【原著】 大学入試研究ジャーナル第 29 号，150-154，2019

伸び率を用いた成績評価指標の改善

石原 正道（郡山女子大学），山口 猛（郡山女子大学短期大学部），古山 幹雄（郡山女子大学短期大学部[1]）

適切な成績評価指標として正答率の伸び率を用いた指標を導入した。導入した指標の妥当性を検討するため，導入した指標を含む３種類の指標を検証した。検証には数学の問題に対する正答率を用いた。正答率から得られる伸び率が成績評価として妥当であるか散布図により検討し，さらに伸び率が１回目の正答率に依存するか評価した。この結果，（１）単純な伸び率は成績の評価には適切でないこと，（２）単純な伸び率に比べ，高正答率であるほど大きく補正された伸び率は，正答率が高くなる場合には成績評価の妥当な指標であること，（３）単純な伸び率や高正答率であるほど大きく補正される伸び率に比べ，本研究で導入した補正された伸び率は，正答率の増減に係わらず妥当な成績評価の指標であること，が判明した。

1 はじめに

これまでに学習について様々な研究や試みが行われてきた。知識の獲得過程が調べられ（Nira, 1993），協同学習（Margarita, 1988）（Cynthia et al., 2003）（権・藤村, 2004）やアクティブラーニング（佐藤, 2015）などの手法が試みられてきた。これらの手法が有効であったかは，何らかの方法でその効果を測定する必要があり，その一つが試験の得点や得点から得られる伸び率などの指標による評価である。

教育において評価は重要な意味をもっている。初等教育から高等教育まで，教育の各段階では学習者の評価が行われる。評価は学習者の学習意欲とも係わっている。とりわけ学校教育では，評価は進級・進学・就職と係わることが多く，生徒や学生のみならず保護者にとっても評価は大きな関心事となっている。評価への関心の高さから，評価をどのように行うかは学習効果と係わっているといってよい。

入学試験や入社試験を実施する側からみると，合格者の能力をいかに伸ばすかが重要事項となる。合格者の能力向上の度合は学校の評価や会社の業績に直結する。このため合格者の学習効果を適切に評価し，合格者の能力を向上させる必要がある。

学習効果を高めるためには，学習者の学習意欲をいかに上げて維持するかが重要である。点数による評価は，点数の上がる余地の少ない学習者や成績の芳しくない学習者の学習意欲の上昇に寄与しがたい。これまでにも，ルーブリックによる評価や成績の伸び率による評価など，学習意欲を上げるよう評価は工夫されてきた（中島, 2018）。

学習効果を測るという観点から評価をみると，単純な点数による評価は問題を有している。一般的に試験では点数に上限があるため，成績のよい学習者は適切に評価されないという問題がある。いわゆる天井効果のために，学習者あるいは学習者群に対する学習効果を見積もることは難しい。この問題を解決するため，補正を加えた伸び率を導入して学習効果を測ることが行われてきた（兼田・新田, 2009）（Nitta, 2010）（西村・新田, 2014）（新田ほか２名, 2014）。

伸び率や補正を加えた伸び率により学習意欲の向上や学習効果の測定が行われてきたが，これらの指標には問題も存在する。天井効果による伸び率抑制の問題や過剰な補正の問題などがある。このため適切に評価を行うには，これらの問題を考慮した指標を用いることが必要である。

本研究では，学習成果を測るために，学習意欲を損なわない指標として伸び率を想定し，伸び率の指標としての妥当性を検討する。実際の理数系の問題に対する成績データを用いて，（１）単純な伸び率が適切な指標になっているか，（２）補正された伸び率が適切な指標になっているか，（３）伸び率や補正された伸び率にも問題がある場合，学力の高低の影響が少ない伸び率を定義し，実際のデータにどのように反映されるか，調べた。

本論文の構成は次の通りである。第２節では本論で用いるデータと分析手法について述べる。第３節ではデータを解析した結果を示す。第４節は結論を述べる。

2 データと分析手法

本研究で用いるデータは理数系の問題に対する正答率である。本データは 2011 年度から 2015 年度における K 大学短期大学部の１学科における講義科目において，１回目（開始時）に行った試験の正答率と２回目（終了時）に行った試験の正答率である。表１に年度毎の受講者数（データ数）を示す。表から分かる

ように全データ数は 90 である。問題数は１回目・２回目とも同じである。各問題の重みは等しいものとし，正答率を算出している。

伸び率を評価に用いることができるか調べるため正答率から伸び率を算出する。この際，単純な伸び率だけでなく，補正を加えた伸び率を定義し，どのような定義の伸び率が評価に用いることができるか検討する。検討のために，（1）１回目の正答率と伸び率による散布図を作成し，偏りがないか確認をする。また１回目の正答率により上位成績群と下位成績群に分け，これらの成績群の標準偏差を考える。上位成績群と下位成績群の伸び率の標準偏差に極端な差異がある場合は伸び率を評価に用いにくいため，（2）伸び率の標準偏差に差異があるか F 検定を行い検討する。

表 1　年度毎のデータ数

年度	2011	2012	2013	2014	2015
データ数	26	13	14	21	16

3 データの解析

3.1 複数の伸び率による散布図を用いた解析

議論を明確にするために次の量を定義する。まず $x_n^{(j)}$ は n 回目の試験における j 番目の人の正答率（$0 \le x_n^{(j)} \le 1$）である。この数値を用いて伸び率 $\zeta^{(j)}$ を次式で定義する。

$$\zeta^{(j)} = \left(x_2^{(j)} - x_1^{(j)}\right)/x_1^{(j)}$$

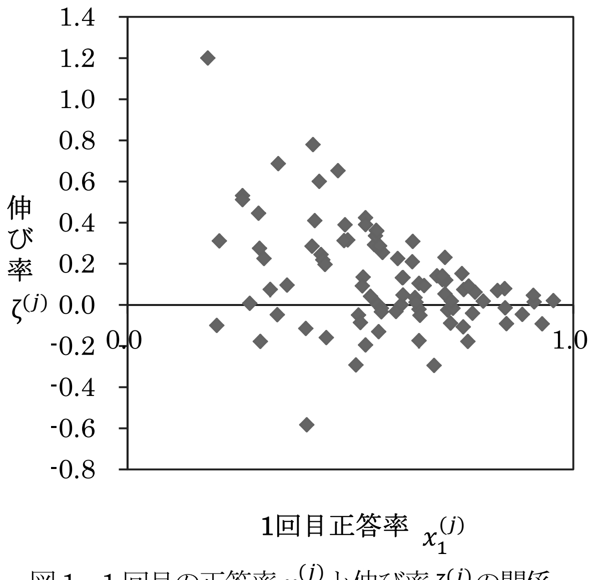

図１　１回目の正答率 $x_1^{(j)}$ と伸び率 $\zeta^{(j)}$ の関係

この伸び率 $\zeta^{(j)}$ と１回目の正答率 $x_1^{(j)}$ との間に何らかの関係があるか調べる。

図１は $x_1^{(j)}$ と $\zeta^{(j)}$ の関係を示した散布図である。図からわかるように，正答率の高い学生は伸び率が明らかに小さくなっている。正答率が高い方が伸び率の広がりは狭い（すなわち標準偏差が小さい）。

この事実は，$\zeta^{(j)}$ で用いられる関数系においては，天井効果（Margarita, 1988）により伸び率が抑制されているためであると考えられる。正答率には上限があるため，１回目の成績が良い学生は伸び率が上がりにくい。したがって伸び率 $\zeta^{(j)}$ によって成績を評価する場合，正答率の高い学生が不利になるという問題が生じうる。

天井効果を取り除くため補正された伸び率を定義する。ここでは Hake らによる伸び率（Hake, 1998）（兼田・新田, 2009）を援用し定義された伸び率（PI 効率）を導入する（兼田・新田, 2009）（Nitta, 2010）（西村・新田, 2014）（新田ほか２名, 2014）。

$$\eta^{(j)} = \left(x_2^{(j)} - x_1^{(j)}\right)/\left(1 - x_1^{(j)}\right)$$

分母は１回目の正答率が高いほど補正因子が大きくなるように調整されている。

図２は $x_1^{(j)}$ と $\eta^{(j)}$ の関係を示した散布図である。伸び率 $\zeta^{(j)}$ に対する図１と比べると，１回目の正答率が高い学生の $\eta^{(j)}$ は，１回目の正答率が低い学生の $\eta^{(j)}$ とほぼ同じ値となっていることが分かる。このことは伸び率による評価をする上で必要なことである。

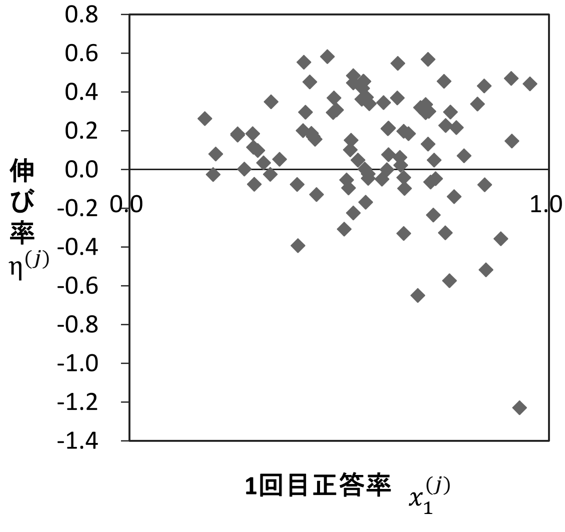

図２　１回目の正答率 $x_1^{(j)}$ と伸び率 $\eta^{(j)}$ の関係

補正された伸び率$\eta^{(j)}$にも問題が残っている。図２では，１回目の評価の高い学生の２回目の正答率が下がった場合に，非常に低い伸び率として観測される。これは$\eta^{(j)}$における補正は成績が下がる場合にも適用されることによる。しかしこの補正は明らかに過剰な補正となっている。伸び率$\eta^{(j)}$の補正は天井効果に対する補正であった。しかし正答率の下降は１回目の正答率が低い方が起こりにくい。すなわち１回目の正答率が高い学生は，下がる余地が大きいことを考慮しないといけない。

そこで，あらたに正答率が下がる可能性を考慮した伸び率$\kappa^{(j)}$を定義する。

$$\kappa^{(j)} = \begin{cases} \left(x_2^{(j)} - x_1^{(j)}\right) / \left| 1 - x_1^{(j)} \right| & \left(x_2^{(j)} \geq x_1^{(j)}\right) \\ \left(x_2^{(j)} - x_1^{(j)}\right) / \left| 0 - x_1^{(j)} \right| & \left(x_2^{(j)} < x_1^{(j)}\right) \end{cases}$$

伸び率$\boldsymbol{\kappa}^{(j)}$では，正答率が上がる場合は上限までの距離による補正を行い，正答率が下がる場合は下限までの距離による補正を行っている。関係式$\boldsymbol{x}_2^{(j)} < \boldsymbol{x}_1^{(j)}$が成り立つ場合は，分母は$\left|\mathbf{0} - \boldsymbol{x}_1^{(j)}\right| = \left|\boldsymbol{x}_1^{(j)} - \mathbf{0}\right| = \boldsymbol{x}_1^{(j)}$となるので，$\boldsymbol{\kappa}^{(j)}$は$\boldsymbol{\zeta}^{(j)}$に一致する。また正答率$\boldsymbol{x}_1^{(j)}$が１または０に一致する可能性がある場合に発散を避けるには，十分小さな$\boldsymbol{\varepsilon} > \mathbf{0}$を用いて，分母の１を$(\mathbf{1} + \boldsymbol{\varepsilon})$に置き換え，０を$(\mathbf{0} - \boldsymbol{\varepsilon})$に置き換えるなどの方法により対応することができる。

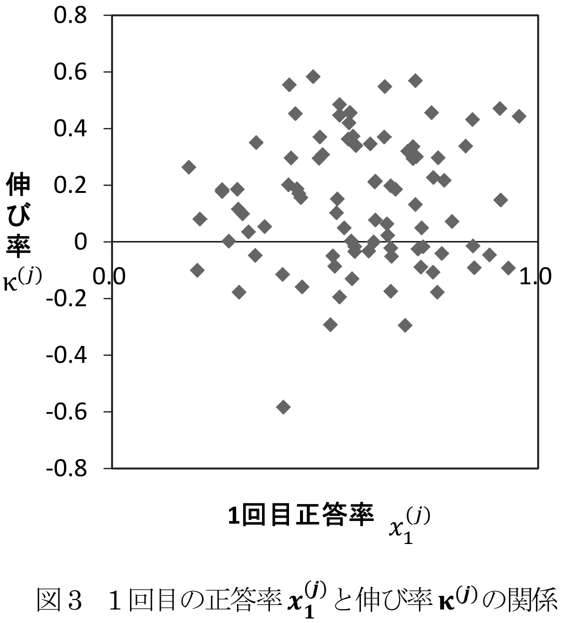

図３　１回目の正答率$\boldsymbol{x}_1^{(j)}$と伸び率$\boldsymbol{\kappa}^{(j)}$の関係

図３は$x_1^{(j)}$と$\kappa^{(j)}$の関係を示した散布図である。図２で問題となった，１回目の正答率が高く２回目で正答率の下がった学生の伸び率が適切に補正されている。

散布図からは評価に用いる伸び率として，伸び率$\kappa^{(j)}$が妥当であるように思える。この事を確認するために，次小節ではF検定により，伸び率$\kappa^{(j)}$が妥当な指標であることを確認する。

3.2 標準偏差を用いた分析

伸び率を評価に用いる場合，１回目の正答率に依らずに伸び率が評価できなければならない。１回目の正答率の高低に，その後の伸び率が依存するかどうかは議論の余地があるが，本論文では「適切な伸び率は，１回目の正答率の高い学生と低い学生間で，伸び率の標準偏差に大きな違いがみられない」ことを適切な指標の条件とする。既述したように，天井効果のために単純な伸び率で評価することは難しい。このことは本研究では標準偏差の違いとして現れる。以下では伸び率$\zeta^{(j)}$，$\eta^{(j)}$，$\kappa^{(j)}$が上記の条件を満たすか調べる。

本分析の手順は次の通りである。まず１回目の正答率に応じてデータを二分する。二分した群のうち，成績上位の群をＡ群，成績下位の群をＢ群とする。本データではA群とB群の境界上に正答率が等しいデータがあったため，Ａ群のデータ数は46，Ｂ群のデータ数は44となった。次に得られた２群の標準偏差について「２群の標準偏差は等しい」という帰無仮説が棄却されるか否か，F検定を行い確認する。本論文では有意水準を５%とする。

表２に，各指標で得られた値から算出される標準偏差を用い，Ａ群とＢ群の標準偏差が等しいとした場合の両側確率を示す。

F検定の結果から，伸び率$\kappa^{(j)}$を成績評価に用いることが妥当であることが分かる。伸び率$\zeta^{(j)}$および$\eta^{(j)}$から得られた標準偏差では「Ａ群とＢ群の標準偏差が等しい」とした帰無仮説は棄却される。一方，伸び率$\kappa^{(j)}$から得られた標準偏差では帰無仮説は棄却されな

表２　正答率の高い群と低い群の標準偏差が等しいとした場合の両側確率

指標	両側確率
$\zeta^{(j)}$	6.16×10^{-10}
$\eta^{(j)}$	8.69×10^{-3}
$\kappa^{(j)}$	3.05×10^{-1}

い。このことは伸び率$\zeta^{(j)}$および$\eta^{(j)}$を成績評価に用いた場合には，歪みが存在する可能性を示している。

4 結論

本研究ではＫ大学短期大学部の講義で得られた理数系科目における１回目の試験の正答率データと２回目の試験の正答率データを用い，伸び率を用いた評価の妥当性を調べた。１回目の正答率と得られた伸び率の関係を調べることで，評価を行う上で適切な伸び率を明らかにした。

まず，素朴に定義された伸び率は，１回目の正答率が高い学生に対しては不利に働くことが示された。この理由は伸び率の関数系と正答率に上限が存在することによる天井効果のためであると考えられる。

天井効果を補正することで，１回目の正答率が高い学生の評価にも伸び率を適用することが考えられる。天井効果を補正した伸び率を用いた結果から，正答率が高くなった学生に対しては適切な補正が行われることが分かった。しかし正答率が低くなった場合は，過剰な補正が行われることが明らかとなった。

そこで本研究では，過剰な補正を抑制するために新たに伸び率を定義した。正答率には上限のみならず下限がある。そこで新たな伸び率では，上限の効果と同様に下限の効果を補正した。新たに定義した伸び率を用いると，正答率が高くなるか低くなるかに依らず，改善された伸び率が得られることが明らかとなった。

次に定義した伸び率の妥当性を検討した。１回目の正答率の上位者と下位者の２群に分けた場合，２群の標準偏差の差がなければ伸び率として妥当であると考えられる。この観点からは本論にて導入した伸び率が適切であることが示された。

以上から，単純な伸び率や正答率が高くなることのみを想定した伸び率は，公平性の観点から評価には用いにくいことが明らかとなった。評価に用いるには本論で新たに定義した伸び率のように，正答率が低くなることも考慮した伸び率を使う必要がある。本研究で定義した伸び率$\kappa^{(j)}$は成績が悪くなることも考慮することで一定の改善がされており，評価の指標として用いることができると考えられる。

本研究では補正を行うことで指標に一定の改善がみられた。しかし補正の問題が完全に解決されたわけではない。例えば，１回目に低正答率（低得点）であり２回目に高正答率（高得点）である場合に分子の差が大きくなることから，伸び率として有利に働く可能性がある。適切に補正をするには $\left|1-x_1^{(j)}\right|$ を $\left(|1-x_1^{(j)}|\right)^{\alpha}$ などの別の関数系に変更する必要がある。また正答率（得点）の差を用いることで信頼性が低下する可能性が高く，検討を要する。以上のように提案した指標で問題点が完全に取り除かれたわけではない。

本論文では新たな伸び率を提案したので，教育現場で用いる場合に留意すべき点に触れておくことにする。伸び率を使う理由の一つは，初期に低正答率（低得点）である学習者も伸び率は高くなりうるため，初期に低正答率である学習者の学習意欲を維持する効果が期待できることである。一方で最初に高正答率（高得点）であった学習者には不利となる可能性がある。このことを解消するために補正した伸び率を考えたのであるが，高正答率を獲得しても伸び率では低評価という状況の生じる可能性が完全に除去されたわけではない。このことは初期に高正答率者の学習意欲を削ぐことになりかねない。したがって，教員側が意識する低正答率者と高正答率者の学習意欲の維持のために，実際の運用では，加工していない正答率（得点）による評価と伸び率（本論で定義された伸び率を含む）による評価を適切な重みによって勘案するなどの方策が必要となる。伸び率を用いた評価を行う際には，被評価者に評価方法を明示的に示すなどの対応が必要だろう。

学習者の学習意欲を維持するために評価の指標は重要である。評価の指標の一つとして，本研究では正答率の伸び率に着目して新たな指標を導入し，その効果を明らかにした。指標には他の定義も考えらえるため，多くの指標を検討し，より適切な指標を得る必要があるだろう。

注

1）2018 年度より，郡山女子大学短期大学部非常勤講師.

参考文献

Cynthia A. R. et al. (2003). Peer-Assisted Learning Interventions With Elementary School Students: A Meta-Analytic Review, *J. Edu. Psy.*, 95, 240-257.

権裕善・藤村宣之（2004）．「同年齢児童の協同はいつ有効であるのか―比例的推理の方略レベルが異なるペアの相互作用―」『教育心理学研究』52, 148-158.

Hake, R.（1998）．Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Ann. J. Phys.*, **66**, 64-74.

兼田真之・新田英雄（2009）．　「クリッカーを用いたピア・インストラクションの授業実践」『物理教育』 **57**, 103-107.

Margarita, A.（1988）. Peer Interaction and Problem Solving: When Are Two Heads Better Than One?, *Child Development* , **59**, 87-96.

中島英博（2018）.「学習評価」 玉川大学出版部.

Nira, G.（1993）. Patterns of Interaction in the Co-Construction of Knowledge: Separate Minds, Joint Effort, and Weird Creatures, In R. H. Wonzniak, & K. W. Fischer (Eds.) *Development in context: Acting and thinking in specific environments*, Lawrence Erlbaum.

西村塁太・新田英雄（2014）.「『振り返り』を導入したピア・インストラクション型授業」『物理教育』 **62**, 7-12.

Nitta, H.（2010）. Mathematical theory of peer-instruction dynamics, *Phys. Rev. St Phys. Educ. Res.,* 6 , 020105.

新田英雄・松浦執・工藤知草（2014）.「ピア・インストラクションを導入した物理入門講義の実践と分析」『科学教育研究』 **38**, 12-19.

佐藤実（2015）.「物理基礎教育におけるアクティブ・ラーニングの試み」『 2015 PC カンファレンス』 259-260.
<http://gakkai.univcoop.or.jp/pcc/2015/papers/pdf/pcc064.pdf> （2016 年 11 月 23 日）