

全国的な学力調査に関する専門家会議（第6回）

1 日時：令和2年8月31日（月曜日）10:00～12:00

2 場所：Web会議（文部科学省 東館9階 総合教育政策局会議室）

※Webexにて公開

3 議事：

(1) 令和3年度全国学力・学習状況調査について

(2) 全国的な学力調査のCBT化検討ワーキンググループ 中間まとめ「論点整理」について

4 出席者：

大津座長代理、青木委員、鎌田委員、川口委員、柴山委員、垂見委員、土屋委員、平川委員、益川委員、村山委員、吉村委員

5 議事要旨：

議事1：令和3年度全国学力・学習状況調査について

・資料1に基づき、事務局より説明があった。関係委員の意見は以下の通り。

【委員】 質問紙調査の項目について、新型コロナウイルス感染症による休業措置の期間中の学習面への影響を明らかにするため、調査項目を見直すことは大いに賛成である。本年度調査の中止の判断理由を考慮しても、様々なデータを得ることは、こういった危機のときほど有用だろう。長期にわたる学校休業で学力や体力が低下した、社会的不安が高まっている、心の温度が低下した、などと言われている。そういった予測に対して、検証によってその根拠を明らかにし、解決のための手段を考えるためには、やはり何らかのデータが必要になるはずである。そのため、来年度はぜひそういった質問項目を実施していただきたい。

議事2：全国的な学力調査のCBT化検討ワーキンググループ中間まとめ「論点整理」について

資料2にもとづき、WGの大津主査より説明があった。関係委員の意見は以下のとおり。

【委員】　　ここ数日、中間まとめに関するマスメディア等の報道を見ていると、議論の一部しか伝わっておらず、サーバの負担やそれに伴う悉皆調査の見直しが必要だ、といった部分が主に取り上げられているように感じる。

悉皆でのCBT調査は、実施方法の可能性として、各学校に一人一人の児童生徒にUSBを配る方式、あるいはオンラインで一斉にアクセスする方式が考えられる。前者の場合は、USBを配るのは学校の先生方の負担が大き過ぎること、後者の場合は、約200万人の小学6年生と中学3年生全員同時にアクセスする通信環境を用意することが可能だとしても、年に1日の調査のために毎年サーバを維持するのはあまりにも無駄が大き過ぎることから、現実的な選択肢ではないだろう。

CBTの場合、一斉実施でなくとも異なる日に受験することや、一人一人の学力に合わせた問題の出題も可能であるため、こうした技術を利用すれば、先述したサーバ問題は解決する。しかし、その場合は、異なる日に受験している児童生徒に同じ問題を出すわけにはいかないため、異なる問題を出題する必要がある。そうなってくると、違う問題を受験した児童生徒の点数をどうやって比べるのか、という新たな問題が発生し、これを解決していくためには、IRT（項目反応理論）というテスト理論が要するという流れで議論が進められている。

項目反応理論を導入するには、それに精通した専門家が必要になり、非常に精度の高いテスト項目が数万題必要である。そして、テスト項目が外部に流出すると、特別な不正をする人が出てくるため、テスト項目を非公表にすることになる。そうすると今度は現場の指導改善のために、調査問題を使うことができなくなる。

全国学力・学習状況調査をCBT化することは構わないが、それを悉皆で毎年実施するとすると、以上のような問題をすべて解決するために、相当なマンパワーが必要になる。このようなマンパワーを確保するのも現実的ではないため、全国学力・学習状況調査のCBT化を検討する上では、そもそも根本的に何のために調査を実施するのかにというところまで立ち戻って考えなくてはいけない、というのがワーキングにおける大きな議論の流れである。少なくとも紙で実施している現行の調査のように、指導改善への活用と政策への活用を両立することは難しいのではないか。

また、ここまで説明してきたように、CBTはいろんな論点が絡んでいるので、一部だけを切り取って報道することも、一部だけ切り取って議論するのも問題ではないか。CBT化することイコールIRTの導入ではないが、項目反応理論というテスト理論がCBTを十全に活かす

ためには必要であり、項目反応理論を導入するには、現在の悉皆実施の体制ではさまざまな問題が生じるということは前提に議論していくべきではないか。

【委員】 大規模な学力調査や社会調査、テスト理論を組み合わせた、いわゆる学力の調査測定技術と、それを裏で支えるヒューマンリソースが必要である。全国学力・学習状況調査のCBT化は、学力の調査測定技術の中に、さらにデジタル化、IT化といった最新の情報処理能力、それから、情報通信技術を組み込んでいかなければならない。

結果、ハードウェアやソフトウェアだけではなく、人的パワーをどうするかを含めて、非常に複雑な連立方程式を解くことが求められている。この中間まとめでは、ワーキンググループで様々議論をしてきたし、今も喧喧諤諤議論しているが、その複雑な連立方程式が論点という形で分かりやすく整理されている。

それぞれの論点を読んでいくと明らかだが、CBT化には、当然のことながらメリット、デメリットがそれぞれある。その中でメリットを最大限生かすには、いかなる技術が必要で、そこにはどのような限界があり、どのような技術をこれから獲得すべきか、という点が報告されている。親会議であるこの専門家会議では、この論点整理を、今後の具体的な議論の前の技術的な見通しを提供している資料、あるいは手引きとして検討していただきたい。例えば、GIGAスクール時代において、全国学力・学習状況調査にどのような機能や役割を持たせるべきかなどの議論の際に活用していただければと考えている。まだまだ詰めないといけない論点もあるため、ひとまずの報告である。

【委員】 ICT化が進んでいる現状を踏まえて、CBTによる学力調査に移行していくことは必然性をもった流れだと認識している。その上で、学校現場の現状を考慮しながら、実現に向かうことが必須であり、この観点から意見を述べる。

各論の1について、新型コロナウイルス感染症拡大を受けて、GIGAスクール構想に向けた取組が各行政において前倒しで進められている。本市においても、今年度内に1人1台端末やネット環境の整備が進められることになった。こうした流れの中で、CBT化を進めることは様々な点で時代のニーズに合った方法であるだろう。

これまで小中学校には、パソコンルームが設置されており、主に技術・家庭科の情報の授業や、各教科、総合的な学習の時間などでの調べ学習で活用しているが、日常的な学習で活用しているというわけではない。本市では、昨年度から各学校に21台のタブレット端

末が配布されているが、Wi-Fi等のネット環境が十分に整備されていない。

一方で、東京都内では、既に1人1台端末やネット環境の整備が完了している地区もあり、地域の整備状況の格差は大きいと言えるだろう。また、同じ地域内でも、学校によってICTの知識や技術に差があって、学校間格差も否めない現状がある。これらの理由から、全ての児童生徒や教師が日常的にICT機器を活用した学習やパソコン操作に慣れることにはまだ時間がかかると想定される。

また、若い世代の教員は比較的ICTに慣れている傾向もあるものの、ベテランの教員の中には、ICTを苦手とする教員も少なくない。教員の働き方改革が進められている中で、これらの習得にもさらなる負担がかかっていくことも懸念される。

各論の2について、実施にかかるコストや手間の軽減、採点の速さや多岐にわたる分析など、CBT化することによるメリットは大きいだろう。また、複数回の実施であれば、突発的な事情で実施できない学校も実施できることや、色覚等の障害のある生徒に対する合理的配慮に関する手立ても必要だが、一方で、不登校生徒や書字障害等でこれまで実施が困難だった生徒も参加することができる可能性も考えられることなど、実施の可能性が広がると思われる。

一方で、全国200万人規模の一斉調査となると、サーバ等の負担が大きく、万が一不具合が生じたとしても、それに対応できる技術的なものは、現状では学校には難しいと思われる。そうした際に遠隔で指示を出したり、画面の遠隔操作を行ったりするヘルプデスクなどのサポート体制を構築することも方策の1つではないか。

また、同一日時で一斉実施するのではなく、一定期間に複数回実施することになると、問題を非公開としたとしても、何らかの形で情報が漏れることがあれば、評価の信憑性に影響するだろう。複数問題セットによる実施が可能であれば、各地区や学校ごとに日程の選択が可能になるが、作問の負担が膨大になるのではないか。

各論の3について、問題用紙等の印刷や配達等のコスト削減には大きなメリットがあり、配送の受渡しや管理などについても学校の負担は軽減される。また、実施後すぐに結果のフィードバックがあれば、短期間で学習の改善につなげることが可能になる。

実施方法のうち、USB方式は、経費の問題のほかに事故防止等の観点から、既に使用できないとしている地域もあると聞いているため、ネットワークを利用する方法のほうが現実的ではないか。

調査設計との関係について、東京都教育委員会では、小学校第5学年と中学校第2学年の

児童生徒に、悉皆で学力向上を図るための調査を実施している。論点整理における様々な課題を考えると、国は全国的な状況把握のための抽出調査を行い、地方自治体との役割分担をすることも方策の一つとなるのではないか。個々の児童生徒の指導改善等に活用するために、例えば、全国的な学力調査を実施する一定期間の後に、問題の一般公開をすることができれば、その問題を各学校で取り組み、参考となる結果を用いることで、指導改善につなげることもできるのではないか。

このコロナ禍の状況下で、学校におけるインターネット環境や端末の普及が進んでいることを考えると、今後ますます児童生徒の学習のオンライン化が進んでいこう。全国的な学力調査についても、様々な課題を克服しながら、国として最善の方法で開発されることを願っている。

**【委員】** 短期間のワーキンググループにもかかわらず、中身の濃い議論があったことを感じる。とにかく大事にしないといけないこととして、円滑な実施が挙げられる。

1つには、現場の先生方への応援、これをしっかり丁寧に分かりやすくやっていく必要がある。先生方の中でもリテラシーの差が大きく、くれぐれも苦手な先生と得意な先生との間で軋轢が起きるようなことのないように配慮が必要である。児童生徒もCBTに不慣れであることが結果に影響することがないように、しっかりとした配慮や工夫、時間の確保、工程表のようなものが必要だろう。

CBTは、多くの問題作成が必要なので、問題作成の体制がしっかりと確立をされて、無理な形にならないようにしていかなければいけない。

期待と懸念を持っているのは、指導改善である。例えばログが分かって多面的な分析ができる、というように指導改善にプラスになる方向で検討していただきたい。この指導改善の側面が弱くなると、調査結果を児童生徒の学力、能力の問題だけにしてしまう可能性が高くなってしまふ。それは避けなければならない。

調査の一つの面として、学校と先生方が自身の指導のどこをどのように修正していけばいいのかという、こういうPDCAの流れの中で見ていくよさがあると考えている。つまり、指導改善というのは、もちろん1つは児童生徒のためであるが、もう一つは、学校の先生方が教師としても人間としても成長していくためにも大事なことである。

私の現場経験では、児童生徒を守り育て続けた偉大な先輩たちというのは、うまくいかなかったときに、それを児童生徒のせいには絶対にしなかった。自分たちのどこをどう直

していけばいいか、こういう分析、思考をいつも大切にされていた。逆に、うまくいかないことを一旦児童生徒のせいしてしまうと、保護者のせい、管理職や教育委員会のせいにもしてしまう。教師が指導改善を通して自分をアップデートできるかということは、学習者理解と並んで、学校の先生の生命線の一つだと考えている。

それだけでなく、正答率、前年度の結果、その公表と、現場にはバイアスがかかっており、その中で過去問練習対策を行ったり、調査問題を分析しているHOW TO本に飛びついてしまったりする傾向も残念ながらある。指導改善を弱めてはならないということに気をつけなければいけない。もし、指導改善の方向性が弱くなるとすれば、どの程度弱くなるのかということ想定し、把握し、別途それを補う手を打たないといけないと考える。

資料3において、今後の検討の中に「全国的な傾向の把握を目指すのか、児童生徒・学校単位などでのきめ細かい指導の改善に活用することを目指すのか、年度間の学力の変化等の把握もできるようにするのか」とあるが、実際はどれも大事である。国と地方自治体との役割分担という意見も出ており、抽出もしくは悉皆、またはこれらの組合せみたいな意見も出ているが、こういった意見も含めて、いろいろ知恵を出して、それぞれを対立させずに、どれも大事にできるような総合的な判断をしていきたい。

【委員】 短期間での論点整理、網羅的に様々な観点がまとめられており、大変参考になった。私の方からは大きく3点申し上げたい。

まず、学力調査の目的について。論点整理の総論であったように、「調査の目的と実施方法などが表裏一体の関係にある」ことから、調査の目的を整理して決める必要があるのではないかと。学力調査は開始以来、全国的な傾向の把握、きめ細かい指導の改善に活用することを目的としてきたが、この2つの目的を1つの調査に、同じ重みづけで盛り込むことには限界があるのではないかと。学力調査が平成19年度から10年以上続けてきたことを考えると、ある意味で別の段階のような位置づけで、思い切って一つに絞った方が良いのではないかと。私自身は、全国的な傾向の把握を目的とするということに絞る、あるいは重みづけをしたほうが良いのではないかと考えている。理由としては3点考えている。1点目は、国がやる意義、地方自治体が行っている学力調査との役割分担を明確にし、国にしかできないことをやるべきではないか。2点目は調査設計をする上で、2つの目的を追うことでどうしても、どちらも中途半端になってしまうところが否めないのではないかと。3点目はきめ細

かい指導の改善の目的を外すことで問題が非公開になり全国一斉開催の必要がなくなった場合、一定期間内に複数回実施することで、より柔軟な実施が可能になり、現場の教員の負担やスケジュール調整の緩和ができるのではないか、という点である。この3つの理由から、学力調査の目的を全国的な傾向の把握に重きを置いた方がいいと考えている。

第二に、IRT導入について。学力調査の目的を全国の傾向把握に絞った場合、データを重視したEBPMの推進という観点から、経年比較、基準集団を定めた年度間比較をしていくことが重要になるのではないか、また、学力の伸びや、学力格差の拡大・縮小というところに焦点を当てていく必要があるのではないか。そういった目的を考えると、今回検討しているIRTを導入することが望ましいのではないかと考えている。IRTを導入していないために、厳密な経年比較ができないというのは、現行の学力調査の大きな弱みではないかと考えている。IRT導入の大きな利点として、1つは、経年比較を厳密にできるようになるという点と、もう一つは、調査問題を複数セット用意するために、幅広い領域の問題が出題できるようになる、こういった利点を強調したい。ただし、その場合は、論点整理でも詳しく書かれているとおり、原則、問題の非公開、そして多くの問題を作成、プール問題を蓄積する必要性が生じる。問題バンクの作成や問題の漏えいの対応策等について、時間をかけて、適切な方法を議論していくことが重要である。

第三に、CBT化に関しては、慎重に、ただ、積極的に検討していくことが望ましいのではないかと考えている。最初の慎重にというのは、調査の規模に関しては、調査目的によって結果的に異なってくると思うが、仮に現行の200万人を対象に一斉実施した場合、現在のスキームのまま実施すると様々なトラブルが想定され、学校現場にも相当負担をかけることになる。小規模の予備調査をしながら、課題を整理し、タイミングや手法を決めていくのが望ましい。

一方で、やはり積極的に議論していく必要がある。1つは、CBT化することは、長い目で見れば、学校現場の負担軽減につながる。コロナ禍での大学のオンライン授業もそうだが、初期はやはり様々な投資、調整、対応が必要で、実施する現場としては、大変手間がかかった。ただ、複数回繰り返していくことによって、かなり大学の授業でも効率化できることもあることがわかってきた。

また、社会的にパソコンを推進するために、取りあえず調査にCBTを取り入れるというスタンスは、もちろん望ましくないだろう。一方、やはりパソコンリテラシーの向上を様々な学びで取り入れていくのが必須であるというメッセージを、学力調査を通して打ち出すことに意義があるのではないかと考えている。例えば、論点整理でも挙げられていたが、画面の切り替

えやスクロールをしながら長い文章を読めることが習慣づいていなければいけない、あるいはローマ字、平仮名の切替えを自由にできなければいけない、そういった意識がメッセージ性の高い全国学力調査を通して、学校現場に浸透、定着していくということには意義があるのではないか。PCリテラシーの向上について、いろいろなところで論じられているが、日本の小中高のPCリテラシーは、パソコンのハード面、ソフト面でも非常に遅れている。ネット上でのチャットやゲームがOECD諸国同様、利用している、あるいは近年非常に増加している一方で、PCなどの学校での利用、あるいは学校外でPCを学びに使う利用というのが、OECD諸国に比べて極めて低い数値となっていることがPISAデータから示されている。この結果が発表されたときに衝撃的な数値と受け止めたが、大学生を見ていると、非常に納得いく数値でもある。スマートフォンは巧みに使えても、PCリテラシーが極めて低い学生が一定数見受けられる。その結果として、大学のオンライン授業においても、現場では驚くほど様々なレベルの支障が生じる結果となった。オンライン化が進む中で、スマートフォンが自由に使えても、PCは自在に使えない層が一定層いるというのは、やはり致命的な問題であり、そういった情報活用格差がこの知識社会においては、生涯の機会の格差にもなるのではないか、そういった意味で、CBT化を積極的に議論する必要があるのではないかと考えている。論点整理の中でも、情報活用能力の差異が調査結果に反映されてしまう、あるいは、ICT機器に慣れていない児童生徒は、測りたい学力を測定できないのではないかという指摘がある。情報活用能力の差異を早急になくし、全ての児童生徒が学校でのカリキュラムを通して、電子機器を用いた日頃の学習がなされ、画面をスクロールして読めるようになる、キーボードを自在に入力できる、自由に情報収集ができる、そのような状況になった上で学力調査のCBTが実施される、そういった筋道を考えることが望ましいのではないか。

最後に、調査目的を決めることがこのCBT化を議論する上でも重要だと思っている。この点を改めて協調しておきたい。調査目的によって、調査の規模や調査設計も大きく変わってくるだろう。例えば、この論点整理において、CBTの実施方式として3パターン挙げられているが、どれにするのかを決める上でも、やはり主要な点は、調査の規模で、その調査の規模を決めるのは調査の目的であるべきだ。調査の規模を決めるのはCBT化の有無ではないという点も強調しておきたい。また、全国学力調査の目的の見直しにしても、あるいはCBT化にしても、それぞれの分野で精通している人材を一定数配置する必要があり、そして、その予算を確保することが不可欠である。そのようなコミットがない限りは、議論の



ための議論に終わってしまうのではないか。

【委員】 論点整理について、非常に多角的な観点からまとめていただいた。

今後の世の中を考えると、CBT化は必然的な流れであり、特に経年変化を見るという観点からIRTの導入は望ましいことだと思っている。これを前提として、3点申し上げたい。

まず、調査の目的と方法という整合性が取られるべきだろう。IRTを導入することによる1つのデメリットとして、これまで指導改善に生かすという大きな目的があったが、具体的な調査問題が提示されないことになると、誤答があった場合、どこの問題が誤答だったのか、あるいはどういう形での誤答だったのかがわからない。誤答の分析は、指導改善に非常に役に立つ情報と思うが、具体的な問題が公表されず、それぞれの児童生徒がどの問題に答えたのかが分からないという形になると、指導改善に生かすには情報不足である。だからといって、IRTを導入すると指導改善に生かすことが難しいから調査目的から外す、つまり方法論から逆に調査の目的を決めるというのは本末転倒である。まずは目的を決める、その上で、具体的な方法を決めるというのが筋ではないか。

次に、IRTを導入したことによる結果のフィードバックについて、述べたい。これまで正答率での結果返却は非常にわかりやすいものだった。いわゆるIRTのシート、能力値をフィードバックする場合、例えば、ある児童生徒のシートが700、学校平均が700、あるいは都道府県の平均が700だったとして、具体的な判断基準がないために、この700の解釈が難しい。1つ考えられるのは、全国平均が500だとすると、700という結果は、平均を上回っていることは分かる。しかし、全国平均を上回っている200というのをどう解釈したらいいのかわからない状態である。これまで全国の学力調査は、学校、あるいは都道府県の序列化を避けるということに非常に気を配っているような工夫をしてきたにもかかわらず、シート、あるいは能力値のみで何の手がかりもない状況だと、隣の県や隣の学校と比べて判断してしまう可能性がある。結果の能力値が分かりにくいということから、避けようとしてきた学校、あるいは都道府県、児童生徒間も含めて、そういった序列化が助長されてしまう可能性があることを懸念している。IRTの結果をどういうふうにフィードバックして、これを理解してもらうのかという点について、検討が非常に必要になってくるだろう。

最後に、質問紙調査について、令和3年度は一部の学校で、オンライン調査で実施予定とのことだが、これまで紙で行っていた調査をオンラインに移行することによって、これまでとのつながりが取れるか、そこで時系列的な断層が生じないかという点について、

あらかじめ慎重な検討が必要ではないか。これまで、小学校は解答欄に選択肢の数字を記入、中学校はマークシートで実施してきた。オンラインで調査する場合、画面上に質問文、あるいは選択肢が表示され、当てはまるところをクリックするという、一般的なウェブ調査の形になるのではないかということイメージしている。例えばアメリカの研究では、解答欄に選択肢の数字を書く解答方式と、選択肢にチェックする、丸をつけるという解答方式では、調査結果が大きく異なるという研究がある。これまでの紙での方式からオンラインに移行することによって、年度間比較、特に時系列的な傾向を見ていこうと思ったときに、そこでの断層が生じないかどうかについて、来年度の試行で、データを活用しながら慎重に検討が必要ではないか。

【委員】 CBT化検討ワーキンググループ中間まとめについては、論点が明確で、期待や課題がよくまとめられていると思った。大きく5点申し上げたい。

1点目、GIGAスクール構想を踏まえての推進について、今後、CBT化に向けて、実施可能な調査方法や実施体制について検討していくというプロセスは、PISA2018の課題解決やGIGAスクール構想推進実現の追い風になるのではないか。平成31年度（令和元年度）の中学校英語「話すこと」調査のPCを用いた調査においては、全国の学校のICT環境の脆弱性などにより大変な苦労や混乱が生じた。次回の調査では、円滑に進められるようにする必要がある。

2点目、調査規模について、CBTを技術的・理論的な論点から突き詰めていくと、恐らく悉皆より抽出の方が親和性が高く、CBT化しやすいという結論に至るのだろうが、私は基本的に全国学力・学習状況調査の悉皆での実施は、ぜひ維持すべきだと思っている。ただ、やり方として、今までのように一斉実施ということにはこだわらなくていいのではないかと考えている。全国学力・学習状況調査というのは、これまで毎年、積極的な改善が加えられてきた。調査開始当初、現場の教師は活用力を問う「B問題」に大変戸惑っていたが、今でこそ当たり前になっている。インプットの改善、求められる学力を現場に浸透させていくという上で、大きな役割を果たしてきた。全国学調の目的は大きく2つあり、一つは、学力の実態を把握・分析して、行政施策を検証していくということ、もう一つは、自治体の教育施策の検証と学校の指導改善につなげるということである。前者を目的とするのであれば、当然、抽出でも可能になるわけだが、後者に重きを置くということになれば、悉皆での実施が求められる。つまり、求められる学力の学校現場への浸透や教師の指導

改善に果たす全国学調の大きな役割というのは、ぜひ今後も維持してほしいと強く思っている。教師の成長、アップデートのためにも重要ではないか。

3点目、学校や教育委員会の負担について、これは十分配慮していかなければならない。先ほど申し上げた昨年度の中学校英語「話すこと」調査、これは初めて学校のPCを利用した音声の録音方式で導入したが、全国の脆弱なICTの状況を鑑みて、CBTとは程遠いスタンドアローン方式、それも現在、学校では使用しないように言われているUSBメモリを使用した。とは言っても、挑戦的な側面もかなり見られ、文科省としても様々なトラブルを予想して、念入りに実施してきたわけだが、それでも想定外の事態が多く発生した。ICT環境が比較的で整備されPCのスペック等も高かった本市でも、ヘッドセットと内臓マイクとのハウリングが発生して、多くの音声データが欠損してしまうという想定外の事態も発生した。今後、PBTからCBTになったとしても、こういった学校や教育委員会等の負担は直ちに軽減するわけではなく、負担の種類や内容が変わるだけ、ということ認識しておく必要がある。

そういったことを鑑みて、大事なことは、小規模からの試行である。一部のCBT、IRT導入、こういったことの諸課題を検証しながら、丁寧に、かつ積極的に進めていってほしい。

4点目、「論点整理」に、地方自治体が行っている学力調査との役割分担について、「地方自治体が行っている学力調査との役割分担等も含め、全国学調の在り方として総合的に検討を行う必要がある」、また「国だけで完結して全国学調の在り方を考えるのではなく」という文言があるが、全国学力・学習状況調査が果たすべき役割や機能について整理することが非常に重要になってくるだろう。これを明らかにした上で、地方自治体の学力調査との役割分担について再確認する必要がある。地方行政のデータの活用というのは、大きな可能性を秘めている。それらは、基本的に悉皆データであり、全ての再配分的施策を実施しているので、様々な領域のデータの結合など、今後、多くの全国展開に向けた政策のトライアルの場にもなるのではないか。そう考えると、国と地方の役割を十分に検討した上で、今までの国から地方へという演繹的な発想だけではなく、地方自治体の実践の蓄積などを国としてより支援すること、さらに、それぞれの立場でステークホルダーが参加しながら、よりよい学力調査に仕立てていくことも必要なのではないか。

最後、5点目、インターネット回線を利用したCBTの調査について、インターネット回線というのは、そもそも相互に接続されているデータ通信網ということであり、ベストエフォートが原則である。これがSINET等とは異なり、音声や動画などの高いリアルタイム性を

要求されるものには不向きであることが明らかになっている。

最近、本市を含め、学校や教育委員会で、積極的にICTを使う状況下で、外部とのビデオ会議等で、映像や音声途切れるという事象が多く起きている。その主な原因は、学校のPC端末が増加し続けて、インターネット回線にはベストエフォートで帯域に余裕がないこと、地域のインターネット利用者が増加することで、プロバイダーに一律帯域を絞られることなどが考えられる。

今後GIGAスクール構想などで、端末の同時利用台数がどんどん増えていくことになると、CBTを実施するには、帯域の使用量の把握と確保や、ネットワークサーバ等の負荷についても、十分考慮していくことが必要になるだろう。

【委員】 小学校現場として最も心配をするのは、児童がコンピュータの扱いに慣れていない状況がかなり見られ、CBTで調査を実施した場合、本来の力を発揮できないケースが出てくるのではないかとということである。現時点でも、ICT環境や取組の自治体間の差はかなり大きなものがあると認識をしている。したがって、大前提として、ICT機器を日常的に使用する授業を実現すること、児童、教員ともにその扱いを習熟していくことが必要である。先進的な学校や自治体での導入はある程度円滑に進んでいくと思うが、現時点で十分に体制が取れていない課題が多い学校、自治体でどういうふうに導入していくか、支援していくかという視点で、さらに検討していただきたい。

CBT化によって、学校現場の負担がどのように変化するのか懸念がある。現行の紙での調査は年数を重ねており、学校もかなり慣れている。CBT化が進む場合、現在の延長線上ということではなく、新しい取組が始まるというイメージも持つ必要があるのではないかと。学校に任されることが増えれば増えるほど、何らかの混乱が生じる可能性は高くなっていくだろう。学校への十分な配慮とサポートが円滑な実施への大きなポイントになってくる。

授業と調査の関係について、調査の結果を日々の指導の改善に役立てるということであれば、日々の授業の内容と調査の内容の関係性が分かりやすく見えている、教員にもよく理解されているということが大切である。実際、授業は授業、調査は調査と乖離しているところもなくはないだろう。CBT化によって、教員のそうした印象がさらに強くなっていくという可能性もある。今年度から新しい学習指導要領による授業が始まっているが、日々の授業と調査の関係をより分かりやすく示していただけるとありがたい。

全国の小学校校長から、悉皆ではなく、抽出の調査で目的を達成することができるので

はないか、6年生の1回の調査をもって、指導方法の改善につなげていくということはなかなか無理があるのではないかという声がある。CBT化に向けて具体的な検討が進んでいく中で、実施方法についてだけではなく、何のために、どのような力をどういう調査で行っていくべきなのかといった、そういう調査の根本を大切にしていきたい。

【委員】 全国学力・学習状況調査が継続的な検証改善サイクルにきちんとつながる形で、CBT化を検討していただきたい。教育に関する継続的な検証改善サイクルを確立するために、国や自治体レベル、それから学校、授業レベルを1つの調査でそれぞれの改善に寄与する国内で唯一の貴重な調査である。多様なレベルにフィードバックするからこそ調査のデザインが難しいが、自分の専門領域である、評価の研究領域においても、国、自治体、学校、それぞれの評価が包括性があると同時に一貫していることが大事であるといわれる。安易にCBTの特性に合わせて、現行の調査を支えている、国と現場の指導改善、両方にフィードバックしつつ改善サイクルを回していくという大きな目的が薄められること、また、特性に合わせて調査方法を変更し、実施していくことで、これまでの調査のよさを潰しかねない。この大きな目的を踏まえて、何をどうCBT化していくと、きちんとよりよく改善サイクルを達成できるようになるのか、整理、位置づけすべきではないか。全国学力・学習状況調査の役割を国レベルに特化する、もしくは自治体や都道府県に任せてしまうと、そこでも分断が起きてしまうのではないか。CBT化の検討を進めていく上で、国、自治体、学校、授業が結びついて、調査結果がフィードバックされている点が失われないような枠組みが大事だと思っている。

作問体制について、IRTの導入や問題バンクの作成にも関わるが、これまで長年の調査による過去問の蓄積というのがある。そのため、今後いろいろCBT化を検討していく上でも、これまでの問題が、どのような領域で、どのような力を測っていった、その成果がどうであったのか、それを整理してまとめていくことが次の問題作成にもつながるのではないか。CBT化の検討や問題バンクの作成だけではなく、国、自治体、学校、授業それぞれに寄与する調査問題、作問としての質を上げていくことにもつながるだろう。調査問題の質の向上については、現在、国研で「学習科学のアプローチによる全国学力・学習状況調査 教科に関する調査の質的向上に関する分析検証委員会」を設けて検討を進めている。その報告書を現在、整理しており、次回以降の専門家会議でも報告させていただきたい。

問題の在り方について、イラストや動画等、CBTの特性を生かした出題が可能になる一方

で、記述式問題の採点も課題があるといった、様々な特性についてまとめられている。また、知識を問う問題を中心にCBT化するという仕組みも含めて検討するなど、様々なアイデアが記載されている。こういう問題の在り方については、慎重に検討していく必要があるだろう。例えば、これまでA問題としていた知識を問う問題の学力が、どれほどB問題、つまり活用に関する学力を支えているのかなど、実際、学習研究者としても、学習論、学力論として、まだまだ分からないところも多い。そう考えると、測定のやりやすさやCBTのなじみやすさだけではなく、学習論、学力論の観点からのテストのデザイン、グランドデザインを描くことが必要ではないか。

論点整理の中には、解答に加えてログも把握することによって、児童生徒のつまずきの把握など、多角的な分析が可能になるというところも利点として書かれているが、こういうログを把握することがきちんとメリットがある問題とはどういったものなのか、問題の中身というのがすごく重要になってくるだろう。これらを踏まえると、体制整備に関する観点から、テスト理論やデータ分析の専門家だけでなく、むしろ学習者、受ける児童生徒の視点に立って、教育と評価を結びつけられるような人材が必要ではないか。また、教科教育とテスト理論との統合、そういうのをリードできる人材がいて、初めてこういう理論と実際に見たいものというものが結びついていくのではないか。

ほかにもIRTに関連して、「大問形式のままプールしていくのか、小問独立形式としてプールするのか」という課題が書かれている。今までの大問形式から小問形式に変えることで、児童生徒に求める資質能力、測りたい力というのを継続的に測ることができるのかどうか、これも実際にデータを集めて検証する必要があるだろう。小問形式の問題を数多く並べて出題したときに、児童生徒がどんな思考過程を發揮して解いているのか、どういう資質能力を使って解いているのか、そういう比較実験等も必要になってくるだろう。こういう観点からも、CBT化の上で必要となる組織の中の人材として、教科教育とテスト理論の統合をリードできるような、学習者の視点に立って、認知過程を分析し、問題の在り方というのを具体的に提案できるような、そういう教育と評価を結びつけられる人材が必要ではないか。

【委員】 調査目的が複数あるために作問が難しいという問題がある。今回、CBT化、IRTを導入するとなれば、この問題はさらに先鋭化していく可能性がある。そういう意味で、CBT、IRT導入ということを生かすのであれば、やはり学力調査として特化していくべきで

はないか。学力調査は、指導改善への活用に比べてほかの手段が限られているため、この目的を優先したい。指導改善への活用という目的を外せば、悉皆調査でなくてもよいだろう。悉皆調査でなければ、問題を集めるメリットがあまり考えられないため、問題漏洩の危険性も少し下がるのではないか。また、たとえば知識を問う問題が中心になっていく形になると、どうしても学習指導要領に定められている学力とずれてくるため、指導改善に活用しにくいという問題がある。そういった観点からも学力調査に特化すべきではないか。

問題が非公表になることで、結果的に1つメリットがあると思っている。今までの調査では、解答できなかった問題について補習を行っていたが、授業改善にはなかなか結びつかなかった。問題単位で対応して終わりという形で使われることが非常に多かった。問題が非公表になり、調査問題がわからない形で結果のフィードバックがある場合、その結果を学校の教育に生かす方法を考える必要がある。そこにちょっと新しい指導改善に関する何か突破口があるのではないかと期待している。

【委員】 「GIGAスクール構想を踏まえ」とあるが、それを土台にした上で学力調査を設計したときに、土台というのは崩れないのかという点が心配である。学力調査に合うようなGIGAスクール構想というのが想定されているのかどうか、CBTに基づく学力調査ができるように、GIGAスクール構想というものが進むのかどうか、懸念がある。

目的について、教育再生実行会議第11次提言では証拠に基づいた政策マネジメントが挙げられており、ここでの学力にはコンピュータリテラシーも含むものとして解釈される。学力調査の内容は、実際に行われる学習指導が反映された内容でないと、証拠に基づく政策の改善には結びつかないだろう。一方で、いわゆるIRTに基づく学力の経年変化といった比較を重視すると、今度は証拠に基づく現場での学習指導には結びつかないというジレンマがある。どちらを優先するのか、何のための調査かというところを決める必要があるだろう。

測定の観点から、CBTでの実施自体に機器の操作スキルと実際に問題を解くという内容の学力とが混在している。これは長崎大学で入学手続き時に提出してもらっている調査の結果である。得意か、苦手か、あるいは経験なしかという5択のうち、「得意」、「どちらかと言えば得意」と回答した学生の割合を足して、数値が高いものから並べられている。そのうち、一番下にあるのが、「コンピュータなどの情報機器を活用する」という項目で、大学入学の手続きの時点でこの結果である。

これらをふまえると、まず、コンピュータリテラシーをそろえてから調査を実施しないと、学力をきっちり測定しているということにならない可能性が非常に高い。例えば、学力調査の前にITスキル検定のようなもので、一定のコンピュータ操作ができることを確保した上で、そして、CBTに基づく調査を実施することにしなければ、学力の調査そのものが成立しないのではないか。そのことは、今度は小学校や中学校の指導を変えることにつながってしまう。学力調査を行うために、コンピュータの操作に特化した何かのトレーニングを入れなければならないことになり、こうなると1つのカリキュラム等を変更せざるを得ない状況をもたらす可能性がある。

CBT化した場合、児童生徒の解答意欲を高めるような出題が可能になるということが書かれていたが、先の問題と同様に、コンピュータの操作を楽しめる児童生徒と、苦手だ、楽しくないという児童生徒がいるため、注意する必要がある。また、エンターテインメント性を持ち込むと、思考力や問題解決力を測定することが、何かゲーム的なものに寄ってしまわないか、という点も危惧している。

IRTの導入について、問題バンクへのプールは小問形式なのか、大問形式なのかという点が記載されているが、項目という概念をどう捉えるのかでパラメーターの解釈が異なる。マウスの動きや反応時間のログを取れるとあるが、実際の児童生徒の動きや認知過程などが分かるような作題をするのか。IRTは局所独立の原則をふまえているが、そういう調査のデザインにすると条件付のモデルになってしまい、非常に複雑なモデルで計算をしていかないといけない。その中で、さらに等質な問題を作り続ける必要が出てくる。

作問を担う人材について、医療系共用試験の事例では、各大学医学部の先生が相当な数の問題を作成している。その体制をこの学力調査に当てはめると、現場の先生が相当な数を作って問題を提供することになる。共用試験の場合は短答式の問題であり、医学部の定期試験でも知識に関する出題が多い。参考にする事例とこの全国学力・学習状況調査は分けて、オリジナルのものを検証しながら作っていく必要がある。限界を見定めて、その上で目的や実施方法を検討する必要がある。

**【委員】** 繰り返しになるが、調査目的を明確にしないと、CBT化検討ワーキンググループの方でも今後の議論が難しくなる。実施方法やそのためにどのような技術が必要なのかといった議論は、前提である調査目的に影響される。

それと、ここまでの議論では、「指導改善のため」「実態把握のため」という論点に加え



て、「メッセージ性」という観点が示されていたように思う。メッセージ性というのは波及効果（Washback effect）のことで、調査をすることで相手に何か影響を与えたいということだと思うが、実態把握とメッセージ性の両立はできないだろう。指導改善への活用と実態把握も根本的には矛盾する目的であるため、まずは調査目的について優先的に決めていく必要がある。目的さえ決まっていれば、技術的なものに関する議論は可能である。先ほど指摘があったIRTだとシータが何を意味しているかよく分からないという問題についても、これ自体は教育心理学などで様々な議論があるので、対応することは不可能ではない。目的に合わせて複数の調査を設計し、それぞれ連携させることは可能だと思うが、現在の全国学力・学習状況調査のように、1つの調査の中ですべての目的を網羅することは難しい。

【委員】 例えば、ログの例など、技術的に何が可能かという観点から目的が決まっていく部分もあると思う。逆からの目的の決め方も考えられるため、何ができて何ができないのかという技術的な側面からも議論していく必要がある。

【委員】 日本の教育全体として、PDCAをやっていくしかないと思っている。その中で、例えば学力調査の目的から指導改善の側面が弱くなった場合、どこか別途、全体として評価していく必要がある。このように全体の設計の中で、日本の教育が、例えば指導要領、それから先生方の努力、PとDがあって、チェックとアクションという、この流れはやっぱり生き生きと展開されないといけない。

実態把握について、児童生徒の実態は当然だが、実は指導の実態も非常に大事である。例えば、A小学校のB先生とC先生の時とで、児童生徒の生き生きさや学力が異なっていた場合、じゃあ、一体そこに何があるのか、どこかで明らかにしていく必要がある。それはPDCAの流れの中で大きい部分だと思っている。

【委員】 理想的なことは色々考えられるが、大体いつ頃までにどういったことをやるのかというプランも必要である。最初から完璧なものではできないので、走りながら、それこそPDCAで直していかなくちゃいけないと思う。例えば4、5年という単位でやるとすると、その中でできることというのも考えないといけないのではないかな。

— 了 —