討 論

学力の評価と測定をめぐって

大 塚 雄 作 柴 山 直 植 阪 友 理 遠 藤 利 彦 野 口 裕 之
(京都大学) (東北大学) (東京大学) (東京大学) (名古屋大学)

On the Evaluation and Measurement of Academic Achievement

YUSAKU OTSUKA, TADASHI SHIBAYAMA, YURI UESAKA, TOSHIHIKO ENDO AND HIROYUKI NOGUCHI
(KYOTO UNIVERSITY, TOHOKU UNIVERSITY, UNIVERSITY OF TOKYO, UNIVERSITY OF TOKYO AND NAGOYA UNIVERSITY)

Amid current developments in high school/university articulation reforms, we frequently experience a sense of lacking a shared, sufficient understanding of how to frame and evaluate “academic achievement.” This arises because we fail to consider differences in the purposes of evaluation and measurement for surveys and selection examinations, and simply evaluate non-cognitive factors of academic achievement (e.g., motivation, perseverance, self-control) in student selection. For evaluation and measurement, it is important to utilize methods appropriate for the corresponding aims and subjects as required by circumstances. This should be the standard for not only high school/university articulation reforms but also educational psychology research. In light of these requirements, the present paper discusses the following points while introducing research cases from the educational psychology research field: (a) Separate use of individual and group scores, data collection design, and so on, in large-scale academic achievement surveys to guarantee the integrity of aims and design specifications; (b) Formative evaluation that is effective in everyday practice of school education; (c) Research trends and tasks related to the “non-cognitive” aspects of the mind, nurtured early in life and thought to influence child development.

Key Words: evidence of academic achievement, formative evaluation, non-cognitive factor

現在進められている高大接続改革の進展の中で、「学力」をどう捉え、どう評価すべきかといった基本的な部分で、十分な理解が共有されているとは言い難いことをしばしば経験する。調査と選抜試験という評価・測定の目的の違いが安易に軽視されたり、形成的評価に適合する手法や、子どもに必要とされる非認知的要因などの評価が、短絡的に選抜における学力評価などに適用されようとしたりする。

評価・測定は、目的や対象にふさわしい評価手法を状況に応じて選択するということが重要であり、それは高大接続改革のみならず、教育心理学研究においても基本とすべきことである。本討論では、以下の諸点に関して、教育心理学研究の領域における研究事例を紹介しつつ論じていくこととする。(a) 大規模学力調査において、目的と設計仕様との整合性の担保、個人スコアと集団スコアの使い分け、データ収集デザイン等が重要という点について。(b) 日常的な学校教育実践において、どのような形成的評価が有効に機能するのかについて。(c) 子どもの発達に影響を及ぼすと思われる人生早期に培われる「非認知」的な心の性質に関わる研究動向と課題について。

キーワード：学力のエビデンス、形成的評価、非認知的要因

討論の趣旨と論点

大塚雄作

高大接続改革の論議

近年、「高大接続改革」論議が巻き起こり、その一環として、大学入試の改革、学習指導要領の改訂など、教育領域における大きな動きが起ころうとしている。その議論の核に据えられている考え方の一つに、いわゆる「学力の3要素」と呼ばれる概念がある。これは、学校教育法に記載された学力の捉え方に基づくものであり、その第30条第2項に、「生涯にわたり学習する基盤が培われるよう、基礎的な知識及び技能を習得させるとともに、これらを活用して課題を解決するために必要な思考力、判断力、表現力その他の能力をはぐくみ、主体的に学習に取り組む態度を養うことに、特に意を用いなければならない。」と記されている。これは、平成19(2007)年に追加されたものであり、また、小学校の章に書かれているものであるが、中学校、高等学校にも準用されることが明記されている。その文中の「知識・技能」、「思考力・判断力・表現力」、「主体的に学習に取り組む態度」という3つが取り上げられて、「学力の3要素」と呼ばれているのである。なお、「主体的に学習に取り組む態度」については、高大接続改革に関わる中央教育審議会(2014)では、「主体的に多様な他者と協働して学習する態度」と表現され、「主体性・多様性・協働性」と呼ばれてもいる。

大学入試改革の議論では、この「学力の3要素」に関わって、今までの大学入試が「知識・技能」の測定に偏っているという前提の下に、「思考力・判断力・表現力」の測定をより重視すると共に、「非認知的」な要素とも言える「主体性・多様性・協働性」も含めて、「学力の3要素」についてバランスよく評価すべきということが強調されてきている。

学力をいかに評価するかということは、大学入試にとどまらず、発達過程の諸相における教育的営みに欠くことのできない課題でもあり、教育心理学の重要な研究テーマにも位置づけられてきている。しかし、大学入試改革の議論においては、学力の評価・測定に関する基本が一般には十分に共有されていないのではないと思われる場面がしばしば散見される。それが、本誌上討論の企画の背景になっている。すなわち、教育心理学における研究において、学力の評価・測定に関わって、どのような状況や対象が取り上げられ、どのようなアプローチがなされてきているのかをあらためて振り返ってみることで、大学入試改革における問

題点を間接的に浮き彫りにしてみることはできないかということである。それはまた、学力に関わる教育心理学的研究においても留意すべき基本でもあることから、日本教育心理学会員の研究や教育実践にも有意義な知見を提供することになることも期待される。そこで、本誌上討論においては、いくつかの状況における学力に関わる評価・測定について紹介すると共に、それぞれの課題を整理することを通して、教育心理学的研究における学力評価のあり方を再考する機会を提供することを本旨としておくことにしたい。

調査と選抜試験の違い

現在の大学入試改革の流れの中では、「思考力・判断力・表現力」の評価・測定が特に求められ、そのための試験問題のサンプルが示されてきているが、その際にしばしば感じられることに、調査と選抜試験の違いが共有されていないということがある。

例えば、「思考力・判断力・表現力」を測定する入試問題の参照すべき例として、PISAや全国学力・学習状況調査の問題が取り上げられることが少なくないのであるが、それらはあくまで調査であって、個人の学力のレベルを識別してそれによって選抜しなければならない入学試験とは目的が異なっている。問題が出されてそれに解答するというテストの形式を両者とも同様に採用しているとしても、調査では基本的に集団の統計量によって「集団」の学力レベルを表現するのが目的となるのに対して、選抜試験では「個人」の学力レベルを識別する必要がある。

信頼性という観点では、集団の統計量は標準誤差によって誤差の大きさの推定がなされるのに対して、個人の得点に関してはテスト得点のモデルに基づいて推定される信頼性係数などが指標として利用されることになる。ある集団における統計量の標準誤差を小さくするためには、例えば、サンプルサイズを大きくすることがまず考えられることの一つであるのに対して、個人の得点に関しては、それに対応して問題の数を増やすことが考慮されるべきことの一つとなる。

個人の得点の信頼性が高くなれば、集団としての統計量も安定するという関係はあるにしても、調査であれば、少数の問題であったとしても、サンプルサイズも含めて適切にサンプルが抽出されていれば、集団の統計量の標準誤差を小さく抑えることがそれなりに可能となる。一方、とりわけ大規模な選抜試験では、選抜に資するために、限られた時間内に一定の問題数を課すことが求められることになり、その制約の下で少数の問題をじっくり考えさせるような試験の得点は信

頼性の観点からは不向きとなる場合もある。

また、公平性という観点からしても、少数の問題において、特定の社会的文脈における問題解決が求められる場合には、その文脈に馴染んでいるかどうかなどによって差異が生じる可能性も大きくなる。

したがって、大規模な共通試験などでは、一定の問題数の確保が望まれることになるが、この種の吟味は、入試改革に関わる議論の流れにおいては十分に押さえられているとは言いがたい。

そこでまず、学力測定に関して、規模の大きい調査や選抜試験を実施する際に、測定論的視点からどのような点に留意したらよいか、柴山直氏に測定論の立場から論じていただくことにする。

形成的評価と総括的評価の違い

わが国でも、教育課程審議会（2000）の答申に「指導と評価は表裏一体をなすものであり、学校においては、学習指導と評価が常に一体となって行われることが求められる。」と明記されるなど、教育実践において、「指導と評価の一体化」ということが求められるようになっていく。その際に適する評価手法として、「オーセンティック・アセスメント（authentic assessment）」と呼ばれる、現実の活動にできるだけ近い状況の下での一連の評価手法の活用が推奨されてきている。そこでは、筆記試験の形をとるよりも、実技系のパフォーマンスを引き出す評価手法や、筆記試験であっても、実際に経験しやすい、より実生活に近い文脈が設定され、そこでの何らかの問題解決が問われたりすることが多い（パフォーマンス評価）。このような流れは、大規模共通試験においても例外ではなく、マークシート問題ではなく記述式問題が、さらに、英語などではスピーキングなどのパフォーマンス型評価の導入の要請が強くなっており、それに応じての改革が模索されているところである。

しかし、元来、評価・測定の手法は、ある目的の達成のために開発されてきているものであって、例えば、教育実践の文脈の中で指導や学習の改善のために向いている評価手法が、必ずしも、大規模に行われる選抜試験などに有効とは限らない。言い換えれば、教育のプロセスの中で利用される形成的評価（formative evaluation）に有効な評価手法が、必ずしも、選抜をはじめとする総括的評価（summative evaluation）にも同じように有効というわけではないということである。

例えば、論述式問題は、大学入試においても各大学の個別試験において広く採用されているし、教育実践における総括的評価に分類される試験にも多く活用さ

れており、書かせることの重要性は誰も否定するものではないが、その際のポイントは、教授者と学習者との相互作用することができる部分にあるということに留意しておく必要があるだろう。ある問いに関して、どういうレベルでどのような視点から解答すればよいかは、出題する側と解答する側で、暗黙のうちにその問いに関するある特殊な文脈が共有されていることが肝要であるということである。論述をどのレベルで、どの視点から行うかは、そのような文脈が何も共有されていない場合には非常に多岐にわたるはずのものであって、その多様性のゆえに、採点基準なども一意には定めることが難しいということになりがちである。その点で、多様な受験生を対象とする大規模共通試験には、論述式の問題が必ずしも適当とは言えないことが帰結されるであろう。

形成的評価において重要なことの一つに、評価情報のフィードバックということをおこなうことができる。何を教え、何を学んだらよいのかといったことが普段の教授・学習の実践の中で自ずとある程度共有されていることで、フィードバックがより的確に機能していくことになる。大規模試験での記述式の導入は、そういう記述式本来の機能を殺ぐことにもなりかねない。

そこで、形成的評価がどのように活用されることによってその機能を有効に発揮しうるかということをおもてて共有するという意味で、教育実践において、教師と生徒との相互作用の中で活性化される教授・学習活動の事例を通して、教育プロセスにおいて有効となりうる形成的評価のあり方について植阪友理氏に論じていただくことにする。

非認知的能力の評価

今般の大学入試改革では、各大学の個別試験において、学力の3要素のうちの「主体性・多様性・協働性」に関わる「非認知的」要素に関する評価を実施することが求められている。しかし、教育心理学研究に携わっている立場からすれば、この種の非認知的要素に関する評価は、測定ツールを開発するだけでも困難であるということのみならず、大学の個別試験であったとしても、比較的規模の大きい受験者数を対象とせざるをえない状況においては、人的にも予算的にも効率的かつ公平に実施することが難しいということも存外議論の遡上に上らない。

そもそも、「主体的に学習に取り組む態度」などは、とりわけ、大学のレベルではすべての学習領域において要求すべきものでもなく、個人個人、領域固有のことであってよいものである。また、ハイスティクスな

選抜試験において評価するという事になったときに、
 どういう波及効果が起こりうるのかということを見ると、
 受験産業などが準備するであろうハウツーの浸透は
 当然予想されることであり、「主体性」等を測定する
 ために開発された測定・評価ツールの持つ当初の妥
 当性は、おそらく年ごとに劣化していつてしまうこと
 にもなるであろう。

大学教育において、学生が主体的に学ぶ態度は、そ
 れが前提となって教育が進められていくことにもなる
 重要な要素と言えるが、それは高校において初めて身
 に付くというものではなく、大学入試による選抜とい
 う目的においてその点を評価すべきかどうかは議論の
 余地が残るところであろう。その種の非認知的要素は、
 むしろより幼少期から形作られていくべきものであ
 って、その主体的な姿勢があるからこそ、思考したり、
 判断したり、表現したりということ積み重ねることが
 でき、より深い理解を伴う知識へと醸成されていく
 べきものである。すなわち、大学入試の段階では、実
 施可能性という点も含めて考えれば、そこまで形作
 られた知識のレベルを問うことに専念すべきという見
 方もできるのである。

そこで、この種の非認知的要素に関して、実際に開
 発された評価手法と、それがどの年齢対象について実
 施されてきているのか、また、それらの評価に関わる
 課題などについて遠藤利彦氏から話題提供していただ
 くことにする。

以上の観点だけで、学力の評価・測定に関する議論
 が十分であるとは言えないが、ここに取り上げた観点
 における課題を共有するだけでも、教育心理学研究そ
 のものにも大きな示唆が得られることであろう。さら
 に、野口裕之氏に最後に総括していただくことを通し
 て、学力の評価・測定に関わる課題を共有し、よりよ
 い教育改革へと結びつけていくことができれば、それ
 が地道で時間がかかることであったとしても、本誌上
 討論の企画者として幸甚至極なことである。

引用文献

- 中央教育審議会 (2014). 新しい時代にふさわしい高大
 接続の実現に向けた高等学校教育、大学教育、大学
 入学選抜の一体的改革について (答申) (中教審第
 177 号) 文部科学省
 教育課程審議会 (2000). 児童生徒の学習と教育課程の
 実施状況の評価の在り方について (答申) 文部科学
 省 初等中等教育局小学校課

大規模学力アセスメント—大学入試センター 「共通テスト」への批判的考察

柴山 直

本稿では、最初で大規模学力アセスメント (large scale
 assessment) をめぐる昨今の社会的背景・動向を考察し、
 あわせて混乱して使われていると思われるいくつかの
 用語の整理を試みる。その上で、すべてのテストには
 目的に応じた設計・開発・実施方法がそれぞれあるこ
 と、またその際に考慮すべき諸点を強調する。さらに
 テスト開発 (test development) の重要な構成要素である
 項目作成 (item development) の際に見落とされがちな、
 それゆえ、それらを知らなければ不必要な混乱と無駄
 なコストと時間、そして何よりも受験者達に不利益を
 招く、長年にわたって蓄積されてきたノウハウのいく
 つかを、特に多肢選択形式に絞って指摘する。最後に
 本稿執筆時 (2017 年 12 月時点) において現在進行形で大
 学入試センターが進めている共通テストの試行試験の
 課題を、上で述べた技術論的な観点から検討する。な
 お本稿では通常、「問題」と呼ばれるものを専門用語で
 ある「項目 (item)」と呼称することにする。

大規模学力アセスメント利用への社会的需要

社会的背景であるが、グローバル教育産業 (Global
 Educational Industry) という言葉 (OECD, 2015) に象徴さ
 れる大規模学力テスト (large scale testing) およびそれを
 サポートするテクノロジー利用の流行の背景を眺めると、
 人口減少、財政難、職業スキルの抽象化の 3 つの
 キーワードに行き着く。2100 年には明治維新の水準ま
 で落ち込みかつ高齢化率約 40% の推計 (国土審議会政
 策部会長期展望委員会, 2011) もある急激な人口減少に対
 しては、これまでの競争原理にもとづく人材確保ではな
 く、質保証の観点からの人材養成が急務となっている。
 やや牽強付会ではあるが、競争原理を内包する相対評
 価から質保証重視の到達度評価への流れにそれを見る
 こともできる。

また、財政難に対しては、納税者への説明責任の観
 点からの公共財投入の際の費用対効果の見積もり、し
 たがって教育であっても何らかの形で数値による評
 価の導入が社会から強く要請されるようになっている。
 これは PDCA サイクルという用語の頻出にみられるよ
 うに、教育からの発想というよりも、企業人の間で標
 語的に語られることの多い「数値で表現できないもの
 は管理できない。管理しようと思えば数値化せよ」と
 いうきわめて経営学的な信念からの影響が大きいと考
 えられる。

さらに、米国の例ではあるが、知的能力を要する職業に就いていた人々の割合が1900年には全就労者人口に占める割合で3%だったものが、今日では35%にまでなり、ますますその割合が増加している (Flynn, 2013)。社会の進展にともない、いわば職業スキルの抽象化とでもいうべき現象がグローバルに起こっていることがここでは指摘できる。

この流れの中で、皮肉な意味で、テストに対する、人間の知的能力を数値化できる手軽な用具 (測定ツール; assessment tool) としての認識の高まりがある。同時に、項目反応理論等の普及、IT 技術等の発展にともない、学力の測定技術自体の高度化・複雑化・大規模化が相まって起こり、技術力・資本力をもった企業の活用、すなわち文部科学省のいうところの「民間活用」の動きが活発化している。

構成概念としての「資質・能力 (コンピテンシー)」

そのいわば端緒となっているのが、現在準備されつつある次期指導要領改訂の検討プロセスなどで出てくる「資質・能力 (コンピテンシー)」という概念である。国立教育政策研究所などの公刊物 (国立教育政策研究所, 2016) を読んでこの概念定義が筆者にはよく分からない。

「資質・能力 (コンピテンシー)」の概念を整理するには、現在、国際標準化機構 (ISO) で検討中の「学習サービスの成果のアセスメント 手引き アウトライン」 (International Organization for Standardization, 2017) の記述が有用である。その中に、

“Competence: an assessment or set of assessments designed to measure an individual's ability to perform a specific job

Ability: an assessment designed to measure an individual's observable production of a specific work product

Skills: an assessment designed to measure an individual's observable performance of a specific psychomotor act”

というように用語の定義案が示されている。コンピテンシーはコンピテンスと辞書的にはほぼ同義であるが、コンピテンスは心理学の専門用語として、またコンピテンシーは経営学上の用語として使われることが多い。

この定義案からも分かるように、“ability, competency, skill” の順に具体的な直接観察可能なものになると整理されている。また文部科学省が行政文書の中で盛んに使っている「資質の向上」といった表現の際の「資質」を、本来の辞書的な意味、すなわち「生来

の性質や才能」ではなく、たとえば教師としての準備状況や専門的特性を示しているなど、比喩的な意味で使われていると解釈すれば、「資質」に対しては英語の disposition を当てはめることも可能であろう。この場合は「資質・能力 (コンピテンシー)」は “dispositions, ability, competency, skill” とスキルを加えた形で概念の整理ができる。

本稿ではそうした整理の上で、それらを何らかの手段で数値に表現することを測定 (measurement)、価値判断抜きにそれらの実態を調べることをアセスメント (assessment)、測定ないしアセスメントされた結果に何らかの価値判断を加えることを評価 (evaluation) とあらためて呼ぶことを提案したい。

「学力の3要素」とテスト理論の不在

一方、このような社会の動きを受けるかのように、学教教育法第4章第30条第2項 (この項が追加された改正は平成19年6月27日法律第九六号: 施行は平成19年12月26日) では「前項の場合においては、生涯にわたり学習する基盤が培われるよう、①基礎的な知識及び技能を習得させるとともに、②これらを活用して課題を解決するために必要な思考力、判断力、表現力その他の能力をはぐくみ、③主体的に学習に取り組む態度を養うことに、特に意を用いなければならない。」とされている (丸囲み数字は筆者挿入)。ここで述べられていることを根拠に文部科学省では「学力の3要素」 (3つは前出の丸囲み文字に対応) という解釈を打ち出し (中央教育審議会, 2014)、以降、この解釈を貫徹する「明治以来の教育改革」を多方面で推し進めてきている。しかし「学力の3要素」については学術的な背景なしに、また審議会等での議論もなしに、行政主導で導入されたことによる弊害が生じる可能性に関する指摘がある (南風原, 2016)。実際に学校現場でも教科教育の中でそれをどう組み込むのか、あるいは具体的な教授・学習、その結果把握のための測定・評価の際に、いかに3つの要素を切り分けて扱うのか、そもそもそれは不可能ではないのかという観点からの戸惑いなし困惑が早くも広がっていると聞く。

さらに池田 (1970)、Arai & Mayekawa (2005)、柴山 (2008) などによって長年にわたり繰り返し指摘されてきたように、国際水準から見れば「逆さま現象」 (前川, 2007) とでもいうべき、大規模テストを裏で支えているテスト理論およびそれに基づく学力の測定技術への無理解・誤解ないし無視の傾向が、わが国の教育行政にはある。それが特に著しいのがここ数年行政主導で推し進められている大学入試センター試験の「改革」で

ある。テスト理論の利用を無視して大規模テストを実施しようとしているこのような動きをもって、わが国の「大規模共通テストのガラパゴス化現象」と批判されることもある。これらの問題点が集約的に出現してしまっている例が、昨今マスメディア等でもとみに迷走が指摘される、2020年度の共通テストに向けた記述式、英語4技能の導入の動きと見ることも可能である。

大規模学力調査を設計する際のポイント

以上を踏まえれば、大規模学力調査においても、一般的な調査デザインに加えて、学力の測定モデルとでもいべき心理計量モデル (psychometric model) も同時に考慮する必要があるということは自然に導かれる。そのため、学力調査の設計ないし利用にあたって念頭に置くべきポイントは、

- (a) 心理計量モデルの考え方への理解、
- (b) 目的と設計仕様との整合性の担保、
- (c) 個人スコアと集団スコアの使い分け、
- (d) データ収集デザインの設定、
- (e) エビデンスとして必要な強度の確保、

とまとめることができる。

心理計量モデルについては個々の定式化への理解を含めてその基本的な考え方を把握しておくことが重要となる。学力、パーソナリティ、適性などを心理学的特性と総称するならばこれらは実体としては存在しない。そこで、それらを仮に存在するものとして、心理学的モデルに組み上げ、それを介して数量化し個人差を表現する。また心理計量モデルは本質的にはきわめて統計的なモデルであって、認知プロセス自体の解明や記述にはほとんど関心がおかれていないことにも注意が必要である。たとえば項目反応理論モデルの中に出てくる能力母数はあくまでも母数であって、その母数が能力を産出する何らかのプロセスにさらに分解されるということはない。したがって共通テストの設計仕様の中に、たとえば認知プロセスを抽出し指導に活かせるような構成主義的なアプローチ (久保田, 2003) を組み込み、テスト得点によってそれを実現しようとすることは、そもそも次元の違う話なのである。

次に、調査目的あるいは選抜なのか指導なのか、それとも診断なのかといった、実施目的にテストの設計仕様を合致させなければならない。たとえば選抜目的なら合否の分割点付近で個人差を識別できる情報が最大となる設計が必要となろう。しかし現実には相反する目的が一つのテストに課せられることはしばしば生じる。

その一例として、得られたスコアを個人の処遇ある

いは指導に利用するのか、それとも、行政判断のようにある集団を総体として見るのに使うのかの問題がある。全国学力学習状況調査結果の向上策としていくつかの自治体がすでに実施しているように、分布の左裾のあたりに位置する集団の平均を上げる方がそうでない部分の集団の平均を上げるよりも効率的な対策となる。しかしそのような対策は、いわゆる学力が十分でない児童生徒達には良いかもしれないが、全国学力学習状況調査が目的として掲げている、わが国のすべての児童生徒達の学力向上に資する指導とは言い難い。

さらに、一つのテストでカバーできる範囲は時間制限のため、本来測定したい内容量の3分の1程度といわれる。その一方でたとえば子どもたちの学習定着度を領域ごとにきめ細かく測定しようとするならば、データ収集デザインとしては、現行の一冊子一斉実施方式ではなく、IRTモデルにもとづくマトリックス・サンプリング (item-matrix sampling; 重複テスト分冊法) などを採用する必要がある (池田, 2010; 柴山, 2014)。

疫学的アプローチによるエビデンスの獲得

また、いかにエビデンスとして強度のある情報を得るかという観点から調査をデザインすることも重要である。その際、対象者にとって不利益となるような調査・実験をおこなってはならない教育において、より客観的で信頼できるエビデンスを得るために、ランダム化比較実験や疫学的アプローチ (中山・津谷, 2016) はもっと利用されてしかるべきものであることを指摘したい。

たとえば、教育では一般に個々の子どもを丁寧に見ていくということが最優先になるため、発達段階を追うときもすべての個人を全期間にわたって追跡しなければならないというような発想に陥りがちである。もちろんこのこと自体は否定すべきことでない。しかし、コストや時間的な制約上どうしてもそのような「きめ細かな」調査が実施できないときにも、集団を比較する疫学的なアプローチをとれば、かなりの情報を科学的エビデンスとして獲得することができるのである。

逆にいうと、そのような実験デザイン・調査デザインなしで収集されたデータ、あるいは後付けであってもそのようなデザインが設定できないデータからは、目的とする情報は得られない。いくら大規模であっても、調査目的に応じた明確なデザインなしに集められたデータからは明確な情報は紡ぎ出せない。統計的データ解析で言われる、いわゆる“Garbage in, garbage out.”現象が起こることが多いのである。

大学入試センター共通テストの課題

さて、以上のようないわば「枠組み論」の陰にかくれてあまりよく認識されていないのがどのように項目を作成していくのかの論点である。特に多肢選択形式の項目に対する表面的な批判は頻繁になされる。選択肢の中に正答があるから当てずっぽうでも正答できる上、マークするだけだから深く考えなくてもすむといった批判がその典型である。残念ながらこれらは表面的な見かけにとらわれた批判といわざるをえない。たとえばここに6肢選択方式の5つの項目があったとしよう。これを当てずっぽうで鉛筆を転がして5点中3点を取れる確率は、すぐに計算できるように216分の1、小数点表示にして約0.005でしかない。実際のテストでは項目数は30程度は確保されているため、当てずっぽうだけでそれなりの得点をするのはほとんど不可能である。

さらに、よく吟味して作題された項目セットを準備すれば多肢選択方式であってもかなり高次の学力まで測定できることはこれまでの長年の実証的研究で分かっている (Haladyna et al., 2002)。このほかにも、受験生に考えさせようとするあまり、多肢選択形式において正答を一つにせずいくつかの正答を組み合わせた選択肢を準備し、その中から正しい組み合わせの選択肢を選ばせるような無理な工夫も散見する。しかしこれも難易度を見かけ上高くするだけでそこから獲得できる情報は、一つ一つ分離して問うていくよりも遙かに少なくなってしまう。受験生に無駄な時間を使わせることで難易度を見かけ上高くしてしまう、間違った項目作成の例としては、資料を大量に与えられた受験生がパズルを解くかのような作業を経たのち、ようやくそのテストが本来測定しようとしている能力が使えるといった構成になっているような場合が典型である。

わが国においては医学共用試験などでも採用されている多肢選択形式の客観式テストは、そもそも大規模なテストにおいて判定したい能力、測定したい学力をいかに低コストで効率よくまた正確に得るかという実務的な観点から考案され、100年あまりの年月をかけ、理論的にも技術的にも洗練されてきたものである。そこには大規模な試験にふさわしい項目の作成のノウハウ・基礎基本というものがすでに確立されてある。大規模テストにおける記述式のコストパフォーマンスの低さ (井本・柴山, 2016)、および記述式を客観式項目群の中に混在させる実施上の不整合なども含めて、現在進められている大学入試センター試験の「改革」は、本稿執筆時においては、大規模テストを支える測定技術の

点から見ても、その基礎・基本から大きく外れた方向に進んでいるとしか筆者には思えないのである。

引用文献

- Arai, S., & Mayekawa, S. (2005). The characteristics of large-scale examinations administered by public institutions in Japan: From the viewpoint of standardization. *Japanese Journal for Research on Testing*, 1, 81-92.
- 中央教育審議会 (2014). 新しい時代にふさわしい高大接続の実現に向けた高等学校教育、大学教育、大学入学者選抜の一体的改革について (答申)
- Flynn, J. (2013). Why our IQ levels are higher than our grandparents. TED.
- 南風原朝和 (2016). 新テストのねらいと予想される帰結 指導と評価, 62(9), 21-23.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15, 309-333.
- 池田 央 (1970). テストの科学 井上健治 (編) テストの話 中央公論社
- 池田 央 (2010). 重複テスト分冊法と学力調査 指導と評価, 56(1), 44-47.
- 井本佳宏・柴山 直 (2016). 「大学入学希望者学力評価テスト (仮称)」の制度設計における限界について 東北大学大学院教育学研究科研究年報, 64(2), 181-192.
- International Organization for Standardization (2017). ISO/DIS 29992 assessment of outcomes of learning services: Guidance (Working draft: ISO/TC 232/WG 4 N51), ISO.
- 国立教育政策研究所 (2016). [国研ライブラリー] 資質・能力 理論編 東洋館出版社
- 国土審議会政策部会長期展望委員会 (2011). 「国土の長期展望」中間とりまとめ 概要
- 久保田賢一 (2003). 構成主義が投げかける新しい教育 コンピュータ&エデュケーション, 15, 12-18.
- 前川眞一 (2007). 医師国家試験改善委員会での取り組み 日本テスト学会第8回研究会 (公開シンポジウム) 発表資料 日本テスト学会
- 中山健夫・津谷喜一郎 (編) (2016). 臨床研究と疫学研究のための国際ルール集 Part2 ライフサイエンス社
- OECD (2015). Global Education Industry Summit

Helsinki, 12-20 October 2015. Retrieved from <http://www.oecd.org/education-industry-summit/> (2017 年 12 月 17 日閲覧)

柴山 直 (2008). 日本のテスト文化について 人事試験研究, 208, 2-13.

柴山 直 (2014). 全国規模の学力調査におけるマトリックス・サンプリングにもとづく集団統計量の推定について 全国的な学力調査平成 24 年度追加分析報告書 文部科学省

日々の授業・学習改善に寄与する評価を考える —心理学を生かした実践的試みを中心に

植阪友理

日々の学習指導における評価の重要性

大規模学力調査や入試等における測定評価は非常に重要な問題である。しかし、テストの利用はそれだけにとどまらない。学校における日々の指導のなかでどのような評価が求められるのかということについては、別個の問題として考える必要がある。そこで本稿では、後者に着目する。

なお、筆者は「評価」の専門家ではない。どのような研究者であるのかと問われれば、「学習改善、指導法改善が専門分野である教授・学習心理学者」と答えるだろう。そこで本稿でも、その視点を踏まえて「評価」の内容をすこし広く捉えて話を進めたい。具体的には、中学校現場での実践を例に話を進めたいと考えているが、定期考査を評価としてとらえるのみならず、授業中に教員がどのような生徒の行動を称揚するのか、また、何を大事と言い、どのような学習課題を与えるのか、そうしたことをすべてを評価として捉えたい。このように広く捉える理由は、後述するように、学習者は丸暗記やパターン学習に陥りがちであったり、丸暗記や結果主義を是とするような信念を持ちがちであったりするが、こうした実態は、教師の行う定期考査や教師の授業中の行動（どのような課題を与え、なにができたときに褒めるのかも含めて）によって大きく方向付けられていると考えられるためである。逆に言えば、本稿は、学習者に期待するような学力、学習行動、学習態度をつけてもらうためには、どのような評価が求められるのかを、テストのみならず授業中の教師の行動も含めた視点から考えることを試みる。

評価を通じて変容させていくべき子どもの実態

ではまず、評価を通じて変えていくべき子どもたちの実態について考えてみよう。心理学的な視点を踏まえると、どのような学習者が理想であり、どのような

点に問題があると言えるのだろうか。学習方略研究を参考にする、リハーサルや単純反復を繰り返すだけでなく、自らの知識と関連付けながら学習することが重要であることが分かる。また、自らの理解状態を客観的に把握し、学習行動に反映させることはメタ認知などと呼ばれ、有効とされる。例えば、間違った部分を中心に覚える、間違った理由を分析して、次に同じ間違いをしないためにはどうすれば良いのかも含めて教訓を引き出すなどが挙げられる。また、こうした学習方法を重視するような考え（学習観）を獲得することの重要性も繰り返し議論されている。

その一方で、私たちの研究グループが長年にわたって行っている心理学を生かした個別学習相談（認知カウンセリング）の実践からは、こうした学習行動が十分に身につけていない実態が見えてくる。認知カウンセリングの実践では、学習に悩む子どもたちの支援を行っている。その中で、知識・技能のつまずきの背景として、日々の授業の受け方や家庭での学習の仕方など、学習方略や学習観に問題が潜んでいる事例が多々見られたためである。

例えば、自分の間違いの理由を分析するというメタ認知的方略が活用できていないために、繰り返し問題を解きながらも、学習成果を上げられず、一種の学習性無力感に陥っている学習者も見受けられる。間違いを振り返らない（振り返れない）理由の背景には、失敗は見たくないという失敗を忌避する不適応的な学習観も伺われた。こうした子どもが悩みから解放されるには、学習観や学習方略についても、改善を促す必要がある。認知カウンセリングでは、個別学習相談の中でこうした試みを行ってきた（例：植阪・市川, 2017）。

世界的な教育動向および日本の教育課程が目指すものとの関係

このように、学校において教科の知識を身につけるのみならず、効果的な学習方略や適応的な学習観についても身につけさせようとする発想は近年の教育界において求められている学習者像とも一致している。例えば、OECD のキー・コンピテンシーや 21 世紀型スキルなどといった議論は、その一つである。こうした考え方では、教科知識の獲得のみならず、社会においてもうまく生きていけるスキルの獲得をも育成することが重視されている。これらの議論を受けて日本でも、「資質・能力の育成」という言葉が中央教育審議会（文部科学省, 2016）の答申の中で重視されており、メタ認知という心理学の用語も答申の中に入っている。上述したような学習者を育てようとすることは、日本も含め、

世界が求める学習者像とも一致していると言えるのである。

では、どのようにすればそうした学習者を育てていくことができるだろうか。実は、キー・コンピテンシーや21世紀型スキルの議論において決定的に欠けているのは、具体的にどうやってこうした学習者を育てていくのかという実践的知見である。特に、具体的な教科の指導のなかでどう育てていくのが明らかにされなければ、目標が達成されていくことは非常に難しい。この点について、具体的な教育実践を伴った提案が求められているのである。

では、具体的に学習観、学習方略を育成していくことを考えてみると、教師が学習方法や学習に対する考え方を重視する発想さえ持っていれば良いのだろうか。もしくは、教師が重要なことを学習者に口で伝えるだけで十分だろうか。どちらも効果的とは言い難いだろう。教師がどんなに意味の理解が重要であると考えても、授業中の課題として手続きを当てはめるだけの問題を出していたとすれば、学習者はそうした学習へと方向付けられることになる。また、たとえ教師が口を酸っぱくして意味を理解すること、途中過程を大事にすることの大切さを語ったとしても、テストで出題される問題にそうしたことが全く含まれていなければ、学習者は暗黙裡に「結局、答えさえ合えばよいのだな」と考えるようになりがちである。以下では、こうした問題意識を持ち、学習観、学習方略も含めた指導を評価と連動させながら行った実践事例を取り上げる。紹介の中心は中学校における実践（岡山市立灘崎中学校、床勝信教諭との実践）であるが、小学校においても同様の発想での実践は可能であることを示すために、2つ目の実践（文京区千駄木小学校、滑澤教諭との実践）についても簡単に触れる。

定期考査、授業、数学通信を一体的に連動させた実践

定期考査の変容 本稿で中心的に紹介する事例は、岡山市立灘崎中学校における数学科の床勝信教諭との実践である（詳細は、ベネッセ（2012）の対談記事や床（2012）などを参考にされたい）。実践当時、灘崎中学校は、地域の4校の小学校とともに、岡山県学力・人間力育成推進会議のIFプランと呼ばれる認知心理学の発想を生かした実践に取り組んでいた。筆者はそのIFプランの研究推進委員として、東京大学の市川伸一教授（同会議、顧問）とともに参加していた。実践の背景として床教諭がしばしば口にしたことは、学習観の改善に取り組む必要性と、口で言うだけでは十分ではないという発想である。すわなち、床教諭は、子どもたちが

ドリル学習を好み、考えることを好まないという学習観のあり方に、非常に大きな疑問を持っていた。その一方で、授業の中で思考過程を考えることの重要性や説明の重要性を口で伝えても、なかなか子どもたちの学習観が変わらないというジレンマも感じていた。これを解決する視点として、授業中に口で重要性を伝えても、実際の定期考査や授業において、考え方が問われることがなければ、子どもは暗黙のうちに、「結局は答えさえ合えば良いのでしょう」という発想を持つ、ということに着目した。そこで、まず、定期考査を変化させることを考えた。

具体的には以下のように変容させた。従来の定期考査は、「六角形の内角の和はいくつか。」といった問題に代表されるように、公式などを使いながら、問題を解くということが中心であった。それを、「六角形の内角の和は720度です。そうなる理由を図で説明せよ。」のように、解かせるだけではなく、なぜそのような公式が導出できるのかを直接的に問う問題へと切り替えた。この他にも、従来であれば「当たり2本、はずれ2本が入っているくじから同時に2本ひく時、当たりくじを引く確率を求めなさい」と確率計算させていた問題において、「以下は先日行なった単元テストで多くの生徒が間違えた問題である。この解き方の間違いを指摘しなさい（誤答例を併せて提示）。」などのように、間違えた理由を解答させるなどの問題を出題している。この他にも、相対度数を計算させた上で、「2つの中学校を比べるのに、度数ではなく相対度数を使っているのはなぜか。その理由を説明しなさい。」というような問題も出題している。こうした、「なぜ」にあたる部分を説明させるような問題を、全体の2分の1から3分の1程度出題する形に切り替えていった。また、採点にあたっては、問題ごとに担当者を決め、記述問題については一貫した評価が行えるように配慮している。

こうした問題に切り替えた背景には、深い理解を促したいという思いはもちろんのこと、学習者の学習観や学習方略を変容させたいという思いがある。そのため、繰り返し述べているように、途中過程が重要であることを口だけで伝えるのではなく、テストでも問い、評価の対象となっていることを明示しているのである。

日々の授業実践の変容 ただし、こうした問題を定期考査に出題するに止ただけで、通常の授業に変化がなければ、こうした実践は十分に機能しなかったであろう。もし授業が従来通りであれば、学習者にとってこうした定期考査は授業でやったこと以外を問

われる、裏切り以外の何物でもなくなってしまう。そこで、日々の授業実践にも変化を加えた。

具体的には、「教えて考えさせる授業」(市川, 2004, 2008; 市川・植阪, 2016) の発想を活用し、教師が意味に着目した説明を行うのみならず、子どもたちにも同様の説明を求めることとした。例えば、多角形の内角の和を教える授業であれば、従来であれば、五角形などを使いながら公式の成り立ちなどを教えた後、六角形や七角形などの内角の和を計算させ、さらに 1080 度になるのは何角形の時かなどといった問題を解かせることになる。すなわち、子ども自身が公式にあてはめて計算をすることが中心で、なぜその公式になるのかを説明する機会はほぼないといつて良い。一方で、床教諭の授業では、まず、五角形などを使いながら公式の成り立ちなどを教えたあと、ペア学習の機会を設けて、全員がそうした公式になる理由を説明することを求める。さらに、他の方法でも公式が導出できる方法をグループで考えるという課題に取り組ませた(具体的には、1つの頂点から線分をひいて三角形に分ける方法ではなく、多角形の中に点を取り、その点と頂点を結んで三角形を作って公式を導出する方法などを子どもたちに考えさせていた)。最後には、振り返りとして大切だと思ったこと、まだ疑問に残っていること、感想などを書かせている。

数学通信による支援 こうした授業のやり方は、子どもたちにとって非常に新しいものであった。また、そうした説明を求められることもこれまでほとんどなかったため、子どもたちの中には、どうしてこんな授業をするのか疑問に思う子どもたちもいた。保護者も同様であった(定期考査後に苦情の電話をかけてきた保護者もいたという)。また、どのような説明を求められているのかが子どもにとって、特に初期の頃には分かりにくいということがおこることも容易に想像できた。そこで、子どもの家庭学習を支援するために数学通信を年間 20—30 枚発行して、授業中にもそれを活用して解説を行った。この数学通信には、効果的な学習方法とその理由が示されている。例えば、なぜ授業中に説明を求めるのかといえ、説明できることが理解と言えること、また、その際にはやり方だけではなくなぜの部分に着目した説明ができることが重要であることなどが記載されていた。さらに、授業中にうまく説明をしていた子どものノート例をコピーして示し、図をうまく使っているなど、なぜ良い説明と言えるのかについても解説を加えていた。また、テスト後には、どのように間違いと向き合うと良いのかといった分析方法についても数学通信で共有していた。当初は、問題を解

くことでテスト対策をしていた子どもたちも、次第に、数学通信をうまく活用しながら勉強するようになっていっている様子が後述するアンケートからも伺える。

子どもたちの反応—2 年間の実践を経て このように、定期考査を変容させ、それに応じて授業も変化させ、数学通信を通じて家庭学習も支援するというような一体的な指導法改善を 2 年間にわたり行った。では、子どもたちはこうした指導に対して、どのように反応したのだろうか。本当に学習観や学習方略は変容したのだろうか。勉強の仕方の変化を問うたアンケートにおいて、実践に参加したあるクラスでは、61.2%もの生徒が、意味理解を重視した学習法、間違いを積極的に分析するような学習法、数学通信を重視するような学習法など、認知心理学的に見て適応的と見える行動に言及していた。例えば、ある生徒は「説明の問題を中心に勉強するようになりました。わからないところを友人に教えてもらい、『どうしてこうなるのか』を考えて勉強しています。1 年の時には『とにかく計算をがむしゃらに解いていこう!』という感じだったので、今のほうが勉強しやすいです。」と記述している。学習観および学習方略の変容が伺える。

ただし、学習観および学習方略の変容に至るには時間もかかることが示唆された。1 年目に行った同様のアンケートでは、変化が見られない子も少なくなかった。5 回にもわたる定期考査を経験したにもかかわらず、認知主義的な学習方法に言及した学習者は 30.3%であった。ある子どもは「課題テストや小学校のテストは『……の答えを求めなさい』というのが多かったけど、定期考査は『なぜ……ですか説明しなさい』ばかりで難しかった。テスト勉強は問題を解くことばかり考えていたから、なんでこういう解き方になるのか考えてやらないとダメだ!」と述べており、この言葉に代表されるように、1 年間をかけてやっと自分の学習方法の問題点に気づいたという状態であることがうかがわれた。子どもたちの丸暗記志向や思考過程軽視志向が根強いことが伺え、継続的に実践を行うことの重要性も示唆された。

小学校での実践可能性について なお、本実践について紹介すると、「中学校では教師がテストを作成しているのでこうした実践が可能であるが、小学校では難しい」という意見を聞くことがある。多くの小学校の場合には、業者テストを使っているためである。しかし、考えてみると、小学校において途中経過がどんなに大事であると先生が強調していても、業者テストで問われるのは式と答えであり、床教諭が指摘した問題

が発生している可能性が十分にある。また中学校段階の根強い丸暗記志向、思考過程軽視志向を見ても、小学校段階でこうした発想が獲得されている可能性が十分にある。では小学校では難しいのだろうか。詳細は割愛するが、文京区千駄木小学校の滑澤雅子教諭は、小学校において業者テストを実施する際に、白紙を1枚与え、そこに途中過程を書くことで部分点を与えるという取り組みも行っている（詳細は滑澤・齊藤, 2014 参照）。またその前提として、授業の中で自分はどの部分を理解していなかったために、問題が解けなかったのかということを十分に共有する実践をしている。床教諭と全く同じ実践ではないものの、評価も意識した実践は、小学校段階からも可能なのである。

考察

床教諭との実践は、評価を定期考査に限定せず、どのような課題を与えるのか、どのような行動を称揚するのかも含めて捉える必要があることを考えさせてくれる。こうした一体的な働きかけの中で、子どもたちは「理解したとは、自分の言葉で『なぜそうなるのか』が説明できることなのだ」という認識を持つようになると期待できる。そしてそうしたことの積み重ねが、昨今の教育界でも求められているような、学習行動や学習に対する考え方の育成につながっていくと考えられる。

これまでも、教師自身は意味理解や思考過程を重視してきたと思われるが、その信念が授業や定期考査にまでは反映されていなかったのが現状であろう。あらためて、自らの信念が子どもの信念として内化されるために必要な広い意味での評価のあり方を問い直す必要がある。

最後に、床教諭の実践についてこれまでの教育心理学の知見との関連を考えてみたい。評価のあり方によって、学習者の学習行動が変化することは、教育心理学においても実験的に確かめられてきている。例えば、村山（2003）は、歴史学習を素材にして穴埋め中心のテストにさらされていると、テストに備えて、表面的な理解を志向するような学習行動を取る一方で、記述的なテストにさらされていると、より深い理解を志向するような学習行動を取ることを実験的に確かめている。この研究は、学習ゼミナールと呼ばれる大学での実験授業を通じて明らかにされており、学校における教科指導とは比較的近い。しかし、やはりそのまま学校における日々の学習指導に活かせるようなスタイルでは提案されていない。また、床教諭の実践が行ったような、授業自体の工夫については考慮されてい

ない。床実践は、村山（2003）が明らかにしたような発想を十分に踏まえながらも、学習観の改善までを視野に入れ、実際の学校教育で利用可能な形にまで高めたものと考えられる。逆に言えば、学習方法を、日々の指導の評価によってどのように変容させていくのかという具体的な教育実践は、これまで十分に明らかにされてこなかったと言える。

また、床教諭の実践は、学習観に変容をもたらしたという点でも特筆すべきだろう。学習観の重要性は、学習方略を取り上げた論文の中で繰り返し議論されている。しかし、その多くは影響関係が見られることを実証的に明らかにしたにとどまり、具体的にどのようにすれば変容させることができるのかという点については、十分に明らかにされてきていない。これまでは、認知カウンセリングの中で学習観の変容を報告している植阪（2010）以外には、具体的に学校での日々の学習における学習観が変容していると分かるような実践は、ほとんど報告されてきていないのである。一方で、今回の床教諭の実践は、定期考査と授業中の教師の行動をも含めた広い意味からの評価を変化させることで、学校での日々の学習における学習観を変容することができる新たな可能性を示したものであった。学習に取り組む態度や、学習方法も含めて学力として捉え、学校教育において保証することが学習指導要領において決められた現代において、重要な実践であったと言えるだろう。

新指導要領では、学習内容の深い理解を達成し、かつ、社会に出た後にも利用できるような資質・能力を育成していくことが求められている。学力保証という観点から考えてみると、どの学校においても、こうした広義の意味での学力を身につけさせる必要が生じている。床教諭の実践や滑澤教諭の実践はあくまで一つの例であるが、現場での工夫を考えるにあたっての参考になれば幸いである。

引用文献

- ベネッセ（2012）. 対談—自律した学習者を育てるために生徒の学習観を変える VIEW21 中学版 2012 年度, Vol. 3, pp. 6-13.
- 中央教育審議会（2016）. 幼稚園、小学校、中学校、高等学校及び特別支援学校の学習指導要領等の改善及び必要な方策等について（答申）
- 市川伸一（2004）. 学ぶ意欲とスキルを育てる 小学館
- 市川伸一（2008）. 「教えて考えさせる授業」を創る—基礎基本の定着・深化・活用を促す「習得型」授業

- 設計 図書文化
市川伸一・植阪友理（編）（2016）. 最新 教えて考えさせる授業 小学校—深い学びとメタ認知を促す授業プラン 図書文化
文部科学省（2016）. 小中一貫した教育課程の編成実施に関する手引き
村山 航（2003）. テスト形式が学習方略に与える影響 教育心理学研究, 51, 1-12.
滑澤雅子・齊藤 純（2014）. つまづきを自覚し、つまづきを克服する教室作り 植阪友理・Emmanuel Manalo（編） 学習方略研究における理論と実践の新たな展開—学習方略プロジェクト H25 年度の研究成果, Working Papers, Vol. 3, pp. 98-101.
床 勝信（2012）. 指導と評価を連動させる定期テストの改善 市川伸一（編） 教えて考えさせる授業 中学校（pp. 114-128） 図書文化社
植阪友理（2010）. 学習方略は教科間でいかに転移するか—「教訓帰納」の自発的な利用を促す事例研究から 教育心理学研究, 58, 80-94.
植阪友理・市川伸一（編）（2017）. 2016 年度学習支援事例—自立的な学習者を育てる認知カウンセリングの実践〈抜粋版〉 東京大学

「非認知」なるものの発達と教育 —殊に学力形成との関わりにおいて

遠藤利彦

心理学における「非認知」なるものへの刮目

近年とみに、教育、殊に幼児教育の分野において、「非認知」なるものの心の性質の発達と教育に対して、多大な関心が寄せられるようになってきている。幼少期に培われる「非認知」的な心が、その後の学力の形成や職業上の成功なども含め、広く様々な心理社会的適応性やウェルビーイングの実現において、重要な鍵を握っているというのである（e.g., Gutman & Schoon, 2013; Stankov & Lee, 2014）。そして、その火付け役となったものが、James Heckman に代表される教育経済学的研究（e.g., Heckman, 2013; Heckman & Rubinstein, 2001）であることは衆目の一致するところであろう。しかし、心理学の領域に目を向ければ、その黎明期から既に「認知」的能力のみならず、それ以外の心の性質への刮目も確かな形で存在していたと考えられる。

例えば Dewey（1909）や Thorndike（1920）などの初期の研究者の理論の中に、既に、今で言うところの「非認知」的な心の性質への着目が明瞭な形で存在していることが読み取れる。Dewey は、人の道徳的な動機や

ふるまいが、社会的状況を正確に理解し、また的確に制御し得る力として在る社会的な知能から発するとし、それを学校教育のカリキュラムの中でいかに育み得るかということに関心を有していたことが知られている。Thorndike も、他者を理解し、自身を管理し、そしてあらゆる人間関係において賢明に行動するための能力を社会的な知能とし、それが抽象的な知能とは明確に区別されるべきものであることを主張していた。また、「認知」の重要性を専ら強調してきたかに見える IQ 論者も、Alfred Binet を始め、人の心理社会的適応性などが純粋に「認知」能力だけでは規定され得ないことを認識し、IQ には含まれない、今で言う「非認知」の要素の枢要な役割を仮定していたと言い得る（Landy, 2005）。現代の代表的な知能論者で Gardner（2006）や Sternberg（2007）もまた然りである。そして、現代の代表的な知能論者である Gardner（2006）や Sternberg（2007）もまた、知能の主要な構成素として、いくつかの「非認知」的な心の性質を仮定している。

しかし、このように、心理学分野においては、「非認知」なるものへの注視は早くから現在に至るまで一貫して存在していたとは言えるものの、教育現場への実践的应用という文脈で言えば、それが、「認知」的能力ほどに、適切かつ十分な扱いを受けてこなかったことは否めないところである。その理由としては様々なことが想定されるが、その主たる要因としては、多くの「認知」的能力が IQ を筆頭に、標準化された測度によってある程度、客観的に測定されるものとして在るのに対して、「非認知」的な心の性質に関しては、一部、パーソナリティなどを除き、その測定評価が必ずしも確かな形ではなされないままになってきたということが挙げられよう。おそらく、教育現場においては、今も昔も変わらずに、直観レベルでは、多くの教育実践者が「非認知」なるものの重要性を看破していたのかもしれない。しかし、測定評価上の障壁によって、それが、体系化された具体的な教育法の開発あるいはそれに基づいた教育実践という形で実を結ぶことは相対的に少なかったと言うべきなのだろう。

IQ のアンチテーゼとしての感情的知性

上述したように、「非認知」的な心の性質に関しては、早くから理論的な関心が払われながら、その教育実践への応用が、長く不十分なものに留まってきたと言わざるを得ない。その一方で、「認知」的能力、殊にその総体として在る IQ が、時に、例えば人種差などに絡み、多分にイデオロギー的とも言える論争を巻き起こしつつも、人の適応性や社会的成功を最も強く予測す

るものとして、良かれ悪しかれ、教育的関心のかなり中核部分に据えられてきたことは否定できないところであろう。

1つ例を挙げれば、今から20年ほど前に、話題を渡った1冊に*Bell curve* (Herrnstein & Murray, 1994) と題した書がある。それは、実証的データに基づきながら、知的能力、すなわちIQの高低が経済的貧富を分ける最も強力な要因であり、社会的階層が、IQに基づく同類交配 (IQ水準のほぼ等しい者同士の結婚や家族形成) によって成り立っている可能性が高いこと、そしてまたIQの遺伝的規定性が強いことを論じ、いくら福祉を手厚くしても (その書では遺伝的にIQが平均的に低いと仮定されている) 貧困層の子どもへの補償教育にあまり効果が期待できないことなどから、貧困層への福祉に重きを置く社会的政策のあり方を見直すべきだと主張したのである。

こうした主張は、大きな物議を醸し、社会からの激しい反発を買うことになった訳であるが、その1年後に、今度はそれとはほぼ真逆の方向で人心を驚づかみにする書、すなわちGoleman (1995) による*Emotional intelligence* が出版されることになるのである。それは、人の社会的な成功や幸福が純然たる知的能力=IQではなく、感情的知性=EI (Emotional Intelligence) によってもたらされること、そしてEIは遺伝よりも多分に教育やしつけあるいは個人の意思によって後天的に獲得可能なものであることを強く印象づけることで、すべての人があらゆる可能性に拓かれて在るという感覚を社会の中にもたらしたのである。

EIの定義は、研究者によってかなり相違があるが、例えば、現今のEI研究の実質的創始者として在るSalovey & Mayer (1990) のいわゆる「4枝モデル」(four branch model) などによれば、EIとは、(a) 感情の知覚・同定、(b) 感情の促進および思考への同化、(c) 感情の理解や推論、(d) 自他の感情の制御と管理という4種の下位要素 (4本の枝) から構成されるものだという。これらは、成人が有すべき心理的特質を、殊に感情の視点から定義づけているものではあるが、昨今の議論の中で、一般的に子どもが備えておいて然るべきとされる「非認知」的な心理的性質と、元来、概念的にかなり大きな重なりを有するものと言える。

Goleman (1995, 1998) に至っては、さらに、EIの中に、自己有能感、動機づけ、共感性、楽観性、優れた道徳的性質、種々の社会的スキルなども広く含めて考えるようになったため、それは、現在「非認知」的な心の性質と言われるものと実質的にほとんど違いがな

いと言っても過言でない。彼によれば、こうしたEIの不足によって、例えば自他の感情を読み違い、適切に自身の感情の制御をし損なうことで、多くの場合、教育上の種々の達成も仕事におけるパフォーマンスや満足感なども低水準に止まることになり、また、その影響下において、少なからず様々な犯罪や精神の病なども生み出されることになるのだという。だが、EIは基本的にすべての人に習得・訓練可能なものとして在るため、たとえば社会的弱者であっても、原理的にそこからの脱出は実現可能なのだと主張したのである。「ペリー就学前計画」が含意するもの

そして、こうしたEIに対する社会的な高まりとほぼ時期を同じくして、*Bell curve* への批判の論陣を張り、別角度から「非認知」的な心の性質の重要性に対して、人の眼を向けさせたのが、先にもふれたHeckmanだったのである。彼は多様なデータに基づきながら、就学前 (乳幼児期) における教育への投資効果が最も大きいことを結論している訳であるが、加えて、その教育が、個人の中に、「認知」的能力の基盤を準備することを通してよりも、むしろそれ以外の力、すなわち「非認知」的な心の基礎をもたらしことを通して、人の生涯に亘るウェルビーイングや社会的成功を可能ならしめると主張したのである (Heckman, 2013; Heckman, Stixrud, & Urzua, 2006)。特にそうした主張の論拠になっているのが、彼自身も関わったいくつかの介入研究、その中でもとりわけ「ペリー就学前計画」ということになる。それは、1962—1967年に、米国ミシガン州ペリー地区において、アフリカ系米国人の貧困層の子どもおよびその養育者を対象に行われた教育的介入の効果を、その後、追跡研究の形で継続的に検証しようとしたものである。

介入の対象となった貧困層の子どもたちは、介入開始時である3, 4歳時点でIQが70—85であり、そこから、経済的な困窮の中で、家庭におけるその子どもたちへの養育や教育はかなり低水準であったことが伺えよう。その子どもたちに対して、その後2年間に亘って、基本的に平日は毎日、午前中に幼稚園に通わせ、初歩的な幼児教育のプログラムや遊びなどを中心とした活動を行わせたのである。この子どもたちは、その後、複数時点で、追跡調査を受け、様々な観点から、同じくアフリカ系アメリカ人の貧困層の子どもからなる対照群と比較分析されることになる。その結果は、被介入群の方が比較対照群に比して、40歳時点において、高校卒業率、収入、持ち家率などにおいてより高く、離婚率、犯罪率、生活保護受給率などにおいて低

いということ、すなわち相対的に経済的に安定しており、健全な市民としての適応的な生活を享受できているケースが多いということを示すものであった。

ただ、ここで着目して然るべきは、「認知」能力の指標たる IQ の群間差ということになろう。確かに介入を受けた子どもたちは幼稚園に通園中の段階は、IQ に瞠目すべき伸びを見せたのだが、介入が終了した後から 2 群の差は徐々に狭まり、10 歳前後には、介入開始時の 3 歳時点と同様、ほとんど違いがなくなってしまったのである。そして、それにも拘わらず、それから 30 年以上の時を隔てた 40 歳の段階で、被介入群が、対照群よりも生活の質がすぐれることから、Heckman は、それを可能ならしめたのが、IQ、すなわち認知的能力ではなく、それ以外のもの、つまりは「非認知」的な心の性質であると言明することになるのである。

もっとも、ここで注意しておくべきことは、ペリー就学前計画が、実のところ「非認知」的な心の性質の重要性をきわめて消極的な形でしか示してはいないということである。Heckman 自身は、被介入群が、「非認知」なるものの中でも殊に、自制心やグリット（粘り強くやり抜く力）のようなものを幼少期に身につけ得た可能性があることにふれてはいるが、この介入研究では、これらの要素も含め、「非認知」的な心の性質に関して、何ら直接的なアセスメントは行われていないのである。また、確かに、被介入群の方で、心理社会的適応や経済的安定の程度は確かに幾分高いのではあるが、介入の何が効果を上げ得たかに関しては、やはり、特にその要因の特定や機序の解明が試みられている訳ではない。効果を生み出した要因の候補としては、子どもたちが受けた特定の幼児教育のプログラムということのみならず、保育者とのアタッチメントを含めた全般的関係性、家庭訪問によって養育者が得たサポートや近隣地域とのつながりなど、実に多様なものが想定される訳である（e.g., Gopnik, 2010; West, 2013）が、それらが現実的に実証的検討に付されてはいないのである。

それともう一点、冷静に考えてみなければならないことは、そこで効果として謳われているものをいかほどのものと評価すべきかということであろう。確かに、対照群と比較すれば、被介入群の方が多くの指標において優っている訳ではあるが、それでも、大概是、同年代の米国人の平均よりも、かなり低水準に止まるのである（例えば、月収や持ち家率は低く、生活保護受給率や拘置所・刑務所収監率は高い；e.g., McCluskey, 2013）。無論、純粋に教育的視座からすれば、某かの介入によって、たとえ微々たる改善であっても、あるいは少数のケース

に限定された形であっても、正の効果が得られたということ自体を前向きに評価すべきことは言うまでもない。ただし、もう少しマクロに、特に政策的視座などからすると、ペリー就学前計画のように、小サンプルに対して高投資で集中的な介入をした場合には、単に効果があったかなかったかという点のみならず、その効果がその介入のコストにどれだけ見合うものであったかについても、当然、正当な評価をしておく必要があるのだろう。

あらためて「非認知」なるものとは何か？

上述したように、殊に 20 世紀の終わり頃から、欧米圏を中心に IQ 信仰に対する疑念が強まる中で、1 つには IQ のアンチテーゼとしての EI 概念に対して理論的・実践的な期待が高まるという流れが生まれ、もう 1 つには Heckman に先導されて教育経済学的知見に基づいて就学前における特に IQ 以外の土台形成の重要性をあらためて見直そうという流れが生まれ、そして、結果として、現今における「非認知」なるものへの関心の高まりがあるのだと見なして然るべきだろう。

さて、ここまでは、あえて「非認知」なるもの、あるいは「非認知」的な心の性質といった婉曲的な表現を用いて議論を行ってきたが、具体的にそれはどのようなものを指し示すものと把握すべきなのだろうか。Heckman も含め、これまでに様々な見方が提示されてきているが、例えば、OECD（経済協力開発機構）は、*Skills for social progress: The power of social and emotional skills* と題したレポートの中で、「非認知」的な心の性質を社会情緒的スキル（social and emotional skills）と言い換えた上で、生産性（productivity：個人の心理社会的適応や社会の経済発展等に寄与し得るか）、測定可能性（measurability：客観的尺度によって測定可能か）、変容可能性（malleability：投資、環境改善、教育などによって変容可能か）という視座から、問うに値する心的要素の絞り込みを行っている。そして、そうした要素が「長期的目標の達成」、「他者との協働」、「感情の管理」という 3 つの側面のいずれかに整理されるとしている（OECD, 2015）。

また、遠藤（2017）は国立教育政策研究所のプロジェクト研究において、現今の研究や実践の文脈において一般的に「非認知」とされているものを、社会情緒的コンピテンスとした上で、『自分と他者・集団との関係に関する社会的適応』および『心身の健康・成長』につながる行動や態度、そしてまた、それらを可能ならしめる心理的特質」と定義づけている。これを、もう少し簡潔に言えば、「社会性」と「自己」に関わる心の性質と換言し得るかも知れない。

本源的に高度に社会的生物として在るヒトにおいては、個人の適応のかかなりのところが、個人に閉じた形ではなく、他者や集団との関係性如何によって規定されるものと言える (e.g., Tooby & Cosmides, 2008)。すなわち、個人がいかに他者との関係性を構築・維持し、また集団の成員として安定した位置を確保し得るかということが、多くの場合、ほぼそのままに個人レベルのウェルビーイングにもつながるものということになる。その意味からすれば、他者との関係性の構築・維持に必然的に深く関わることになる、例えば心的理解、共感性、道徳性、規範意識、向社会的行動といった、一般的に総じて「社会性」と言われてきたものが、社会情緒的コンピテンスの1本の重要な柱ということになるだろう。

一方、人は元来、自己実現、自己高揚、自己保全等の欲求に強く駆られる存在としても在り (e.g., Dweck, 2000, 2006)、その意味からすれば、自らの心身の健康や成長において、そうした「自己」に関わる欲求の充足を効果的に可能ならしめる心の性質、例えば自尊心、自制心、グリット、内発的動機づけ、自律性といった特質もまた、社会情緒的コンピテンスのもう1本の重要な柱として仮定されて然るべきであろう。

ちなみに、実のところ、こうした一連の心的要素は、基本的に、いずれも、一般的に社会的認知 (social cognition) と言われるものを基盤に成り立っているものと言え、その意味からすれば、本来、心理学的に言えば、厳密な意味で、これらを「非認知」という言葉で括することは、不適切ということになるだろう。ノーベル経済学賞受賞者たる Heckman が用いたということで、「非認知」という言葉が、現在、殊に経済学や幼児教育・保育の領域を中心に、社会の中に広く浸透しつつある訳だが、この言葉の使用に関しては、今後、慎重なスタンスが求められよう。

「非認知」なるものと学力との関連性

さて、上述したような「非認知」的な心の性質は、学力といかなる関連性を有していると言えるのだろうか。その詳細に関しては、乳幼児期から児童期・青年期に至るまでの「非認知」的な心の性質とその規定因や発達の帰結などに関わる実証研究を包括的にレビューした遠藤 (2017) に委ねるが、それを概括すると、特に「自己」領域に関わる諸変数の学力に対する関与が多く認められると言えるかもしれない。「社会性」に関わる諸変数の関与を認めている研究も一部にはあるが、それは主にクラス適応などを介しての効果と言えるようである。

これまで教育心理学分野で精力的に問われてきた「自己」に潜在的に深く関わる諸変数、例えば自己効力感、期待価値、原因帰属スタイルなどが動機づけの高低を左右することを介して、学業達成に影響を及ぼすことは、ある意味、言わずもがなと言うべきであるが、幼少期の心理学的特質に着目して言えば、とりわけ自己制御や自己抑制に関わる変数の予測力が相対的に大きいようである。より具体的には、実行機能 (注意の統制・切替など)、衝動および感情の制御、満足感遅延、エフォートフル・コントロールなどが、同時期およびその後の学力水準と関連するという結果を得ている研究が比較的多いと考えられる。また、これらの心的要素に関しては、それらを高めるための介入も、学力形成に対して一定の効果をもたらし得ることが、明らかになっているようである。

もっとも、日本も含め、世界各国において、教育分野での関心がとりわけ高いと言える自尊心あるいは自己肯定感に関しては、その学力との関連はかなり微妙なものと言わなければならない。確かに同時相関的な研究からは、それらの間の有意な相関関係が数多く報告されているが、その因果関係は、学力の高さが自尊心や自己肯定感を高め得るのであって、その逆では必ずしもないようなのである。例えば、米国では、1980年代から自尊心を社会的ワクチンと称し、それを高める介入によって、学力の向上や反社会的行動の減少を企図する試みが精力的になされたことがあった訳であるが、今では、その効果は実質的にはほとんどなかったこと、そしてまた、根拠に裏打ちされない形でいたずらに自尊心が高められることが時に逆効果をもたらす場合もあることが結論されている (Baumeister et al., 2005)。ただし、この場合の自尊心は、全般的自尊心であって、個別の領域に関わる自尊心ではないということに注意を払う必要があろう。むしろ、自身の学力等に関して正当な自己評価を伴う、それこそ根拠ある学業的自己概念などは、学業上のパフォーマンスと双方向的因果関係を有していると言えるようであり、それに対する介入には一定の効果が見込まれるものと考えられる (e.g., Marsh & O'mara, 2008)。

なお、教育経済学的研究においては、「非認知」的な心の構成要素の1つとしてパーソナリティを取り上げ、それと学歴、収入、社会的地位等との関連性を問い、現に特定パーソナリティ次元とそれらの有意な連関を明らかにしているものも少なからずあるようである (e.g., 李, 2014)。学力に焦点化した研究ではないが、例えば、日本および米国のサンプルを用いた Lee & Ohtake

(2014)の研究では、所得に関しては男性では勤勉性と、女性では外向性や情緒安定性と正の相関が認められたこと、また、昇進に関しては男性のみで外向性との間に正の相関が認められたことが報告されている。そして、教育経済学の領域では、こうした知見に基づいて、勤勉性や外向性などの促進に向けた教育的働きかけを、例えば学力、延いては将来的な経済的安定性の具現に向けて行うということの有効性が説かれることもあるようである。

確かに、原理的に言えば、こうした試みは不可能ではないのだろう。しかし、パーソナリティの基盤には、元来、個々人に固有の遺伝的要因が深く関わっており、しかも人生の早期段階における変化可能性はそれなりにあるものの、年齢とともにそれは大きく減じていくということも知られており (e.g., Ashton, 2013)、パーソナリティを教育的介入の対象とすることは、現実的な意味で、また倫理的な意味でも、あまり適切ではないと考えられる。パーソナリティやその先駆体として在る気質などに関しては、それ自体の変容を企図するというよりは、むしろそれらを個々人に固有の安定した属性と見なし、それに応じて、いかなる「自己」や「社会性」に絡む社会情緒的コンピテンスを身につけさせることが、より高い心理社会的適応に結びつきやすいかを問う方が、はるかに現実的かつ有効なのだろう。

学力の背景に潜んで在る「非認知」に対する見極め

先にもふれたOECDレポート (OECD, 2015) は、世界9カ国の青少年を対象とした調査結果に基づき、「認知」スキルと「非認知」(社会情緒的) スキルが、一部、重なりを見せつつも、それぞれ異なる発達の帰結を招来することを見出した上で、経済的収入を始め成人期に至ってからの心理社会的適応性やウェルビーイングが少なくとも「認知」スキルだけでは到底、説明され得ないことを強調している。確かに、種々の「認知」スキルは、学力を含めた教育的達成により強い予測力を有しているものの、それが「非認知」スキルを十分に伴わない場合には、必ずしも個人のその後のより高い適応性にはつながらないというのである。また、時間的に先行する「非認知」スキルが、その後の「認知」スキルを予測するという因果関係は認められても、その逆方向の関連性は認められないとしている。そして、「スキルがスキルを生む (skills beget skills)」のだと結論し、殊に幼少期の段階に「非認知」スキルの基礎を身につけておくことが、その後の「認知」「非認知」両側面に亘る、より高次のスキル獲得を可能ならしめ、延いては人の生涯全体に及ぶ経済的安定性や心身の健康

などを実現させ得ることを示唆している。

おそらく、学力の増進だけに限って言えば、殊に短期的に大きな成果を目論む場合には、「非認知」、すなわち社会情緒的なスキルやコンピテンスに焦点化した教育的介入は、必ずしも効率的だとは言えないのかもしれない。例えば、教授法や教材などを工夫し、また外発的動機づけに訴えるような教育的営為の方が、短期的には、少なくとも数値から見える学力の引き上げにはより功を奏することが想定される。しかし、短期間に目に見えた効果を上げないからといって、それらに対する働きかけは決して軽視されてはならないのだろう。OECD レポートが示しているように、より長期的目線で見た場合には、社会情緒的なスキルやコンピテンスに裏打ちされた学力でなければ、個人の生涯に亘る心理社会的適応性の具現には必ずしも直結しない可能性があるからである。

その意味からすると、学力の測定と評価は、それだけに閉じてなされてはならないと言えるのかも知れない。それと同時に、種々の社会情緒的スキルやコンピテンスが個々人において、どのような水準のものとして在るのかの正当な測定と評価がなされて然るべきであろう。個々人の学力が、いったい、どのような「認知」「非認知」の心的要素に支えられて在るのか、あるいはないのかといったことの、精度の高い見極めこそが、本来、強く求められるものと考えられる。

実のところ、社会情緒的なスキルやコンピテンスの測定と評価は、学力や「認知」スキルの測定と評価以上に、テクニカルな課題を多く抱えていると言える。先に見たEIに関しても、現在、急速な勢いで、様々な測度の開発が進められているが、未だ決定打は出ていない状況である (遠藤, 2013)。そうした意味において、その解釈に関しては無論、慎重なスタンスが必要となるが、種々の社会情緒的なスキルやコンピテンスの測定を伴わせることで、個々人の学力が、今後に活きるものか否かについての、ある程度、信頼できる判断が可能となるものと言えよう。当然のことながら、学力が、あるテストにおいて、たとえ同じ値で示されるにしても、その背景に潜んで在るメカニズムが違えば、その値の将来に対する意味合いは個々人によって相当に異質なものと言わざるを得ないのである。個々人の学力の背景に潜む、殊に「非認知」的な心的要素の測定と評価、そしてまた、そうした見極めに応じた、可能な限り個別化された教育 (personalized education) という方向性の追究が、これからの教育心理学における1つの重要な課題なのかも知れない。

引用文献

- Ashton, M. C. (2013). *Individual differences and personality*. New York: Academic Press.
- Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2005). Exploding the self-esteem myth. *Scientific American*, 292, 84-92.
- Dewey, J. (1909). *Moral principles in education*. Boston: Houghton Mifflin.
- Dweck, C. (2000). *Self-theories: Their role in motivation, personality, and development*. New York: Psychology Press.
- Dweck, C. (2006). *Mindset: Changing the way you think to fulfil your potential*. New York: Random House.
- 遠藤利彦 (2013). 「情の理」論—情動の合理性をめぐる心理学的考究 東京大学出版会
- 遠藤利彦 [統括] (2017). 非認知的 (社会情緒的) 能力の発達と科学的研究手法についての研究についての報告書 国立教育政策研究所・平成 27 年度研究プロジェクト報告書
- Gardner, H. (2006). *Multiple intelligences: New horizons*. New York: Basic Books.
- Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. New York: Bantam Books.
- Goleman, D. (1998). *Working with emotional intelligence*. New York: Bantam Books.
- Gopnik, A. (2010). *The philosophical baby: What children's minds tell us about truth, love, and the meaning of life*. New York: Farrar, Straus and Giroux.
- Gutman, L. M., & Schoon, I. (2013). *The impact of non-cognitive skills on outcomes for young people*. London: Institute of Education, University of London.
- Heckman, J. J. (2013). *Giving kids a fair chance*. Cambridge, MA: MIT Press.
- Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *The American Economic Review*, 91(2), 145-149.
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24, 411-482.
- Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.
- Landy, F. J. (2005). Some historical and scientific issues related to research on emotional intelligence. *Journal of Organizational Behavior*, 26, 411-424.
- 李 嬋娟 (2014). 非認知能力が労働市場の成果に与える影響について 日本労働研究雑誌, 650, 30-43.
- Lee, S. Y., & Ohtake, F. (2014). The effects of personality traits and behavioral characteristics on schooling, earnings, and career promotion, *RIETI DP*, 14-E-023.
- Marsh, H. W., & O'mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives on self-concept. *Personality and Social Psychology Bulletin*, 34, 542-552.
- McCluskey, N. (2013). Forum commentary. In J. J. Heckman, *Giving kids a fair chance* (pp. 83-90). Cambridge, MA: MIT Press.
- OECD (2015). *Skills for social progress: The power of social and emotional skills*. OECD Publishing.
- Salovey, P., & Mayer, J. D. (1990). Emotional Intelligence. *Imagination, Cognition, and Personality*, 9, 185-211.
- Stankov, L., & Lee, J. (2014). Quest for the best non-cognitive predictor of academic achievement. *Educational Psychology*, 34, 1-8.
- Sternberg, R. J. (2007). *Wisdom, intelligence, and creativity synthesized*. New York: Cambridge University Press.
- Thorndike, E. L. (1920). Intelligence and its use. *Harper's Magazine*, 140, 227-235.
- Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed, pp. 114-137). New York: Guilford Press.
- West, R. (2013). Forum commentary. In J. J. Heckman, *Giving kids a fair chance* (pp. 55-62). Cambridge, MA: MIT Press.

学力の評価と測定とその周辺の話題

野口裕之

ここでは、この誌上討論「学力の評価と測定をめぐって」の総括に加えて、現在世の中から注目されている「高大接続改革」の中で「大学入学共通テスト」に外部英語試験を導入する際に「異なる外部試験間で測定結果を比較可能にする」という「幻想」で急に有名になった CEFR (Common European Framework of Reference for Languages: Learning, teaching, assessment) について評価と測定の観点から少し取り上げる。CEFR とは後に述べるように、単に英語学力水準を表現する能力規準ではなく、背後に欧州評議会の理想像があって、その実現に向けて開発された複数言語を対象とした「言語能力参照枠組」である。

最初に、大塚雄作氏は本誌上討論の趣旨として、「本誌上討論においては、いくつかの状況における学力に関わる評価・測定について紹介すると共に、それぞれの課題を整理することを通して、教育心理学研究における学力評価のあり方を再考する機会を提供することを本旨としておくことにしたい。」として、「調査と選抜試験の違い」、「形成的評価と総括的評価の違い」、「非認知的能力の評価」という3つの状況を設定された。そして、それぞれについてコンパクトにまとめた解説を加えた上で、具体的な事例として柴山直氏、植阪友理氏、遠藤利彦氏の論文への導入を図っている。筆者(野口)には総括の役割を託された。柴山論文、植阪論文、遠藤論文のいずれもかなり奥が深いテーマで、執筆の方々時間がかけて実施し、成果を得られた研究がベースになり、かつ、相互に重なりを持つが異なる広範囲な研究領域にわたり、正直言って筆者には荷が重すぎる役目ではある。

柴山論文について

柴山直氏は教育測定・心理計量学のエキスパートでテスト理論面だけではなく、新潟県の学力調査をはじめとする各種の学力調査や公的な大規模試験に携わって来た豊富な経験をもとに議論を展開している。

最初に、「測定」(measurement)、「アセスメント」(assessment)、「評価」(evaluation)という基礎用語を定義した上で、測定・評価に関係する教育行政の風土を指摘した。「大規模テストを裏で支えているテスト理論およびそれに基づく学力の測定技術への無理解・誤解ないし無視の傾向が、わが国の教育行政にはある。」ということである。確かに、例えば TOEFL が(主として北米地域の)大学入学に必要な英語能力を測定する大規模

試験として優れていることを主張する人は少なくないが、その背後でしっかり支えているテスト理論の存在があるということを知る人は少ない。念のために言うと、筆者は TOEFL が大学入試を目的とした大規模英語試験で最も優れたものであると主張しているわけではない、あくまで例えばの話である。

次に、大規模学力調査を取り上げて、「大規模学力調査においても、一般的な調査デザインに加えて、学力の測定モデルとでもいうべき心理計量モデル (psychometric model) も同時に考慮する必要がある」とし、さらに「調査目的あるいは選抜なのか指導なのか、それとも診断なのかといった、実施目的にテストの設計仕様を合致させなければならない。」と述べている。「調査目的」というのは、「実態の把握」とも言えるが、この場合には調査対象は「集団」であって、「個人」ではない。したがって、調査対象者全員が同一の問題冊子に解答する必要はなく、一般に調査に使える時間は限られているので、相互に重複した項目を含むが、それ以外は異なる項目から構成された問題冊子を複数用意して調査する重複テスト分冊法を用いることによって、一度の調査でカバーできる測定範囲が広く、得られる情報が多くなるという利点がある。このことは、既に米国の学力調査 NAEP でも実績があり、いまずぐにでも実施体制を整備すればよいことであるが現実にはそうはなっていない。

また、「実験デザイン・調査デザインなしで収集されたデータ、あるいは後付けであってもそのようなデザインが設定できないデータからは、目的とする情報は得られない。それがいくら大規模であっても……」として「ランダム化比較実験や疫学的アプローチ」の利用を提案している。

最後に「大学入試センター共通テストの課題」として、「多枝選択形式」が十分使えるものであり、中途半端な「記述式」にする必要性に疑問を呈している。

いずれも、テスト理論に基づく制度設計、項目開発、テストの編集、スコアリング、目的に応じた統計分析の必要性に繋がっている。

大規模能力(学力)試験を開発し、維持するためには、資金と人材と時間とが必要である。英語能力試験の代表的な開発機関である Cambridge Assessment English や Educational Testing Service をはじめ各開発機関はテスト理論をはじめとして、測定内容、テスト編集、テスト実施など広い意味でテストに関する専門家集団が機関内に存在している。また、妥当性検証も時間をかけて実施している。一見短期間に開発されたように見える試験でも、それ以前の膨大なデータの蓄積が活

かされていることが多い。

書き出せばきりがないが、柴山氏が言う「大規模テストを裏で支えているテスト理論およびそれに基づく学力の測定技術への無理解・誤解ないし無視の傾向」は今すぐにも正さなければならないし、教育心理学という学問が関与できることが多数あると考える。

植阪論文について

植阪友理氏は教授・学習心理学が専門で、理論的研究だけではなく、実践的な研究も精力的にやって来た実績がある。今回は、「学習者が教授者が期待するような学力、学習行動、学習態度を身に付けるには、どのような評価が求められるのか」について、最初に、「日々の学習指導における評価の重要性」「評価を通じて変容させていくべき子どもの実態」を述べ、「世界的な教育動向および日本の教育課程が目指すものとの関係」の中で「学校における日々の指導のなかでどのような評価が求められるのか」について問題提起をした上で、続いて植阪氏が直接関係する学校現場での実践に即して論を展開している。具体的には中学校と小学校における実践例が紹介されている。

まず、「定期考査、授業、数学通信を一体的に連動させた実践」で、中学校における実践について詳細に紹介された。数学で定期考査に公式を適用して解を求める形式ではなく、なぜそのような解に至ったのかを説明させる形式の問題も加えて、そこが評価されることを明示的に示した。それとともに、日々の授業も「教師が意味に着目した説明を行うのみならず、子どもたちにも同様の説明を求めることとし」、授業も「考えさせること」を重視するように変化させた。さらに、「数学通信」を発行して、家庭学習の支援、間違いから学べる支援を行った。「定期考査」「授業」「数学通信」が一体となって、子どもたちの「学習観および学習方略の変容」を図ることを試みている。その結果、確かに変容が見られたが、それには時間も必要であることも明らかになったとしている。

この実践では、狙った意図が非常にうまく実現した実践であると高く評価できるし、こういう指導と評価のサイクルが十分に機能すること、それが、特定の教員に留まらず広く、他の数学教員に、他の教科教員に普及して行くことが望ましいと考える。

ただ、このことがドリル軽視にならないように配慮することが必要のように思える。公式の導出過程を理解することは大切だし、応用問題でも解答の過程が説明できることは大切である。ただ、公式を自動的に使えることもまた現実世界では大切な事である。もちろ

ん植阪氏や床教諭はそのようなことは十分分かって、このような実践をし、研究を進めていることは疑うものではない。ただ、これがいわば創始者の手を離れて広く普及して行く過程で思わぬ副作用があることに気を配らないといけないと考える。筆者（野口）は教育における「振り子現象」と勝手に呼んでいるが、どうもわが国では教育に関しては議論が極端から極端に振れる傾向がある。「ゆとり教育」に大きく振れたかと思うと、こんどは逆に振れている。振り子が振り切れるまで振れるのが特徴で、AかBかの二者択一になる傾向がある。AもBもそれぞれ理由があって出て来たものであり、それぞれの長所と短所とがあるはずである。また、適用する状況によって最適解は変わるはずである。AからBへ移行してもAのよかった点は十分に踏まえ活かすべきである。

そう考えると、ここで提案された「過程重視」と「ドリル学習」とをうまくバランスを取って教育を進めることが大切であろう。「原理」が分かることは重要だが、常に「原理」にもどるのではなく、自動的に適用できることもまた大切であろう。

そして、「小学校の実践可能性について」では、業者テストが使われることが多い小学校においては中学校での実践をそのまま取り入れること難しいと考えられるが、その点を解答の筋道を記述するための白紙をテスト用紙と同時に配布して、その解答に至った過程を児童に記述させ、部分点を設定することで解決することを試みている。これは優れたアイデアだと評価するものであるが、小学校と中学校との間の発達段階の違いや教科内容の扱い方、また、教科書の文体の違いなども考慮に入れる必要があるかもしれない。「考え方」「考える過程」が大切なことはその通りであるが、「書く」という作業が逆にそちらに児童の意識が集中して自由に考えることの妨害要因として働いてしまう可能性もあるのではないだろうか。白紙ではなく、若干誘導する文言を入れることも考えられるし、算数の文章題などでは、普段の授業時から複数名の児童に黒板に解答を書かせて、なぜそう考えたのかを口頭で発表させることで、いわば「くせ」をつけておくことも必要かもしれない。

また、学年により、教科により、単元により「考える過程」の重要度は異なって来るであろうし、児童の日本語能力や学級集団の特性を考えあわせながら、きめの細かい運用が望まれると思われるし、それは今後の課題と言うべきことだと言えよう。

ここで提案された実践例を今後どのように広げて行くかが課題になる。本当にうまく行った例であるので

どこでも、誰でもこのようにうまく行くとは限らない。

これは筆者(野口)が否定的に考えているのではない。そうではなくて、全国の中学校や小学校はさまざまな固有の事情を抱えているはずで、そこにどのように「調整」して導入し、普及させるかということが大きな課題であろう。教科による違い、学校環境の違い、教師集団の考え方、児童・生徒集団の違い、保護者集団の違い、が関係する。しかし、言い出せばきりがない。今後の展開を期待したい。

遠藤論文について

遠藤利彦氏は、発達心理学、中でも幼少期の育ちが人生のその後に及ぼす影響について研究を進めて来た方である。今回は、柴山論文および植阪論文が人の「認知的側面」の測定・評価について扱っていたのに対して、それとは異なる側面の測定・評価について、教育経済学の最近の展開と対比しながら論を進めている。

最初に「心理学における『非認知』なるものへの刮目」で、最近の教育経済学で「幼少期に培われる『非認知』的な心が、その後の学力の形成や職業上の成功なども含め、広く様々な心理社会的適応性やウェルビーイングの実現において、重要な鍵を握っている」ことが注目されているのに対して、心理学ではThorndike (1920) などの研究者が既に「認知」的能力以外の「心の性質」にも関心を向けていた。ただし、教育現場への実践的应用という文脈では「認知」的能力ほどには注目されて来なかったことを指摘している。

そして、「IQのアンチテーゼとしての感情的知性」で、「人間の感情的知性=EI (Emotional Intelligence)」が人の社会的な成功や幸福に純然たる知的能力よりも影響を及ぼすというEIの流れについて紹介し、これが、「一般的に子どもが備えておいて然るべきとされる『非認知』的な心理的性質と、元来、概念的にかなり大きな重なりを有する」としている。

さらに、Heckman (2013) の考え方による縦断的調査研究「ペリー就学前計画」の成果が「非認知」的な心の性質の重要性について極めて消極的な形でしか示していないことを紹介している。

以上の議論を踏まえて、「あらためて『非認知』なるものとは何か？」で『『非認知』的なもの』に対して、より具体的な定義を与えている。遠藤(2017)が、社会情緒的コンピテンス(一般に非認知的とされているもの)を「社会性」に関わる心の性質と「自己」に関わる心の性質と定義しているが、これらが社会的認知と言われるものを基盤に成り立っていることから、心理学的に厳密には「非認知」ということが不適切であることを指

摘している。(教育)経済学的な考え方と(教育)心理学的な考え方、あるいは、現象のとらえ方に違いがあり、心理学的には「非認知」という術語の使用に慎重であるべきだというのである。

「『非認知』なるものと学力との関連性」で、遠藤(2017)を概括して、教育経済学的研究で、特定のパーソナリティ次元と学歴、収入、社会的地位との関連性を明らかにしているものがあるのに対して、「パーソナリティを教育的介入の対象とすることは、現実的な意味で、また、倫理的な意味でも、あまり適切ではないと考えられる。」という批判的見解を出し、パーソナリティを変えるのではなく、「安定した属性と見なし、それに応じた社会情緒的コンピテンスを教育的介入によって身につけさせることの方が現実的に有効であるとしている。

確かに経済効果を目的変数として個人のパーソナリティに介入することは、個人の生き方や生活に大きな負担を強いることになり、そういうケースが増えると目的とした経済効果がむしろ社会全体として低下してしまうことになりかねない危惧がある。また、国際化(敢えて「グローバル化」とは言わない)が進む時代とは言え、国・地域の文化差や価値観の違いもあり、確認された効果が当該研究の対象とした国・地域を超えて有効性あるいは一般化可能性を持ち得るかにも注意を払う必要があると考える。

最後に遠藤氏は「学力の背景に潜んで在る『非認知』に対する見極め」で、「長期目線で見た場合には、社会情緒的なスキルやコンピテンスに裏打ちされた学力でなければ、個人の生涯に亘る心理社会的適応性の具現には必ずしも直結しない可能性がある」としている。それと同時に、社会情緒的なスキルやコンピテンスの測定と評価の困難さも指摘しているが、それでもこれらを踏まえた個別化された教育を目指すことをこれからの教育心理学における重要な課題の1つであると、結んでいる。

概して短期的な効果を問題にしがちな(それが悪いということではない)測定と評価であるが、そこに発達という軸を入れて測定・評価することの必要性、意義は十分に理解できる。ただ、「ペリー就学前計画」のように、限られた研究参加者を長期に追跡しても、個人の置かれた環境をすべて知ることはできないし、その間の社会変動なども考慮する必要があるが、その影響は個人によって大きく異なる可能性が高い。そうすると、どのような変数を調査研究の中に取り込むのが結果を解釈する際に大きく影響してしまうし、想定した変数を測定道具が適切に測定しているかという妥当性の問題、そもそも、各時点で一定の精度を持つかと言う信

頼性の問題がある。といって、妥当性と信頼性の高い測定結果が得られる変数だけを問題にしていると、何か大切な要因を落としてしまう可能性もある。遠藤論文に見られるような理論的に考察を進めることが重要である。

大規模な長期にわたる追跡は OECD のような機関が主導して研究を進めるしかないように思うが、経済的な豊かさが必ずしも個人の幸福、あるいは、幸福感に対応するわけではなく、心理学、特に、発達心理学的な考え方を基礎にした調査研究を大きな公的機関が核になって実施することが必要で、そこで、evidence based な成果が得られる、という体制が組まれることが望まれる。

総括すると

柴山論文、植阪論文、遠藤論文と見てくると測定と評価とは教育心理学が対象とする広範囲な場面で関係して来ることが分かる。「測定・評価」と一括りにされることが多いが、測定では客観性が重視されるのに対して、評価は測定された結果に対して価値判断が加わるという意味で主観性が強い。同じ測定結果から異なる評価があり得るということである。そして、何のために測定・評価するのかという目的に応じて、測定道具に要求される条件は異なって来る。測定する構成概念を明確に規定して、精緻な測定（計量）モデルが必要な場合もあれば、測定結果から何らかの診断結果が得られるような診断（分類）モデルが必要な場合もある。また、構成概念自体が理論的に確定できない段階にある研究で、それを明確化して行く過程で測定する場合もある。ただ、測定目的によらず、測定道具は妥当性と信頼性について何らかの証拠（evidence）が示されていることが必要である。

また、評価をするためには、その評価の適切さについて、理論的に何らかの根拠をつける必要がある。

測定結果や評価結果をどのように用いるかによって、その厳密度は異なるが、一般に「心理尺度の開発」という系統の研究論文のすべてが、これらのことに十分配慮し、evidence を十分に公表しているという訳ではないように思われる。通常の手順で統計分析をして、しかるべき指標を示してはあがあるが、測定している構成概念が確かに測定結果に反映しているかどうかに関する検討が十分でないものも見られる。もちろん、さらに研究を進めて行く途上での報告という場合もあるかと思うが、その場合は、そのことを明示した上で、今後の課題について、論文中にしっかり記述されていることが必要である。

また、測定や評価の道具は正しく理解して使用する

ことが必要である。最近話題になることが多い CEFR、すなわち、「欧州共通言語参照枠組」（Council of Europe, 2001）であるが、これは欧州域内の諸言語の学習者が学習している言語（外国語）でどの程度の熟達度にあるかを言語間で共通の 6 レベル（A1, A2, B1, B2, C1, C2）で表し、レベル毎に学習している外国語で具体的にどのような言語行動が可能であるかを Can-do statements で例示しているものである。いわばレベルイメージを表していて、学習している外国語によらず学習者の熟達度を表わすことができる。この CEFR はあくまでふわっとした「レベルイメージ」であって、あいまい性を残したものである。そのことが「参照枠組」という名称に反映していて、決して、測定や評価の「規準」「基準」を表わしているような厳密性を持ったものではない。現在わが国では、CEFR を用いて特定の外国語に関する大規模テスト間の測定結果（得点）を相互に対応づけることができるような誤解が一部であるようだが決してそのようなものではない。あくまで、外国語学習者の熟達度レベルがイメージできるように開発された「欧州言語共通参照枠組」である。言語は異なってもよい。例えば英語など、特定言語のテストの測定結果を評価したり、特定言語の複数のテスト間の得点を対応付ける道具ではないということに注意して欲しい。

「測定・評価」は大切だが、その道具は目的を持って作られている。目的に照らして適切に使用してこそ意味があることを肝に銘じなければならない。

引用文献

- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Strasbourg: Council of Europe. (吉島茂・大橋理枝（訳・編）(2004). 外国語教育Ⅱ—外国語の学習、教授、評価のためのヨーロッパ共通参照枠 朝日出版社〔原著の Appendix C: The DIAL-ANG self-assessment scales と Appendix D: The ALTE 'Can Do' statements は含まれていない〕)
- 遠藤利彦〔統括〕(2017). 非認知的（社会情緒的）能力の発達と科学研究手法についての研究についての報告書 国立教育政策研究所・平成 27 年度研究プロジェクト報告書
- Heckman, J. J. (2013). *Giving kids a fair chance*. Cambridge, MA: MIT Press.
- Thorndike, E. L. (1920). Intelligence and its use. *Harper's Magazine*, 140, 227-235.