

中高生は英語の能力についてどのくらい正確に自己評価をしているか?¹⁾ －学習動機づけと事後の成績との関連も踏まえた検討－

Accuracy of Japanese Junior High and High school student Self-Assessment of Performance on English Test

西 村 洋 一

Abstract

The aim of this study was to investigate how student self-assessment of their performance on English test differed from their real ones. It was predicted that students who score low on a performance test would overestimate their performances, relative to others, whereas high scorers would underestimate their own performances (the Dunning-Kruger effect). Japanese junior high school students and high school students took an English test (Young Learners English) and answered a questionnaire based on their performances on the test. The results supported the prediction. Moreover, the discrepancies between their estimates and real performances were positively related to their learning motivation and future test performances, especially in good performers. Furthermore, the implication of these results for English education are discussed.

キーワード：自己評価 (self-assessment, self-evaluation) / 学習動機づけ (learning motivation) /
ダニング＝クルーガー効果 (Dunning-Kruger effect)

I 問題

私たちは、自身の能力やパフォーマンスについてどの程度正確に評価を行うことができるのだろうか。近年は教育場面においても自分自身での評価を課されることが多くなっている。これは自身の能力、パフォーマンスを把握することが、事後の学習の成果に結び付くよう期待されているのであろう。このように、私たちは自身の能力やパフォーマンスの自己評価を行うことは珍しいことではないが、その正確性については必ずしも高くはないことが指摘されている (例えばレビューとして、Dunning, Heath, & Suls, 2004)。

両者の関係の具体的検討として、Mabe & West (1982) は、能力の自己評価とパフォーマンスとの間の相関についてメタ分析を行っている。特定の能力やスキルの自己評価と他の測度との関連を行っている55の研究を取り上げたが、相関係数の

範囲としては、 -0.26 から 0.80 であり、平均では $r = 0.29$ (標準偏差は 0.25) という結果であった。このような低い相関係数が得られた理由として測定における誤差の問題が考えられるが、その点を考慮して修正したとしても、 $r = 0.42$ と中程度の相関でしかないとの推定がなされている。

Dunning et al. (2004) は、このような乏しい自己評価の正確性 (特に過大評価—非現実的な楽観主義) につながる心理学的要因を列挙し、総合して2つの観点にまとめている。1つに、正確な査定を行うのに必要なすべての情報を手に入れられない、そしてそのことを考慮しないという点である。そしてもう1つが、手に入れている有用な情報を無視する点である。さらに、別の考慮すべき観点として、自己高揚的な動機づけが働いている可能性も挙げている。

ここまで述べてきたように、自己評価は様々な場面で活用されているものの、心理学的なメカニズムにより、能力やパフォーマンスについて、客観的な評価と一致する正確な自己評価を行うこと

NISHIMURA, Youichi

北陸学院大学 人間総合学部 社会学科
社会心理学A・B

は困難であることが理解される。

教育場面における自己評価

本研究では、教育場面における自己評価のあり方に焦点を当てる。そこで、教育場面での自己評価と実際の客観的なパフォーマンスとの間の関連に関する既存の知見を振り返る。

Hansford & Hattie (1982) は、教育関連の研究を選定して、教育的な達成やパフォーマンスと自己関連の測度(例えば、自己概念や自尊感情など)の関連を調べた研究のメタ分析を行った。両者の間の相関係数は、範囲としては -0.77 から 0.96 の間にあり、その平均は $r=0.21$ (標準偏差は 0.007)という結果であった。また、高等教育における両者の関連についてメタ分析を行った結果では、学生自身の自己評価と教員による評価の相関係数の平均として、 $r=0.39$ という値が得られている(Falchikov & Boud, 1989)。

本研究では「英語」に対する自己評価を取り扱うが、英語に焦点を当てた自己評価と客観的評価との間の関連についてはどのような結果が得られているだろうか。ベネッセ総合教育研究所が実施した「第3回学習基本調査」では、小中高生を対象に数学・国語・英語の到達度テストの結果と成績の自己評価との間の関連を見ている。高校生においては、相関係数を高校のランク別に算出がされており、英語については、 $0.50 \leq r \leq 0.61$ であり、国語が $0.32 \leq r \leq 0.49$ 、数学が $0.47 \leq r \leq 0.59$ 、そして総合成績で、 $0.50 \leq r \leq 0.64$ という結果であった。学校ランクによる違いに着目すると、上位校において相関が高く、下位校が低いという傾向が見られた。中学生についての結果は、相関係数は算出されていない。ただし、自己評価と実際の成績区分の割合を見てみると、国語、数学で両者の間の関連が見て取れる。英語については国語、数学との成績区分との関連についての結果が示されているが、教科間での同様の関連も見られている。小学生では、調査を行ったのが国語と算数のみであるが、調査結果を見ると、中学生ほどにはきれいな関連が観察できないという結果であった(ベネッセ総合教育研究所, 2001a, 2001b, 2001c)。他に同様の検討を行った物井(2016)においては、小学生5, 6年生を対象に英語運用能力と自己評価

の関連を検討し、両者の間に、 $r=0.30$ という相関係数が得られた。これらの結果は英語の能力についての自己評価には一定の正確性があることを示しているものの、相当のずれがあるということも言えるということである。

それでは、そのずれのあり方になにか特徴はあるのだろうか。米田・細川・西村・物井(2013)では、小学校から大学生を対象に、英語の学力と自己評価との間の関連を検討している。両者の関係について学校種の違いによる影響はあまり見られなかったが、学力別に検討してみると、学力が高いほど自己評価との一致が多くなるという結果が得られている。また、この研究で興味深いのは、一致と不一致(過大評価、過小評価)の割合が示されている点である。学力の下位群(CEFRレベルでのPre-A1)においてはその大部分が過大評価を行い(具体的には97.8%が過大評価)、中位群(同レベルでのA1)ではその割合が減り(68.0%)、上位群(同レベルでのA2, B1)では過大評価は少ないかほとんどおらず、一致か過小評価が多くなるという傾向が見られた²。この学力と自己評価について、クラメールの連関係数は $V=0.50$ という値を示し、有意な関連が示されたが、自己評価のあり方は、学力の下位群が過大評価、上位群が過小評価というような形でずれが見られたということである。そして、さらに注目される点として、この自己評価のずれのあり方は、いわゆるDunning-Kruger効果と同様の形であることが挙げられる。

Dunning-Kruger効果

能力やパフォーマンスの自己評価について特に能力や成績の低い人に顕著に過大評価する傾向がKruger & Dunning(1999)の研究により明らかにされている。そこでは、ある領域の自身の能力やスキルなどを適切に評価するためには、その領域の能力やスキルが必要であり、その能力を欠く人は、メタ認知が適切に機能しないと考えられている。また、このような人々は二重の重荷を背負い込むようなものであるとも説明される(Dunning, 2011)。すなわち、まずスキルや能力を欠くために多くの誤りを犯してしまう。そして、その同じスキルや能力の欠如は、自分が間違いを

犯したこと、そして他者はより適切に実行していることに対する認識の失敗へとつながるということである。結果としてこれらのことは、自身のスキルを見直し、向上・改善を図るきっかけを失わせてしまう。

Kruger & Dunning (1999) では、ユーモア、論理的推論の能力、英語の文法の力を題材に、それぞれのテストのパフォーマンスについての相対的評価（順位の推定）と絶対的評価（テストの得点の推定）を測定した。その結果、能力を欠く人は相当な過大評価を行い、他者の能力との対比の中で自身の能力を認識することに困難を抱えることを明らかにしている。具体的には、全体の平均として、自身の集団内の相対値とをパーセンタイル順位で66位と回答するといった平均以上効果を示し、特に下位25%の位置にある人たちは、実際のパーセンタイル順位が12位のところ平均して自身を58位というように理論上の平均（50位）よりも高く自己評価を行っていた（Study 1の結果）。加えて、良いパフォーマンスを示した人においては、自身のパフォーマンスを過小評価する傾向も見られた。その過小評価の大きさはパフォーマンスの乏しい人に比べると小さいが（課題によって異なるという点も見られたが）、一貫した傾向として結果が得られている。続く研究（Study 3, 4）において、パフォーマンスの乏しい人がメタ認知的なスキルを欠き、他者の能力、パフォーマンスの情報を自身の評価に活用できていないこと、そして、能力を高める介入により自己評価のずれが低減することを示した。

このような自己評価のパターンは、その後の研究でも頑健に示されている。例えば、学生が受ける試験（Dunning, Johnson, Ehrlinger, & Kruger; Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008）やディベート大会の出場者（Ehrlinger et al., 2008）、教習中のドライバーの運転に関する自己評価（Mynttinen, Sundström, Vissers, Koivukoski, Hakuli, & Keskinen, 2009）など、多様な領域においていずれも同様の結果が得られている。

本研究の目的

本研究では、英語のパフォーマンスを題材として、その自己評価の実態について理解することを

目的とする。先行研究からは、自己評価と実際のパフォーマンスの間にはある程度の相関が見られるものの、ずれが確実に存在することも示されている。さらにそのずれのあり方について上記の Dunning-Kruger 効果がみられるかという点に焦点を当てる。メタ認知が発揮され、自己評価が適切になされることで、学習が制御されるということが考えられる（例えば、Thiede, Anderson, & Therriault, 2003）。そうであるならば、自己評価のあり方を把握し、つまずきがある場合には介入をする必要があるだろう。その際、Dunning-Kruger 効果の研究で示されているように、パフォーマンスの良い人と乏しい人との間では、ずれのあり方とその要因が異なることになり、その介入のあり方、例えば、教育場面での指導法なども変えていく必要があることが考えられる。そこで、本研究では、検討手法として Kruger & Dunning (1999) の方法に則りながら、英語のパフォーマンスについての自己評価と実際のパフォーマンスとの関連のあり方について検討を行う。

先行研究を踏まえ、本研究の仮説は以下に示すとおりである。

1. 英語の成績の下位グループは、自身のパフォーマンスについて過剰評価を行う。それは、相対的評価（パーセンタイル順位）、および絶対的評価（テストの得点）の両方において見られる。
2. 英語の成績の上位グループは、自身のパフォーマンスについて過小評価を行う。先行研究の知見を踏まえれば、特に相対的評価において過小評価が見られ、絶対的評価においては大きなずれは見られないと予測される。

また、本研究では、中高生を対象に調査を行う。中学生と高校生との間に、Dunning-Kruger 効果のあり方に差が見られるかについては、明確な仮説はない。英語に関する自己評価と能力の先行研究からは、小学生においては相関が低い傾向が見られるものの、中学生、高校生では明確な違いが見られていない。ただし、Kruger & Dunning (1999) で考えられているように、メタ認知スキルの差によるところが大きいとするのであれば、青年期におけるその発達的な違い（例えば、van der Stel & Veenman, 2010）や英語に対する意識の差（ベネッセ総合教育研究所, 2014）により、中学生と高

校生の間に違いが見られることも考えられる。そこで、中学生、高校生の中に自己評価と実際のパフォーマンスの関連パターンに違いが見られるのかという点をリサーチクエスチョンとし、検討を行いたい。

さらに、本研究では、自己評価と実際のパフォーマンスの間に不一致が見られるとして、それが学習過程、あるいはその後の学習の結果にどのような帰結をもたらすのかという点も視野に入れる。特に、パフォーマンスの下位群と上位群の間で自己評価と実際のパフォーマンスとの間のずれのあり方に違いが見られ、その要因も異なるということであれば、学習過程に関わる要因や自己の学習の成果に異なる形で関連を示すことも考えられる。具体的にどのような関連が見られるかということについて明確な仮説はなく、探索的な検討でしかないが、何らかの傾向がつかめるのであれば、そこから教育的な示唆が得られることも期待される。そこで、本研究では、英語に対する学習動機づけと調査後に行われた英語のテスト結果を用いて、検討を行うこととする。

II 方法

調査対象

関東地方および北陸地方の中学校A, B校, そして高等学校C, D, E校の計5校の協力により、調査を実施した。中学校はA校が1年生, B校が2年生であり, 高校は3校とも2年生であった。英語力を測るためのテスト結果, および質問項目にすべて回答した中学生309名, 高校生412名を対象に分析を行った。また, そのうち, およそ4ヶ月後に実施された同レベルの英語テストの点数が得られた中学校A校187名, 高校C校200名分のデータを自己評価と将来の英語テスト得点との関連の分析に用いた。

調査内容

英語力の測定

本研究において児童, 生徒の英語力を測るためにYoung Learners English Test (以下, YLE) を使用した。本テストは本来7~12歳を対象としたものであるが, 調査対象の学力と試験実施時間を考慮し, 本テストを使用することとした。YLEはレ

ベルによってFlyers, Movers, Startersの3つに分かれる。調査対象となった各協力校と相談し, 用いるテストの種類を決定した。その結果, 本研究においては, 中学校がStarters (A校, B校) を, 高校が3校ともFlyersを受験した生徒のデータを採用した。

また, 英語力は本来4技能の観点から評価されるべきものであるが, テストの実施時間を考慮し, 中学校, 高校においてはReadingとWritingのテストのみ実施した。

採点については全てのテストについて正答数を問題数で除し, 英語力の指標とした。

英語力の自己評価

Kruger & Dunning (1999) を参考に, 自身のYLEの結果, 総合的な英語力に関する同じ学年内におけるパーセンタイル順位, およびYLEの試験の正解率の見積りを尋ねた。その際, YLEと英語力の順位については図を添付して考えやすくなるよう配慮した。回答方法はすべて数値を直接記載してもらう形で回答してもらった。

英語学習動機づけ

米田・西村・細川 (2014) で使用した英語学習に対する動機づけを尋ねる質問項目を使用した。本尺度は英語学習に対する意欲や関心について測定を行う際に用いられることの多いAMTBを参照しながら項目内容を検討したものである。12項目からなるが, 因子分析の結果, 「英語学習動機づけ」(10項目) と「コミュニケーション懸念」(2項目) に分かれた。本研究では「英語学習動機づけ」を用いた。「いいえ」(1) から「はい」(5) の5件法で回答を求めた。

III 結果

英語テストの結果, 英語力の自己評価の記述統計, および実際の結果との関連

英語のパフォーマンスについての自己評価の平均及び標準偏差を算出し, Table 1 に結果を示した。中学生, 高校生ともに順位の推定の平均は50位を超えることはなく, それよりやや低いという結果であった。

次に, テスト順位や英語力順位, ならびにテス

ト得点推定と実際のテスト結果（順位，得点）との相関係数を算出した。順位について中学生においては，実際のテスト結果の順位とテスト順位 ($r = .51, p < .001$) や英語力順位の推定 ($r = .50, p < .001$) となり，中程度の相関がみられた。高校生もほぼ同様であり，実際のテスト結果の順位とテスト順位 ($r = .61, p < .001$) や英語力順位の推定 ($r = .57, p < .001$) となり，中程度の相関がみられた。テスト得点の実際と推定の相関係数は，中学生で ($r = .60, p < .001$)，高校生で ($r = .59, p < .001$) であり，中程度の相関であった。

テスト順位の推定と実際の順位の差

YLEの結果について，自身の相対的位置（順位）の推定と実際の順位がどのような関係にあるかについて検討を行った。まず，調査対象者のYLEの得点を用いて，学校ごとにパーセンタイル順位を算出した。そして，得点の四分位数を割り出し，4つの群の実際のテスト得点のパーセンタイル順位，そしてテスト得点のパーセンタイル順位の推定，および総合的な英語力の順位の推定について平均を算出した。中高各校の結果は，ほぼ同様の結果を示したため，ここでは中学校，高校それぞれの学校を併合して算出した結果を示す。結果は

Figure 1, 2に示した。

中学生，高校生のどちらにおいても，成績の下位25%の群（Bottom Quartile）において実際のテスト順位よりも推定の方が高い順位となるような差が見られた。中学生においては，実際のテスト順位が平均で11位となっているところ，推定では32位と見立てているという結果であった。高校では実際には平均で12位のところ，32位という推定であった。この差について対応のあるサンプルのt検定を行ったところ，統計的に有意な差であることが示された（中学生： $t(71) = 7.3, p < .001$ ；高校生： $t(101) = 15.2, p < .001$ ）。

2nd Quartileの群においては，Bottom Quartileの群ほどではないが，やはり推定された順位の方が実際の順位よりも高かった（中学生：推定＝平均46位，実際＝32位 $t(65) = 5.4, p < .001$ ；高校生：推定＝41位，実際＝35位 $t(105) = 4.7, p < .001$ ）。

3rd Quartileの群では，中学生は実際が平均52位のところ，推定は49位と少し推定が下回ったものの，統計的に有意な差は見られなかった。高校生においても同様に推定の方が実際の順位よりも下回り（実際＝60位，推定＝48位），両者の差は統計的に有意であった ($t(98) = 8.5, p < .001$)。

最後にTop Quartileの群では，中学生において

Table 1 英語力（順位，得点）の自己評価の平均

	テスト順位の推定	英語力順位の推定	テスト得点の推定
中学生	47.2 (21.1)	45.1 (21.2)	51.9 (22.2)
高校生	44.8 (15.8)	43.3 (17.4)	50.9 (17.4)

注) () 内は標準偏差

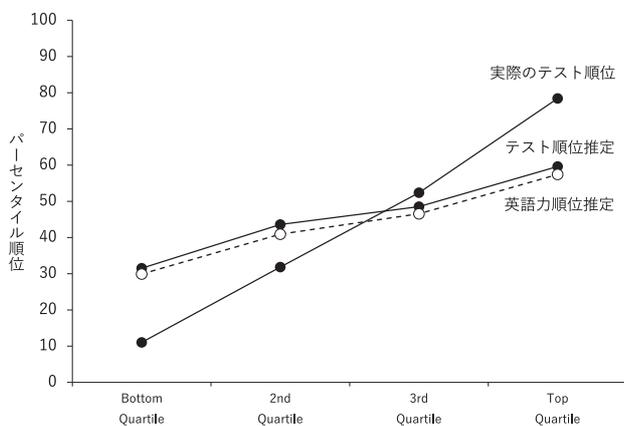


Figure 1 中学生のテスト得点，英語力の実際と推定

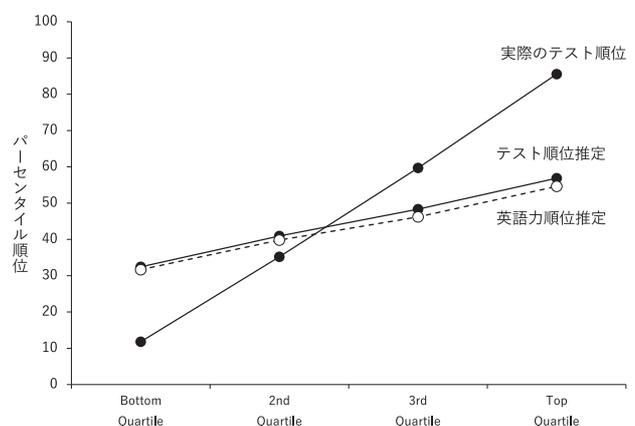


Figure 2 高校生のテスト得点，英語力の実際と推定

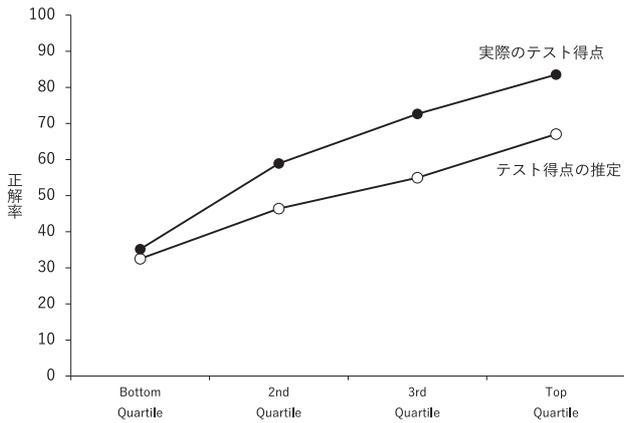


Figure 3 中学生のテスト得点の推定と実際のテスト得点

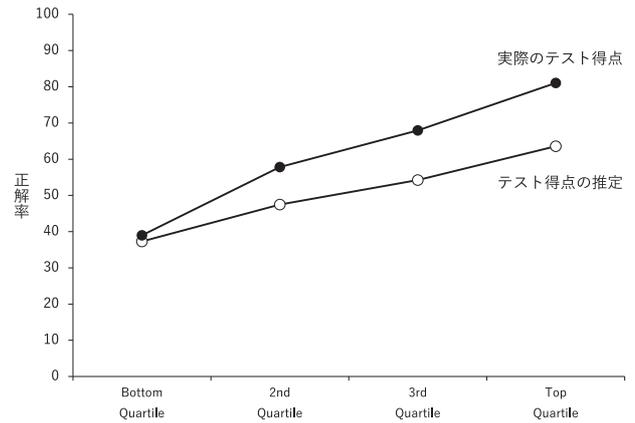


Figure 4 高校生のテスト得点の推定と実際のテスト得点

Table 2 実際のテスト得点と推定の学校種、群ごとの分析結果

群		得点の推定と実際の差	t
中学生	Bottom Quartile	2.7	1.0
	2nd Quartile	12.5	5.4***
	3rd Quartile	17.7	8.3***
	Top Quartile	16.5	12.2***
高校生	Bottom Quartile	1.7	1.0
	2nd Quartile	10.4	8.6***
	3rd Quartile	13.7	9.4***
	Top Quartile	17.5	14.9***

注) *** $p < .001$

実際に平均78位のところ、推定が60位、高校生は平均86位のところ、推定が59位であった。これらの差は統計的に有意であった(中学生： $t(101) = 10.8, p < .001$ ；高校生： $t(110) = 22.6, p < .001$)。

次に、各群でのテスト順位の推定と実際の順位の差について、中学生、高校生という学校種による違いが見られるのかを検討するために、学校種(中学、高校)と群(パーセンタイル順位による4つの群)を要因とした2要因の分散分析を行った。その結果、学校種と群の交互作用が有意であった($F(3, 719) = 3.2, p < .05$)。単純主効果の検定を行ったところ、Bottom Quartileでは推定と実際の順位差の差は有意でなかったが(中学生： $M = 20.5$ ，高校生： $M = 20.7$)，その他の群では有意であり，3rd Quartile以上では中学生における差のほうが高校生よりも小さいという結果であった(2nd Quartile 中学生： $M = 11.8$ ，高校生： $M = 5.7$ ；3rd Quartile 中学生： $M = -3.9$ ，高校生：

$M = -11.3$ ；Top Quartile 中学生： $M = -18.8$ ，高校生： $M = -28.7$)。

テスト得点の推定と実際の得点の差

YLEのテスト得点について推定と実際の得点の差について検討を行うために、得点のパーセンタイル順位により4つの群に分け、得点の推定と実際の得点の平均を算出した。結果をFigure 3と4に図示した。

Figure 3および4より、中学生、高校生ともにBottom Quartileの群においては推定と実際のテスト得点に乖離は見られなかった。しかし、2nd Quartileより上位の群については、実際にテスト得点よりもテスト得点の推定は下回っており、上位の群はその乖離が大きくなっていった(Table 2参照)。

得点の推定と実際の差についても、学校種(中学、高校)と群(パーセンタイル順位による4つ

の群)を要因とした2要因の分散分析を行った。その結果、交互作用効果は有意ではなかった ($F(3, 719) = 0.8, ns$)。

自己評価のずれと英語学習動機づけの関連

ここまでの分析により、英語のテスト得点の実際と推定(自己評価)との間には相対的(パーセントイル順位)にも絶対的(テスト得点)にも成績の上位群、下位群によって差(ずれ)が見られることが明らかにされた。これらのずれが学習及びその成果にもたらす影響を調べる一環として、まずこの実際と自己評価とのずれと英語学習動機づけとの関連を検討した。具体的には順位の差(順位の推定-実際の順位)、および得点の差(実際の得点-得点の推定)と英語学習動機づけ尺度の得点との相関係数を算出した。その際、ずれのあり方には成績の上位群と下位群とで違いが見られたことから、上位群(Top Quartileと3rd Quartile)、下位群(Bottom Quartileと2nd Quartile)を設定し、それぞれ相関係数を算出した(Table 3)。

Table 3 テストの順位と得点の推定と実際の差と英語学習動機づけとの関連

		下位群	上位群
中学生	順位差	.14	.21**
	得点差	.12	.24**
高校生	順位差	.03	.08
	得点差	.10	.26**

注) ** $p < .01$ 中学生 上位群 $n = 171$, 下位群 $n = 135$,
高校生 上位群 $n = 207$, 下位群 $n = 205$

相関係数を見ると、中学生も高校生も上位群において有意な関連が示された。中学生は推定と実際の順位差も得点差も有意な正の関連を示した。この結果は、自身の順位を実際よりも高く、テスト得点を高く見積もることが、高い動機づけを持つとも考えられる。しかし、順位差と学習動機づけの散布図(Figure 5)を見てみると、上位群は自身の成績を順位も得点も低く見ていることから、大部分が負の値を示しており、自身の実際の成績に近づくことが学習動機づけにつながるといっても解釈される結果であった。得点差についても同様の傾向であった。また、高校生の得点差と学習動機づけの散布図(Figure 6)を見ても、やはり同様の傾向が見て取れる結果であった。

順位と得点の推定と実際の差とテスト成績変化との関連

順位と得点の推定と実際の差が何をもたらすかを検討するためのさらなる指標として、本調査が行われた事後(およそ4ヶ月後)に行われた同レベルのテスト結果が得られた調査対象のデータを用いて両者の関連を検討した。その際、学習動機づけの検討と同様に事前のテスト結果による上位群、下位群の区分を設定して、事前事後のテスト得点の差(大きいほど事後のテスト得点が高くなるよう得点化を行った)との間の相関係数を算出した。結果をTable 4に示した。

有意な相関を示したのは、中学生、高校生ともに上位群のみであった。自身の順位を実際よりも高く、事後のテストの得点が高くなる方向への変

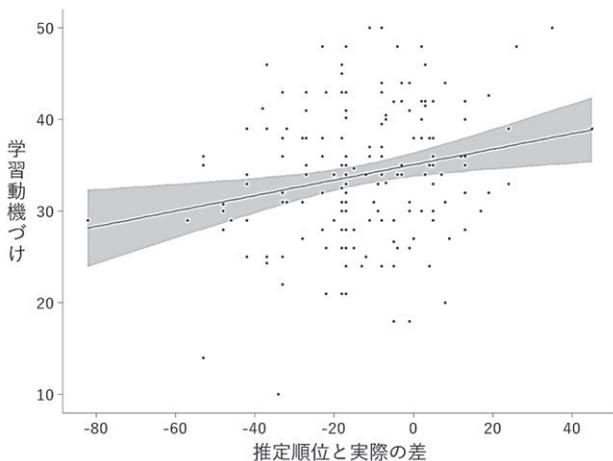


Figure 5 中生成績上位群の順位差と動機づけとの関連

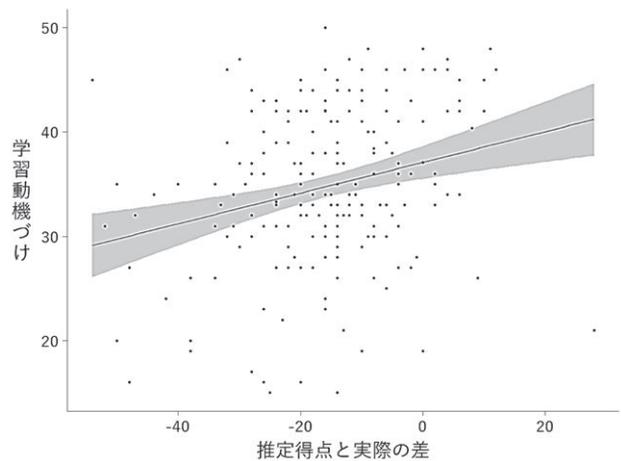


Figure 6 高生成績上位群の得点差と動機づけとの関連

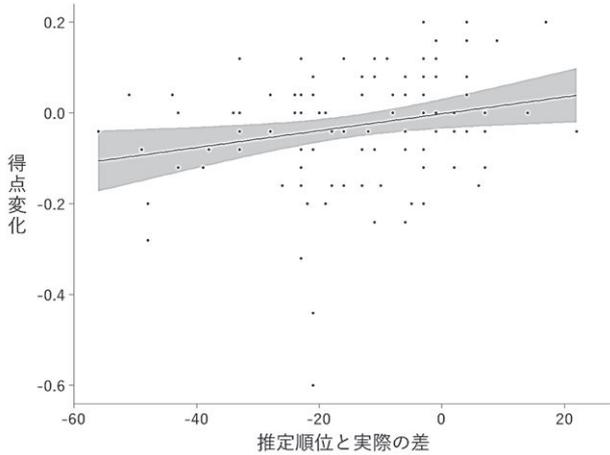


Figure 7 中学生上位群の順位差とテスト得点変化との関連

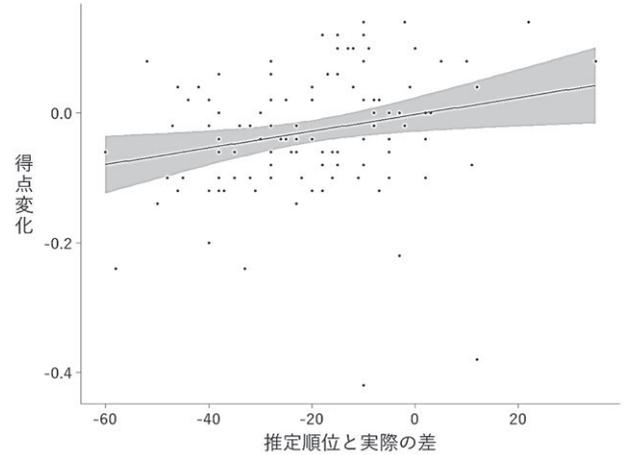


Figure 8 高校上位群の順位差とテスト得点変化との関連

Table 4 テストの順位と得点の推定とテスト得点の変化(事前事後)の関連

		下位群	上位群
中学生	順位差	.21	.23*
	得点差	-.12	-.11
高校生	順位差	.06	.23*
	得点差	.20	.14

注) $p < .05$ 中学生 上位群 $n=107$, 下位群 $n=80$,
高校生 上位群 $n=118$, 下位群 $n=82$

化と関連が見られたと考えられる。ただし、学習動機づけの分析と同様に両者の散布図を作成したところ (Figure 7, 8), 多くが自身の順位を実際よりも低くみており, 自身の実際の順位に近づくことが, 事後のテスト得点を高める方向の変化とつながるといふようにも解釈される結果であった。

IV 考察

本研究では, 中高生の英語パフォーマンスについての自己評価が実際のものとどのくらい一致しているのか (あるいは, どの程度, どのようにずれているのか) を検討するために, 英語のテストの実施と自己評価を尋ねる調査を行い, 両者の関係について検討を行った。まずは, 本研究の主要な仮説に従って考察を進めていく。

テスト順位, 得点の推定と実際の差について

中高生の自己評価と実際の結果との関係はどの

ようなものであったのだろうか。まず言えることは, 先行研究と同様に, 両者の間には中程度の相関が見られたということである。すなわち両者はまったく乖離したものというわけではない。ただ, 本研究ではずれのあり方にも着目し, Dunning-Kruger効果の研究パラダイムに則り, テスト得点から算出したパーセンタイル順位に基づき4つの群にまとめ, 順位や得点の推定と実際の差の検討を行った。その際, 特に注目されるのが成績下位25%の群 (Bottom Quartile) である (仮説1)。順位についての結果を見ると, 中学生においても, 高校生においても, 実際の順位よりも高い順位を推定しており, 過大評価を行っていた。これは仮説1を支持するものである。すなわち, 日本の中高生の英語のパフォーマンスにおいても, Dunning-Kruger効果の主要な点が結果として見られた。

ただし, その過大評価の状況はDunningらの研究グループで示された知見の多くとは異なっている点も指摘される。具体的には, 過大評価の幅が狭く, 中学生も高校生も50位を超えておらず, 40位程度にとどまっていた。Dunningらの結果の多くは50位を超えて推定されている場合が高く, 一番低い群でも平均以上と評価していたことが示されている。また, このような平均以上効果が見られないのは調査対象全体に言えることであり, 全体の平均でも50位を下回っていることから, 本研究での調査対象はかなり控えめに自己評価を行っていたということである。

他方, パフォーマンスの良かった群においては

どうだったであろうか。特に上位の群においては、順位についても過小評価することが予測された(仮説2)。結果は仮説を支持しており、中学生ともに、最上位25%の群 (Top Quartile) で顕著に過小評価を行っているという結果であった。Dunningらの研究グループでも過小評価は見られており、一致する結果である。

ただし、Dunningらのグループで示された結果においては、その過小評価の幅は下位群と比べると小さくなく、その点は、本研究の結果とは一致していない。日本の小中高生の英語の自己評価と英語力との関係をまとめた米田他 (2013) の結果は、比較的上位群は多くが過小評価をしていることが示されており、本研究と同様の結果であると考えられる。そのため、能力を測るためのテストの題材が英語であったということが、Dunningらの結果との違いを生み出したとも解釈される。

順位についての自己評価と実際の差については仮説を支持する結果であったが、得点についての結果は予測と異なるものであった。成績下位群においては、得点についても過大評価が行われ、成績上位群においてはある程度正確に自己評価が行われるという予測であったが、得られた結果は反対の結果であった。すなわち、成績下位群は正確に自己のパフォーマンスを評価しており、成績上位群は実際の得点と一致せず、順位と同様に過小評価を行っているという結果であった。この結果は、Kruger & Dunning (1999) の結果とも一致していない。

得点の自己評価については仮説や先行研究の知見と一致しないという結果であったが、順位についても、成績下位群の過大評価が比較的小さく、成績上位群の過小評価が大きいという一致しない結果も含めて解釈する必要がある。元々、Kruger & Dunning (1999) においては、成績上位群は自身のパフォーマンスについて、その正誤の認識が働くため、絶対的な評価においては正確になることが示されている(成績下位群は逆にその判別ができない)。同様の結果が本研究で示されなかったのは、本研究で用いられた能力・課題の性質(すなわち英語力)によるかもしれない³。すなわち、英語のテストについての中高生の苦手意識が反映されたことが考えられる。例えば、英語に対する

中高生の苦手意識は、ベネッセ総合教育研究所 (2014) の調査でも示されており、テストに対する自己評価も全体的に過小評価を行う方向に力が働いたということもありうるであろう。その点については、英語パフォーマンスの調査対象全体において平均以上効果が見られなかったという結果から推測することができる。そして、英語が苦手であるという自己観を持つのであれば、自身のテスト時のパフォーマンスについては正誤の判別が正確に働きながらも、テスト内での体験が自己観に合わせたように認識されることもありうる (Cricher & Dunning, 2009; Ehrling & Dunning, 2003)。このようにして、特に上位群において順位の評価についても絶対的な評価が全体に低くなったことが反映されたと考えられる。ベネッセ総合教育研究所 (2014) の調査結果では、中学生よりも高校生においてより英語を苦にしているさまがうかがえ、本研究での成績上位群の過小評価が中学生よりも高校生でより大きかったという知見とも整合するものである。成績下位群において見られた順位における過大評価については、元々社会的比較がうまく機能しないところに (Kruger & Dunning, 1999), フォールス・コンセンサス効果が働き、他者も問題を多く正解できないと推測することで、相対的な順位が高まったと考えられる。しかしながら、この解釈は、絶対的な評価の正確さも他者の情報(社会的比較)、そしてフォールス・コンセンサス効果 (Ross, Greene, & House, 1977) の働きについても、上位群、下位群で先行研究の知見、解釈とは異なるものになっている。すなわち、パフォーマンスの乏しいものは、成績が低いだけでなく、パフォーマンスの相対的な低さには気づかない。そして、パフォーマンスの良好なものは自分の成績の良さを認識しておらず、周りの他者を過大に評価してしまっている。

本研究の得られた結果が、メタ認知スキルや社会的比較のあり方など心理的メカニズムの観点から確かなものと言えるかはさらなる検討が必要である。また、自身の能力について尋ねられるという調査内容であったことから、日本人において比較的多く見られる自己卑下傾向 (Heine, Takata, & Lehman, 2000) の影響といった別の観点から解釈されるものであるかもしれない。

動機づけ、事後のテスト得点との関連

本研究では、探索的にではあるが、英語のパフォーマンスに対する自己評価の正確さと学習動機づけ、事後に行われる英語テストとの関連が見られるのかについて検討を行った。両指標について、特に成績上位群において、自己評価と実際の差の大きさと学習動機づけ、事前事後のテスト得点の変化との間に有意の正の相関が見られた。ここで注意をしなければいけないのは、成績上位群は多くが自己評価として過小評価を行っていたという点である。本研究の結果は、例えば、学習動機づけとの間に正の相関が見られたという点について、過大評価を行うことで学習動機づけも高くなるという結果とも解釈可能ではある。しかし、実際のところは過小評価を行っている人よりも自身の実際の順位（あるいは得点）と一致する方向で認識することが動機づけの高さにつながると解釈することも可能であり、より妥当であるかもしれない。これは、自己評価と実際の差とテスト得点（その上昇）との関連についても同様の解釈が可能であると考えられる。これらの解釈の妥当性は、Zimmerman & Moylan (2009) の循環的段階モデル (cyclical phase model) から考える。同モデルは、自己調整的フィードバックのモデルとされ、動機づけなどが含まれる予見段階と、遂行段階、そして自己評価や原因帰属を含む自己内省段階が置かれている。それぞれの段階が循環的に影響して学習が進められていく。自己評価が適切に行われることが、自身の強みや弱みを把握し、動機づけと学習への取り組みへとつながるのであれば、その結果としての成果につながるということになる。そのようなプロセスが働くというのであれば、成績下位群においては、動機づけにも事後のテスト結果との間にも有意な相関が見られなかったという点も重要である。これらの点については、後の節において改めて触れたい。

自己評価についての中高生の差

本研究では、中高生を対象に調査を行ったため、その発達の違いについても視野に入れた。本研究の中高生の年齢差はおよそ3歳であった。メタ認知スキルの発達の变化については、青年期内においても変化が見られることが明らかにされて

おり（例えば、van der Stel & Veenman, 2010）、英語に対する意識にも多少の差が見られている（ベネッセ総合教育研究所, 2014）。しかしながら、全体としては中高生で大きな違いは見られず、成績上位、下位の自己評価の実際とのずれのあり方も動機づけや成果との関連についてもほぼ同様の結果であった。違いとしては、先にも触れた、成績上位群の過小評価が中学生よりも高校生で大きいというように、自己評価が高校生のほうが控えめであったということである。ただし、この結果はメタ認知スキルの違いが生み出したというよりも、英語に対する認識、すなわち高校生の苦手意識の強さや同質性が高い中で社会的比較が課される状況による違いと考えたほうがより妥当であると思われる。自己評価についての発達の違いについての知見を深めるためには、小学生や大学生などより幅を広げ、本研究と同様の検討を行うことが求められる。

本研究によって得られた知見の教育への示唆

本研究によって得られた知見が教育の実践に寄与すると考えられることについてまとめる。まずは自己評価の正確さに向けてどのように介入するかという点である。成績上位群と成績下位群で自己評価の正確さに違いが見られた。成績上位群では、順位について他者のパフォーマンスを過度に高く評価すると同時に、自身の得点を実際よりも低く見積もるといった傾向が示された。Kruger & Dunning (1999) では、他者の情報を提供することで、成績上位群はその情報をうまく活用し、自身の能力の相対的位置づけ修正できたことが報告されている。日本の中高生においてもまずそのような情報を提供することで、推定と実際の一致が期待される。また、絶対的評価としての得点についてはテスト得点のフィードバックが行われることで修正が行われると考えられるが、本研究での解釈のように、英語に対する苦手意識などからネガティブな自己観を有している場合は、ただ点数をフィードバックするだけでは、認識が変わりにくいことも考えられる (Crichter & Dunning, 2009; Ehrling & Dunning, 2003)。その場合、テスト得点だけではなく、自己の能力観についてのあり方にまで配慮する必要があるだろう。

成績下位群では特に順位について、つまり相対的な評価の部分に過大評価が見られた。また、そのような過大評価は他の学習プロセス（例えば学習動機づけなど）に寄与するようなポジティブな作用が得られていない。それは他者からは過度に楽観的になっているだけと捉えられるが、それがメタ認知スキルの不足が背景にあるということであれば、その点についての介入が必要であろう。ただし、メタ認知スキルの不足により自己評価にずれがある場合、フィードバックが有効に働かない（フィードバックが正確でない、自身との関連性が薄いと認識される）という知見（Sheldon, Dunning, & Ames, 2014）もあるため、その点を考慮した形でのフィードバックを行うことが重要である。

本研究の限界と今後の課題

本研究では、中高生を対象に英語を題材として自己評価の正確性をDunning-Kruger効果パラダイムに則り検討を行い、いくつかの点で違いはあるものの、全体として同様の結果が得られた。日本においては同様の検討はほとんどされていないことから、本知見には一定の意義があると思われる。ただし、その結果について基本的に心理的メカニズムの観点から解釈を行ったが、この効果はアーティファクト（例えば、平均への回帰、課題の難易度の観点）による結果ではないかという議論も存在する（例えば、Burson, Larrick, & Klayman, 2006; Krueger & Mueller, 2002）。本研究では、Dunning-Kruger効果と同様の現象が見られるかという点にまず着目したため、その原因については十分検討がなされていないことは留意する必要がある（注の3も参照）。

また、ずれのあり方に先行研究といくつかの違いも見られたこともあり、自己評価と実際のずれについての成績上位群と下位群の心理的メカニズムについて、さらに詳細な検討が必要であろう。その際、Kruger & Dunning（1999）などにある実験的操作を加えた検討を行うことで、メカニズムに迫ることが可能となり、それは、さらなる教育的示唆を得ることにつながることも期待される。

さらに、その基本的な心理メカニズムの検証だけでなく、英語以外の他の教科、あるいは他の様々

な能力でも同様の検討・分析を積み重ねることにより、本研究での知見が英語というものに特有であるのか、あるいは、文化的要因が働いているのかを明らかにしていくこともできるであろう。

〈注〉

- 1 本研究は、日本教育心理学会第60回総会において発表した内容を再分析したものである。
- 2 ここで示した結果はリーディングについてのものがあるが、ライティングについても同様の結果であった。
- 3 メタ認知スキルといった心理的メカニズムの問題というよりも、課題の難易度により、Dunning-Kruger効果のような自己評価のずれが生じるという知見もある（Burson et al., 2006）。具体的には、課題が容易なときには、全体的に自身のパフォーマンスを高く評価し、成績下位群は過大評価を行う。過大の難易度が高い場合には、自身のパフォーマンスを低く評価し、成績上位群の方が自身のパフォーマンスについての推定に困難が生じるという主張である。この主張、すなわち課題の難易度の観点は、本研究の結果を解釈するには特に、傾聴に値する部分があると思われる。

〈謝辞〉

本研究は平成23年度科学研究費助成事業基盤研究(C) 課題番号23520765（研究代表者：米田佐紀子）において行われた研究の一部である。本論文の執筆に際しデータの使用をご承諾いただきました、米田佐紀子先生（玉川大学）に感謝申し上げます。

〈引用文献〉

- ベネッセ総合教育研究所（2001a）. 第3回学習基本調査報告書・高校生版 Retrieved from <https://www.crn.or.jp/LIBRARY/GAKUSHU/KOU.HTM>（2018年9月29日）
- ベネッセ総合教育研究所（2001b）. 第3回学習基本調査報告書・中学生版 Retrieved from <https://www.crn.or.jp/LIBRARY/GAKUSHU/CYU.HTM>（2018年9月29日）
- ベネッセ総合教育研究所（2001c）. 第3回学習基本調査報告書・小学生版 Retrieved from <https://www.crn.or.jp/LIBRARY/GAKUSHU/SYOU.HTM>（2018年9月29日）
- ベネッセ総合教育研究所（2014）. 中高生の英語学習に関する実態調査 2014 Retrieved from https://berd.benesse.jp/up_images/research/Teenagers_English_learning_Survey-2014_ALL.pdf（2018年9月29日）

- Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology, 90*, 60–77.
- Critcher, C. R., & Dunning, D. (2009). How chronic self-views influence (and Mislead) self-assessments of task performance: Self-views shape bottom-up experiences with the task. *Journal of Personality and Social Psychology, 97*, 931–945.
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*, 69–106.
- Dunning, D. (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. *Advances in Experimental Social Psychology, 44*, 247–296.
- Dunning, D., Johnson, K., Ehrlinger, E., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*, 83–87.
- Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology, 84*, 5–17.
- Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior & Human Decision Process, 105*, 98–121.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*, 395–430.
- Hansford, B. C., & Hattie, J. A. (1982). The Relationship between self and achievement/performance measures. *Review of Educational Research, 52*, 123–142.
- Heine, S. J., Takata, T., & Lehman, D. R. (2000). Beyond self-presentation: Evidence for self-criticism among Japanese. *Personality and Social Psychology Bulletin, 21*, 1200–1209.
- Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology, 82*, 180–188.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessment. *Journal of Personality and Social Psychology, 77*, 1121–1134.
- Mabe, P., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 17*, 280–296.
- 物井尚子 (2016). 高学年児童の英語運用能力と自己評価の関係日本児童英語教育学会 (JASTEC) 研究紀要, 35, 1–16.
- Myntinen, S., Sundström, A., Vissers, J., Koivukoski, M., Hakuli, K., & Keskinen, E. (2009). Self-assessed driver competence among novice drivers – a comparison of driving test candidate assessments and examiner assessments in a Dutch and Finnish sample. *Journal of Safety Research, 40*, 301–309.
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Psychology, 13*, 279–301.
- Sheldon, O. J., Dunning, D., & Ames, D. R. (2014). Emotionally unskilled, unaware, and uninterested in learning more: Reactions to feedback about deficits in emotional intelligence. *Journal of Applied Psychology, 99*, 125–137.
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66–73.
- van der Stel, M., & Veenman, M. (2010). Development of metacognitive skillfulness: A longitudinal study. *Learning and Individual Differences, 20*, 220–224.
- 米田佐紀子・細川真衣・西村洋一・物井尚子 (2013). 小・中・高・大学生の自己評価と英語力に関する研究—CEFRに基づくCan Doリストとケンブリッジ英検模試を用いて—中部地区英語教育学会紀要, 42, 131–138.
- 米田佐紀子・西村洋一・細川真衣 (2014). 公立中学校における英語学習ポートフォリオ使用の効果と課題—石川県と千葉県における3年間の追跡調査から—北陸学院大学・北陸学院大学短期大学部研究紀要, 7, 219–233.
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation interest. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Handbook of metacognition in education*. New York: Routledge. pp. 300–305.