （参考，lnβij=-1.09+0.72の時のβijは約0.69）
予測力の比較的大きかった計算課題の一部（10,9）や，基本文章題（3）などでは，図表2課題（12,11）などに比べて，lnβ ijが大きかった。
※予測力が大きいほどPikα ijは小さくなる。実態によらない効果であるβ ijによって補正している。

図１０

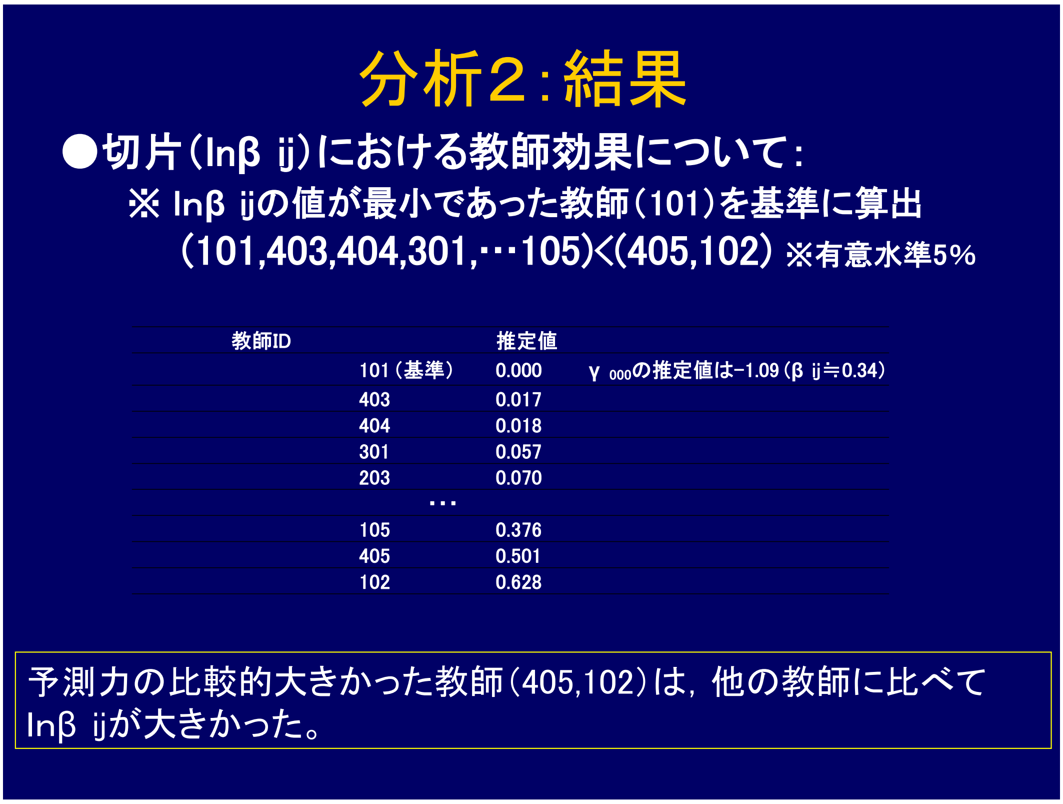
分析2：結果
●切片（lnβ ij）における教師効果について：
※ lnβ ijの値が最小であった教師（101）を基準に算出
（101,403,404,301,･･･105）<（405,102） ※有意水準5%
教師ID 推定値
101（基準） 0.000 γ 000の推定値は−1.09（β ij≒0.34）
403 0.017
404 0.018
301 0.057
203 0.070
…
105 0.376
405 0.501
102 0.628
予測力の比較的大きかった教師（405,102）は，他の教師に比べて lnβ ijが大きかった。

図１１

考察

教師の予測の精度を解析するモデルを提案するとともに，実際のデータに適用した。その結果，教師の予測が不正確な課題が明らかとなった。これまでのテスト得点関する実証研究と本論文で提案した方法を組み合わせて考えることによって，従来の学校現場の指導では「見過ごされてきた力」を実証的に明らかにすることができる。

ここで，本論文において提案した手法がどのような点で既存の数理モデルよりも優れているのかについて改めて論じる。分布間の類似度を評価する数理指標としては，前述したように KL 情報量が挙げられる。KL 情報量は，相対エントロピーとも呼ばれ，ある分布を基準として，それとの類似を複数の分布について計算する指標である。すなわち，KL 情報量を使うと，ある同一のテストについて，同一の集団をベースラインと，複数の予測を相互に比較することが可能である。具体的には，ある教科のテストについて，あるクラスの実測をベースラインとして，複数の教師の予測を相互に比較するような状況では利用が可能である。その一方で，ベースラインを固定する必要があることから，他の教科や，他のクラスの予測を統一的に扱うことができない。これに対して本研究で提案した解析モデルは，全く異なるベースラインの場合や，異なるテスト実施した場合であって適用可能である。

また，KL 情報量を用いた場合には，図１２に示したように，どのような点で似ているのかを明確にできないという問題点もある（図１２に示しているように，分布の形状が異なる教師 C において，最も類似度が高くなると判定されてしまうことも生じる)。一方，本論文の数理モデルは，αij を軸として，どの程度実態を反映した予測を行っているのかということを解析し，上述の問題を解消している。様々な状況で得られた予測を，統一した観点から解析でき，より実用可能性が高いモデルになっている点や，どのような意味で類似度が高いのかを明確にしているという点で従来の数理指標とは異なる利点があると言える。

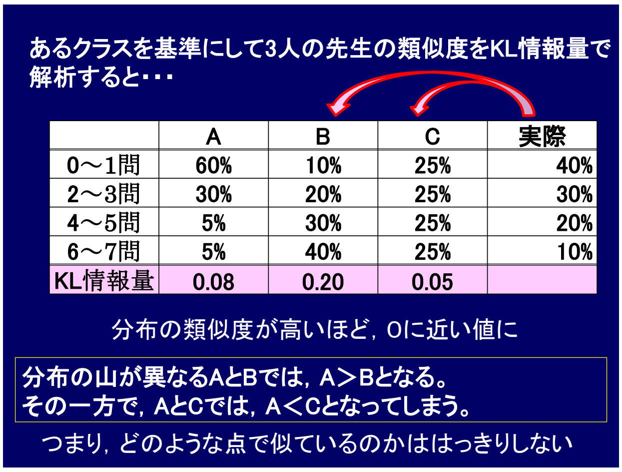

図１２

また，COMPASS のように，問題を解くために求められる学力を構成要素に分けて評価し，これらを予測させることで，単なる正答率の予測を超えたより有意義な分析を行うことが可能である。テストを行う際には，そこにおける正誤を明らかにするとともに，「なぜつまずいているのか」「どうして解けないのか」を明らかにすることが大切である。COMPASS では，それらを明らかにするような課題を設定している。こうした課題を用いて教師に予測を求め，本モデルを適用することによって，「解けるようにするために，従来不十分であったのはどのような指導か」といった具体的な示唆が得られる。COMPASS は算数・数学に特化してこうしたことを明らかにするツールであるが，この他の課題についてもどこにつまずいているのかを診断するテスト課題を用意し，予測を解析することで，より指導に生かしやすい知見を得ることができると考えられる。

引用文献

市川伸一・南風原朝和・杉澤武俊・瀬尾美紀子・清河幸子・犬塚美輪・村山航・植阪友理・小林寛子・篠ヶ谷圭太 (2009) 数学の学力・学習力診断テスト COMPASS の開発 認知科学, 16(3), 333-347

植阪友理・清河幸子・市川伸一 (2010) 構成要素型テスト COMPASS で見る日本の数学的基礎学力の実態—「基礎基本は十分，活用は不十分」は本当か— 第 74 回日本心理学会大会発表論文集，p775.