

全国的な学力調査の CBT 化検討ワーキンググループにおける検討について

1. 基本的観点

以下の観点を踏まえ、CBT 化に関する様々な論点についてどのようなパターンや選択肢等が考えられるか検討。

- (1) CBT の利点を活かすこと
- (2) 全国どこにおいても、安定的かつ継続的に実施できること
- (3) 学校現場の負担をできるだけ軽減すること

2. 主な検討事項例 ※準備・検討期間を要すると思われる事項から、順次検討

- (1) 日々の学習における ICT の活用と全国的な学力調査との関係
- (2) CBT の利点を活かした学力調査の在り方

①実施の仕方

- ・全国一斉実施の場合、一定期間内(複数回)実施の場合の論点

②調査問題

- ・全国同一の問題セットとする場合、複数の問題セットとする場合の論点
- ・CBT における出題方式・調査問題作成のパターンや選択肢について

③項目反応理論 (IRT)

④採点の在り方

- ・選択式・短答式・記述式問題の採点や、フィードバックの即時性に関する論点

⑤特別な配慮が必要な児童生徒への対応

- ・より丁寧な配慮を行うために、どのような方法が可能か

⑥問題作成の体制や工程

- ・CBT 化にあたり、必要となる基本的体制や工程はどのようなものか

(3) CBT 特有の課題・論点【第4回 議題(1)】

①調査資材の印刷、配送・回収、採点、集計、分析など、現在の一連の工程の効率化

- ・各作業工程について、どの程度効率化されるか

(4) 実施体制等

①CBT システムの開発 (業務管理、CBT、採点、集計・分析、問題プールなど)

- ・実施する際に必要となるシステムやパターンはどのようなものがあるか

②実証実験、予備調査、試行等を含めたスケジュールや具体的な進め方

等

＜各論＞ ※準備・検討期間を要すると思われる事項から、順次検討。

2. 主な検討事項例

(1) 日々の学習における ICT の活用と全国的な学力調査との関係

○GIGA スクール構想を踏まえ、全国規模の学力調査を学校 PC 端末を使用して行うためには、全国的に PC 端末やネットワークが整備されることが大前提であり、その上で、児童生徒や教師が日常的に ICT 機器を使用した学習やパソコン操作に慣れることが重要ではないか。

○情報活用能力という新たな能力が学力の把握・測定に影響することとなるが、ICT の整備状況や普段の利用状況が児童生徒によって様々である。こうした点については、全国学力・学状況調査の本格的な CBT 化の前に、試行や予備調査等を通じて、測定したい学力と情報活用能力などの関係について分析することが望ましいのではないか(例えば、記述式問題の取扱いについても、キーボード入力とするのか手書き入力とするのか等の論点があるのではないか)。

(2) CBT の利点を活かした学力調査の在り方

①実施の仕方

・全国一斉実施の場合、一定期間内(複数回)実施の場合の論点

○ 全国一斉実施の場合、前年度までの学習状況について、全国で同じタイミングで評価できるが、調査日の調整(各学校への周知)は、約1年半前から行う必要がある。また、自然災害や感染症等の突発的な事情で調査日に実施できない場合、改めて数ヶ月以内に調査日を調整することが難しく、中止とせざるを得ない。また、調査日に修学旅行等の学校行事や学級・学校閉鎖等で実施できない場合、後日実施期間中に実施することとなるが、集計の対象外となる。

○ 他方、一定期間内(複数回)実施の場合は、各学校において都合の良い調査日を選択することができるようになるが、その場合、調査問題を複数セット用意する必要があるのではないか。

○ 全国学力・学習状況調査をオンラインで CBT 化する場合、参加児童生徒数が約200万人と大規模であり、サーバの負荷はかなり大きく、解答時間中に調査を続けられないなど、トラブルに脆弱となることが想定され、現在のように、同一日時で一斉実施するのではなく、抽出または複数回に分けて実施することが必要ではないか。

②調査問題

- ・全国同一の問題セットとする場合、複数の問題セットとする場合の論点
- ・CBTにおける出題方式・調査問題作成のパターンや選択肢について

(問題セット)

- 現行の全国学力・学習状況調査は、出題される問題数が非常に限られている。複数の問題セットを用意することにより、幅広い領域の問題が出題されるというのは、教育委員会から見た時に、自治体内の全体的な児童・生徒の傾向について得られる知見が多く、指導改善への活用という点から大きなメリットである。
- (全6時間で320問を出題している医療系大学間共用試験と異なり) 全国学力・学習状況調査は、一人の児童生徒が多くの問題を受けるということは想定できない。幅広い領域・内容の学力について情報を取得するためには複数問題セットを用意して、全児童生徒で全領域を手分けして受けてもらうような形になるのではないか。
- 複数問題セットが用意できれば、複数回調査を実施することが可能となり、社会状況と学力の関連について、各自治体の状況に合わせた日程で柔軟に調査を実施できるようになるのではないか。

(出題方式)

- TIMSS(国際数学・理科教育動向調査)における PSI¹(問題解決と探究)のように、CBT 特有の新しい問題が導入できることとなるが、調査問題の開発に関しては、どのような問題でどのような能力を測るのか、研究が必要ではないか。
- CBT では紙による筆記方式ではできなかった様々な出題が可能になるが、作問にあたっては、現行のように教科教育の専門家だけではなく、CBT の出題方法や解答形式などに詳しい専門人材が必要になり、作問体制を抜本的に変える必要があるのではないか。
- 医療系大学間共用試験では、コンピュータの機能を活用し、医療面接から始まる一連の思考過程を問う(後戻りできない) 順次解答型問題を用いて、実際の臨床の流れに即した問題を出題している。また、マルチメディア対応問題として、例えば、心雑音の音声を出して、どのような症状か判断させたり、写真をクリックして動画で患者の状態を見て、意識レベルを判断させるなど、より一層の臨床の現場に即した問題の出題を検討している。
- CBT 化によって、例えば算数・数学の作図など一部の分野や、国語の縦書きで出題されてきた問題文をどうするかなど、出題方法の工夫が必要になるのではないか。

¹ PSI: Problem Solving and Inquiry (問題解決と探究) : TIMSS2019 から加わった新類型の問題で、研究室での実験や日常生活の文脈を模した問題を出題。児童生徒の回答に応じて、次の問が変化するという従来の筆記型調査ではできない出題方式となっている。また、問題はカラフルで動画を含んだものとなっている。

- 全国学力・学習状況調査は、これまで、主として、知識・技能を問う A 問題、活用を問う B 問題があり、平成31年度調査で、知識・技能や活用を一体的に問う形式(いわゆる「AB 一体化」)となったが、CBT 化に伴い、例えば知識・技能を問う問題を CBT 化するという仕組みも考えられるのではないか。
- 現行の全国学力・学習状況調査の国語では、大問の中に、いくつかの問が設けられている形で出題されている。問題をプールしていく際、現行の調査問題がそのままプールされるのか、あるいは特定の資質・能力に焦点化し、小さな分量の問をプールしていくのか、問題の選別方法と各問題の資質・能力に関する等化の方法について検討する必要があるのではないか。

(解答方式)

- 記述式問題については、キーボード入力とするのか手書き入力とするのか検討する必要がある。キーボードによる解答は、一部の児童生徒には負担があり困難ではないか。例えば、記述式問題の解答は紙による筆記方式とするなど、対応を検討する必要がある。

※引き続き、ご議論いただきたい論点

- ・児童・生徒の学力を測定する際の精度や領域別学力のフィードバックのあり方
- ・問題数、作問体制、作問スケジュール等への効果や影響
- ・問題の公開/非公開及びそれに伴う教育現場へのメッセージ性や指導の在り方への効果や影響
- ・年度間比較の必要性について
- ・CBT における出題形式(分冊方式、LOFT 方式(※)、適応型とマルチステージ方式等)

(※)LOFT:Linear On-the-Fly Testing 方式:出題領域や問題の難易度、解答時間などが同じようになるように問題セットを自動編集して配布する方式。

③項目反応理論 (IRT) ²

- ・項目反応理論を導入する場合の利点や課題

(IRT の導入)

- 現行の全国学力・学習状況調査は、全国の児童生徒が同じ問題を解くことにより、その結果から傾向を分析するなどして、教育現場の指導改善に活かしている。また、配点や部分点がなく、どの問題も、正答・誤答のいずれかとなり、調査結果は、正答数や正答率で表している。また、調査問題は、毎年度、限られた問題数の中で、身につけておいて欲しい事項を問うており、年度間で難易度を調整しているわけではな

² 項目反応理論(IRT: Item Response Theory):問題の特性(難易度や識別力等)と各児童生徒の学力を分けて考える枠組みであり、例えば、児童生徒の正答・誤答が、問題の難易度(簡単・難しい)によるのか、問題の質によるのか(測りたい能力を正確に測れる問題か)、児童生徒の学力によるのか等を区別して考えることができる。これにより、いくつかの手続きを経ることで、異なる時点、冊子、児童生徒集団等で実施した場合であっても、その結果を相互に比較できるようになる。

いため、基本的には、年度間の比較はできない仕組みとなっている。

- IRTを導入する場合、調査前に各問題の難易度や識別力を把握することが可能なため、一斉実施ではなく、難易度を揃えた複数の問題セットによる実施や、年度間の結果の比較が可能となる。一方で、問題を継続的に使用する必要があるため、問題は原則として非公開とすること、また、作成する問題数を増やすこと、調査前に問題の特性(難易度や識別力等)を確認する手順をとること、問題を長期間にわたって管理することなどが求められるのではないかな。
- 指導改善に活かすのであれば、問題公開と結果のフィードバックは重要である。IRTを導入する場合も、調査問題は一律非公開とするのではなく、何らかの工夫が必要ではないかな。
- IRTを導入する場合、適切な項目特性値を持ったプール問題を非公表で蓄積する必要がある。問題漏洩については、医療系大学間共用試験で行っているように、正答率などの項目特性値を確認し、大幅に数値が変化した問題は削除するなど、対策を検討する必要がある。
- 医療系大学間共用試験では、異なるテスト(異なる日程で、異なる問題を用いて、異なる場所で受験する)を受ける者の能力を高い信頼性で測定する、という要件を設定し、それらに対応するために異なるテストの成績を同一の尺度(能力値)で評価できるIRTが採用された。現状の全国学力・学習状況調査では、現状1つの問題セットを全国一斉で実施しているため、そのままの要件であればIRTの導入は必須ではないが、基準集団を定めた年度間比較を行うのであれば、IRTが必要となるのではないかな。

※引き続き、ご議論いただきたい論点

- ・IRTを導入する場合、現行の調査工程で変わる点は何かな(作問や調査問題の公表など)。
- ・IRTを導入する場合、測定機能の強化、指導改善、EBPM推進の観点からどのようなメリットがあるかな。
- ・IRTを導入しない場合、例えば、年度間比較や現行より正確な学力測定をしたり、一定期間内に複数回実施したりするために、どのような方策があるのかな。その場合、IRTと比較して、どのようなメリット・デメリットがあるかな。

④採点の在り方

- ・選択式・短答式・記述式問題の採点や、フィードバックの即時性に関する論点

(採点)

- 選択式、短答式問題は自動採点や即時の結果のフィードバックが可能となるが、記述式問題は、技術的、予算的に自動採点が可能かどうか、検討を要するのではないかな。

(結果提供)

- 仮に問題が非公開の場合、児童生徒が問題そのものを見ることなく、数値だけをもって自分の学習のどこを、どのように変えていく必要があるかを理解することは困難であり、大学生を対象としている医療系大学間共用試験と異なる点として、留意する必要があるのではないか。
- 具体的な問題を示し、その回答に対するフィードバックを行うやり方も当然あるが、その問題がどの領域の、どういった能力を評価するものなのかが明確になっていれば、個々の問題内容に具体的に触れなくとも、どのような能力に課題があるかフィードバックすることは可能ではないか。

(解答データの取り扱い)

- CBT では、解答に加えてログ(児童生徒が解答までに要した時間やクリック等の動きなどの詳細なデータ)も把握することが可能となるが、それらの分析にあたっては、データの前処理や解答時刻の分析などを行うことができる専門的な技術を持った人材が不可欠となる。
- 解答データについては様々な利用・分析が可能になると考えられるが、解答データの取扱いやセキュリティ等についても十分に検討しておく必要がある。

(医療系大学間共用試験の事例)

- CBT・IRT で実施している医療系大学間共用試験では、モデル・コア・カリキュラムに準拠した出題を行っており、その領域別にレーダーチャートを示して、成績を示している。また、基準集団を設定し、IRT 標準スコアに基づく学内での位置、全国での位置がどれくらいかを示している。さらに、各大学で統計処理が可能な個別データ、全国集計もフィードバックすることにより、課題のある分野を把握し、指導に役立てられるような対応をしている。

⑤特別な配慮が必要な児童生徒への対応

- ・より丁寧な配慮を行うために、どのような方法が可能か
- 現在は、調査日に登校している児童生徒しか調査を受けることができないが、CBT 化により、家庭の端末も使用できる場合に、例えば、不登校の児童生徒も調査に参加できるようになり、調査を通じた学力・学習状況の把握、指導改善に活かすことが可能となる。

※引き続き、ご議論いただきたい論点

現行の全国学力・学習状況調査は、点時版、拡大文字版、ルビ振り版の3種類の問題を用意し、実施において、別室実施や時間延長などの配慮を行っているが、CBT 化によって、新たにどのようなことが可能となるか。

⑥問題作成の体制や工程

・CBT化にあたり、必要となる基本的体制や工程はどのようなものか

(問題作成の体制・工程)

- 全国学力・学習状況調査の問題は、現在、国立教育政策研究所に作問委員会を設置し、約1年かけて、学習指導要領の理念・目標・内容等に基づき学習指導上特に重視される点や身に着けるべき力を具体的に示すメッセージとなるよう専門家による検討を重ね、入念に作成されている。仮に医療系大学間共用試験のような規模の問題バンク³を構築する場合、現在の作問体制は見直す必要がある。
- 医療系大学間共用試験では、各大学から広く様々な問題を集め、精査、吟味していくというプロセスで作問を行い、問題バンクを構築している。教師は、調査問題を作成していく過程で指導力を高めていく面があり、全国学力・学習状況調査の作問体制の構築にあたって、こうした仕組みを取り入れていく必要があるのではないか。
- 各自治体で行っている学力調査の問題傾向を見ると、全国学力・学習状況調査の問題を作問の基準として様々な工夫をして出題しているものが多い。全国学力・学習状況調査については、各学校や教委等が作成した問題を収集して精査していくという作問体制は取りにくいのではないか。
- 現在、全国学力・学習状況調査の補完調査として実施している経年比較分析調査は、問題は非公開で実施されており、その問題を一部組み合わせ出題していくことが考えられるのではないか。

(医療系大学間共用試験の事例)

- CBTで実施している医療系大学間共用試験は良質な問題の作成のために全医学部に問題の作成依頼をしている。問題作成の標準化のために作問システム・標準化テンプレート・作問マニュアルを提供するとともに、説明会の開催や、各大学から3名の委員を推薦し、委員会を編成し、問題のブラッシュアップを行っている。また、質保証のために、プール問題を管理する専門委員会を設置することに加え、CBT、教育測定学、統計学、テスト理論の専門家による定期的な検討を行うことで、問題の質を担保している。

(3) CBT 特有の課題・論点

①調査資料の印刷、配送・回収、採点、集計、分析など、現在の一連の工程の効率化【第4回】

・各作業工程について、どの程度効率化されるか

※今回、ご議論いただきたい論点

- ・現行の全国学力・学習状況調査は、調査資料の印刷、配送・回収、採点、集計、分析などを委託事業者が行っているが、CBT化した場合、各作業の現在の工程と比較し、効率化される部分・内容は、具体的にどの

³ 医療系大学間共用試験では、項目特性値を持った問題を約27,000題（2018年時点）プールしている。

ようなことか。

・CBT化により、新たに必要となる作業・工程等はどのようなものがあるか。

(4) 実施体制等

① CBT システムの開発 (業務管理、CBT、採点、集計・分析、問題プールなど)

・実施する際に必要となるシステムやパターンはどのようなものがあるか

(システム設計や開発)

- システム設計や開発は、限られた予算の範囲内でどの程度の技術水準を要求するかを明確にする必要がある。例えば、実施方式に関するシステム、問題作成や管理に関するシステム、解答データの回収、採点、集計などに関するシステムなど、様々なシステムの技術開発を同時並行で行おうとすると、予算を案分することになるため、優先順位を定めて設計・開発する必要があるのではないか。
- 自治体や学校によって整備されている端末が異なるので、どのような OS やブラウザでも同じような画面表示になるように設計する必要があるのではないか。
- システム開発には、児童生徒、教師、保護者などのユーザーがどのような要望を持っているかを把握した上で要件定義することが望ましいのではないか。

(実施方式)

- CBT の実施方式として、主に①USB 方式、②中間サーバ+LAN 接続方式、③WAN 接続方式があるが、全国学力・学習状況調査を CBT 化する場合、②の中間サーバ+LAN 接続方式 (児童・生徒の受験端末は、調査実施中は校内 LAN にのみ接続し、外部ネットワークには接続しない。解答は一時的に中間サーバに送信しておき、解答終了後に時間差でメインサーバに送信する方式) が現実的ではないか。①USB 方式の場合、大量の USB が必要となるため経費がかかり、取扱いにも手間がかかること、③WAN 方式の場合、国がサーバを用意する必要があり多額の経費がかかることが主な課題ではないか。
- 全国学力・学習状況調査を CBT で実施する場合、監督機能を持たせることが重要である。例えば、問題セットが正しく配信されたことを確認する機能や、解答が確実に送信されたことを確認する機能を備える必要がある。
- 医療系大学間共用試験では、問題を管理し、試験の実施・評価を担う機構⁴にセンターサーバがあり、各大学にサイトサーバと試験端末が置かれる構成となっている。CBT 導入時点で、実施大学のネットワークインフラが十分ではなかったこと、セキュリティ確保の観点から、センターサーバとサイトサーバの間は現時点でもオフラインとなっている。試験実施情報や受験者名簿、問題集、体験テスト、実施マニュアルなどを含む「実施キット」を CD 等で大学に送付し、実施後に解答データを

⁴ 公益社団法人医療系大学間共用試験実施評価機構

CDで機構に送付する運用をとっている。

- 現行調査では、毎年度入札で委託事業者を決定しているが、CBTシステムに関しては継続的なシステム運用が不可欠であり、国と事業者の役割分担等について検討すべきではないか。

(医療系大学間共用試験の事例)

- 医療系大学間共用試験の場合、機構が、試験問題の作成から試験の実施、改善までを組織的に行い、試験を実施する大学は、ICT環境整備と試験実施時の体制を整備している。また、試験当日は機構から試験監督者を派遣するとともに、機構内に実施本部を開設し、試験の実施を監督している。コンピュータシステム上のトラブルについてのみ、システムサポートセンターを開設し、業者に委託して対応している。

※引き続き、ご議論いただきたい論点

- ・ 児童生徒の端末上での操作について、画面の操作方法や、記述式の場合、キーボード入力または手書き入力とするか、調査中の誤操作や、漢字変換機能の制御等の在り方について

②実証実験、予備調査、試行等を含めたスケジュールや具体的な進め方

(実証実験・準備の必要性等)

- 全国学力・学習状況調査をCBT化するという事は、現行、小学6年及び中学3年の約200万人規模の調査であるため、大規模なシステムを開発・整備することとなり、相応の費用がかかる。また、準備期間について、TIMSS(国際数学・理科教育動向調査)は、2017年以前から準備を始め、2023年に完全実施というスケジュールで進めていることを踏まえると、相応の時間が必要となることが見込まれる。CBTシステムの構築に向けて、まず小規模の実験をスタートさせ、検討を進めることが必要である。
- 全国学力・学習状況調査のような大規模調査をCBT化する場合、当初から設計しておくべき点と実証研究を通して改善できる点を区別して進めることが重要である。後者に関しては、小規模な予備調査や実証研究などを行い、その過程で得られた課題から大規模調査で実現可能な方法を検討していくという流れになるのではないか。
- 全国学力・学習状況調査のCBT化は、問題作成・実施方式・結果のフィードバックなど大きな変革となる一大プロジェクトである。着実に実施するためには、小規模の実験から実績を積み重ねていくことが重要である。

(医療系大学間共用試験の事例)

- 医療系大学間共用試験の場合、CBT化にあたり、①試験システムの根幹(IRT採用、非同時実施、セキュリティなど)、②学生の適切な能力評価のための必要な配慮(問題形式、マルチメディア対応等)、③フィードバック(対受験生、実施大学、問題作成等を行う者、行政)、④フィージビリティ(大学におけるICT環境、ICT支援員の必要性、コンピュータ設置スペース等)の観点から検討を

行った。その他、運用手順、障害のある学生への対応、運用マニュアル、手引き等についても検討を行った。

- 医療系大学間共用試験では、正式実施に至るまで、受験生 100 人以上（1 試験会場分）を対象にして、試行を「トライアル0」～「トライアル4（最終トライアル）」と称して、4年間かけて段階的に実施している。トライアルは全国 80 医学部（当時）が参加して行われている。「トライアル0」として、CBT 導入可能性の検討のためのプロトタイプシステムを開発し、「トライアル1」として、100人以上を対象にし、改めて問題作成システムの開発、出題管理システムの開発等を行い、「トライアル2」として、全受験生に同じ問題を出すアンカー問題を20%出題しIRTの精緻化、追・再試験の実施機能の追加等の対応を行い、「トライアル3」として、視覚素材の出典管理機能や問題ブラッシュアップシステムの構築等を行い、「トライアル4（最終トライアル）」として、正式実施を想定した運用のチェック等を行った。これらのトライアルの実施を通じて、コンピュータの障害内容を分析し、どのような機能が必要か検討し、サーバがダウンしたときの端末の動作、サーバが故障しても、端末だけで試験が継続できる仕組みの構築等の対応、学生の能力を適切に評価するための問題形式の開発・確認を行った。

※引き続きご議論いただきたい論点

- ・実証研究の実施に向けて留意すべき点、初期段階の実証研究で確認すべき事項について
- ・今後、予備調査、試行等の規模や実施にあたっての留意点等について
- ・PBT から CBT への具体的な移行方法について（移行方法、PBT と比較した場合の費用について）