

杉並区独自の学力調査である「特定の課題に対する調査」は、潜在変数モデル(LVM: Latent Variables Model)のうち潜在変数に順序(尺度)を仮定した「潜在ランク理論(LRT: Latent Rank Theory)」を設計思想とし、「潜在ランク尺度」で「目標準拠の絶対評価」を実現するメカニズムをもって運用しています。今年度の「編集後記」は、本報告書の参考資料(pp. 177-196.)にも記したこの点を学術的に補説します。

### (1) テスト理論における三つのパラダイム

さて、学力をはじめとした潜在特性を測定するテスト、その標準化＝尺度化と等化のためのテスト理論には、大別して三つのパラダイムがあります。提唱順に並べるなら、「古典的テスト理論(CTT: Classical Test Theory)」「項目反応理論(IRT: Item Response Theory)」「ニューラルテスト理論(NTT: Neural Test Theory)」となり、LVMで分類すると、CTTとIRTは連続(尺度)、NTTは順序(尺度)を潜在変数に仮定する理論です。なお、NTTはLRTの下位モデルの一つであり、統計メカニズムとしては自己組織化マップや生成トポグラフィックマッピングを利用します。

詳細な説明は省きますが、CTTには、標準化の知見が被調査者集団の特性分布＝当該のデータに完全に依存する限界があります。この限界を乗り越えるIRTは、異なる被調査者が、異なる日時・場所で、異なる設問の調査を受けても結果を比較可能な尺度を構成できるため、経時変化の考察などに優れた効力を発揮します。最後発のNTTは「潜在変数に順序を仮定したIRT」と解説される場合があることから、IRTとほぼ同じ利点を備えます。ここで、学力など認知能力を測定するテストないし調査は、誤差Eの混入によって測定値Xの1単位の差が真値Tの能力差と(ニア)イコールで結べるほど解像度が高くないと考えることもできます。NTTがランクの間隔の等しさを前提しない「潜在ランク尺度」を仮定し、対象者をざっくり段階分けする理由はここにあります。

しかしながら、私たちは、IRTとNTT、あるいはCTTさえも含み、三者のいずれが最も優れているのかという素朴な議論に回収されてはなりません。まずもってここでは、三つの理論が、テスト・調査の「目的」が何であるか、どの程度の時間や設問数で実施するか、利用可能な計算機資源はどの程度なのかといった「状況」に応じて選択又は併用されるべきものであることを押さえておきたいと思えます。

### (2) 集団準拠の相対評価と目標準拠の絶対評価

上記は、テスト・調査を運用するに当たっての「測定」「統計解析」の段階を担う理論群です。ここではさらに、測定・統計解析結果の教育的価値付け、つまり「教育評価」についても理解しなければなりません。私たちにも馴染み深い「集団準拠の相対評価」と「目標準拠の絶対評価」がその考え方です。言うまでもありませんが、後者の「目標」は一般に「学習指導要領」(が定める目標や内容・事項)を意味します。観点別学習状況評価や総括評価としての評定は目標準拠の絶対評価を基本としますが、集団準拠の相対評価が教育現場から完全に排除されるわけではありません。少なくとも原理的には、「自己内評価」も合わせ、目的と状況に応じて選択・併用されるべきものです。

このことを踏まえて解説すると、CTT が根底に置く連続尺度上の  $X=T+E$  における偏差値 = 70、IRT が推定する潜在的な連続変数  $\theta = 2.0$  といった値は、それ自体が教育的に絶対的な意味をもっていません。もう少し具体的に言うと、偏差値 70 という値からは、当該個人が調査集団内で相対的に上位だろうという価値しか導けません。 $\theta$  は多くの場合 -4.0 から +4.0 に収まるため、2.0 と推定された個人は当該尺度上で高度な能力を有すと評価できそうです。しかし、 $\theta$  自体に目標準拠の絶対的な意味はなく、したがって習得・育成すべき能力が目標に達しているかどうかを価値付けしたことにはなりません。そもそも測定値の(僅)差は、教育的に意味のない誤差の可能性もあります。

### (3) 潜在ランク理論と目標準拠の絶対評価による学力調査

そこで、杉並区の学力調査です。LRT を設計思想とし、NTT の統計メカニズムを踏まえつつ、潜在ランク(尺度)としての学力段階 R が目標 = 学習指導要領準拠の絶対的な意味をもつ特徴を備えます。調査の主な目的は、報告書冒頭にも記すように、児童生徒、教員、教育行政の三者が結果を活用することであり、入学試験のように受験者を順序付けたり選抜したりすることではありません。状況として特に考慮した制約は、年度ごとに各教科等の調査を「1 単位時間で実施する」ことから「最大 25 問ずつで構成する」こと、それゆえ「測定値には多分に誤差が含まれる」ことです。LRT と目標準拠の絶対評価は、これら目的と状況から選択した調査運用の考え方ということになります。

そして、このことは、「学校評価」や「教育行政における点検及び評価」の KPI とするため、調査結果を一つの値に集約する場合などに決定的な意味をもちます。本区調査のメカニズムなら、測定誤差を十分考慮し、「全児童生徒が到達すべき最低限の水準を『R3 おおむね定着がみられる』とする」「中学校第 3 学年における R3 以上の割合を(最低でも) 80%以上にする」といった KPI を導出できます。しかし偏差値や  $\theta$  は、何らかの絶対的基準がない限り代表値や散布度を算出して相対評価をするか、あるいは自己内評価を行うほかに主だった選択肢がありません。普遍福祉という正当性の原理から“目標準拠”に“全て”の児童生徒の状況を問う、すなわち、中下位層に集中投資して平均を上げるといった“ごまかし”のきかない調査の設計と KPI の設定。学校や教育行政の別なく第一義に重視すべきものがここにあることは、論を待たないでしょう。

### (4) 学力調査の構造転換、新しい無限の領野へ

もちろん、連続尺度や集団準拠の相対評価に全く価値がないわけではなく、連続尺度型のテストが動機づけによいことなどは参考資料にも記したとおりです。そして、計算機の存在が“自然”となる未来では、調査自体が日常の学びに“遍在”し、テスト理論などもエンドユーザが意識することなくその都度最適なものが自動選択・併用されるようになるでしょう。これはすなわち、既知の理論群をメタから捉えた新たなパラダイムへの移行、学力調査にまつわる「原罪」をも克服する構造転換を意味します。

杉並区では、CTT、IRT、NTT を経て、2012(平成 24)年度に現在のメカニズムで調査の運用を始めました。ここまでの道のりも険しいものでしたが、今、眼前には、公教育の構造転換の一端となる学力調査 = 日常の学習評価の「無限の領野」が広がっています。その新たな地の探究へと進み行く一步を、私たちは、力強く踏み出します。

平成 30 年度 杉並区  
特定の課題に対する調査、意識・実態調査 報告書

平成 30 年 11 月 発行

登録印刷物番号  
30-0059

調査機関・編集発行 杉並区立済美教育センター  
〒166-0013  
杉並区堀ノ内二丁目 5 番 26 号  
電 話 03(3311)0021



