

IRT導入のメリット

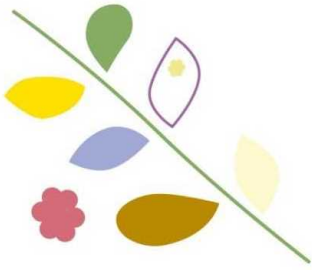
第9回CBT化検討WG
別府先生説明資料より
抜粋

出題する問題が異なっても、児童・生徒の得点を比較できる。

- 複数種類の問題セットの難度を、ほぼ同じ程度に揃えることが可能
- 複数種類の問題セットを使い、複数日程で調査を実施することが可能
- 問題セットの種類を増やし、教育課程をより広くカバーすることが可能
- 学力の変化を時系列的に追跡することが可能

⇐ 共通尺度
⇐ 等化

※以下、IRTを導入した（する）調査のことを、「IRT調査」と呼びます。



IRT調査の業務の流れ

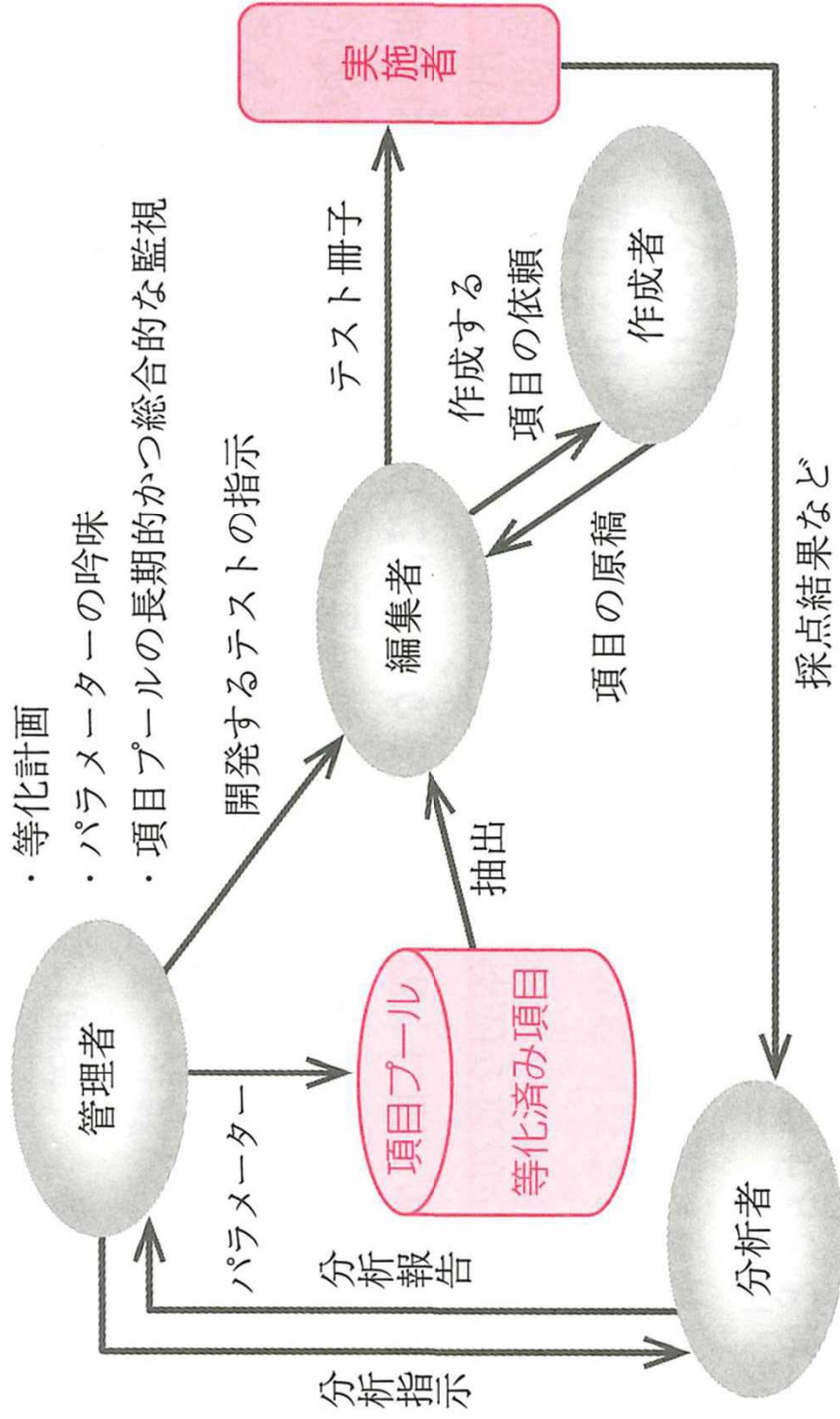
- 1) 調査を実施する目的の明確化
 - 2) 目的を達成するための計画立案*と体制整備
 - 3) 問題セットの作成
 - 4) 調査実施
 - 5) 分析・等化
 - 6) 結果の活用
 - 7) 検証
- ⇨ 3) ~ 5) を「調査の実務」と呼びます

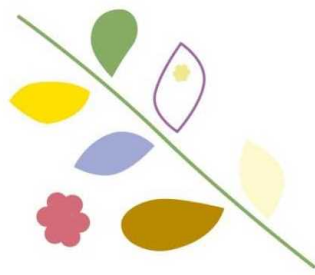
*：調査の出題計画、等化計画、実施計画、分析計画、活用計画、収支計画など

IRTを導入する場合、3) 問題セットの作成は、4) 調査実施や、5) 調査の分析・等化と深く関係しますので、ここでは、3) ~ 5) をまとめて、「調査の実務」とし、以下、必要な役割や体制について取りあげます。



調査・テストに必要な役割





各役割に求められること

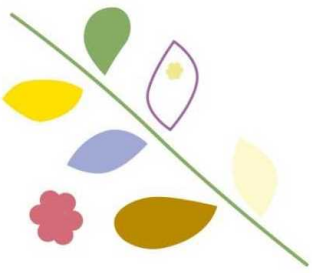
| 役割 | 技術 | | | 知見 | | | 資質 | | |
|-----|----|----|----|------|------|------|--------------|--------------|--------------------|
| | 分析 | 編集 | 作問 | 測定理論 | 教科知識 | 教育情勢 | 他 | 共通 | 個別 |
| 管理者 | ○ | △ | △ | ○ | ○ | ◎ | 教育に関する総合的な知見 | ・判断力 ・計画性 | |
| 編集者 | ○ | ◎ | △ | △ | ○ | ○ | | 適度な柔軟性 | ・バランス感覚 ・調整力 |
| 作成者 | — | — | ◎ | △ | ◎ | ○ | | コミュニケーション力 | ・協調性 ・教科内容への考察力 |
| 分析者 | ◎ | — | — | ◎ | △ | ○ | | | ・探究心 ・現実感覚 |

補足：記号は、技術や知見のそれぞれにおいて、各役割にどの程度求められるかの目安。

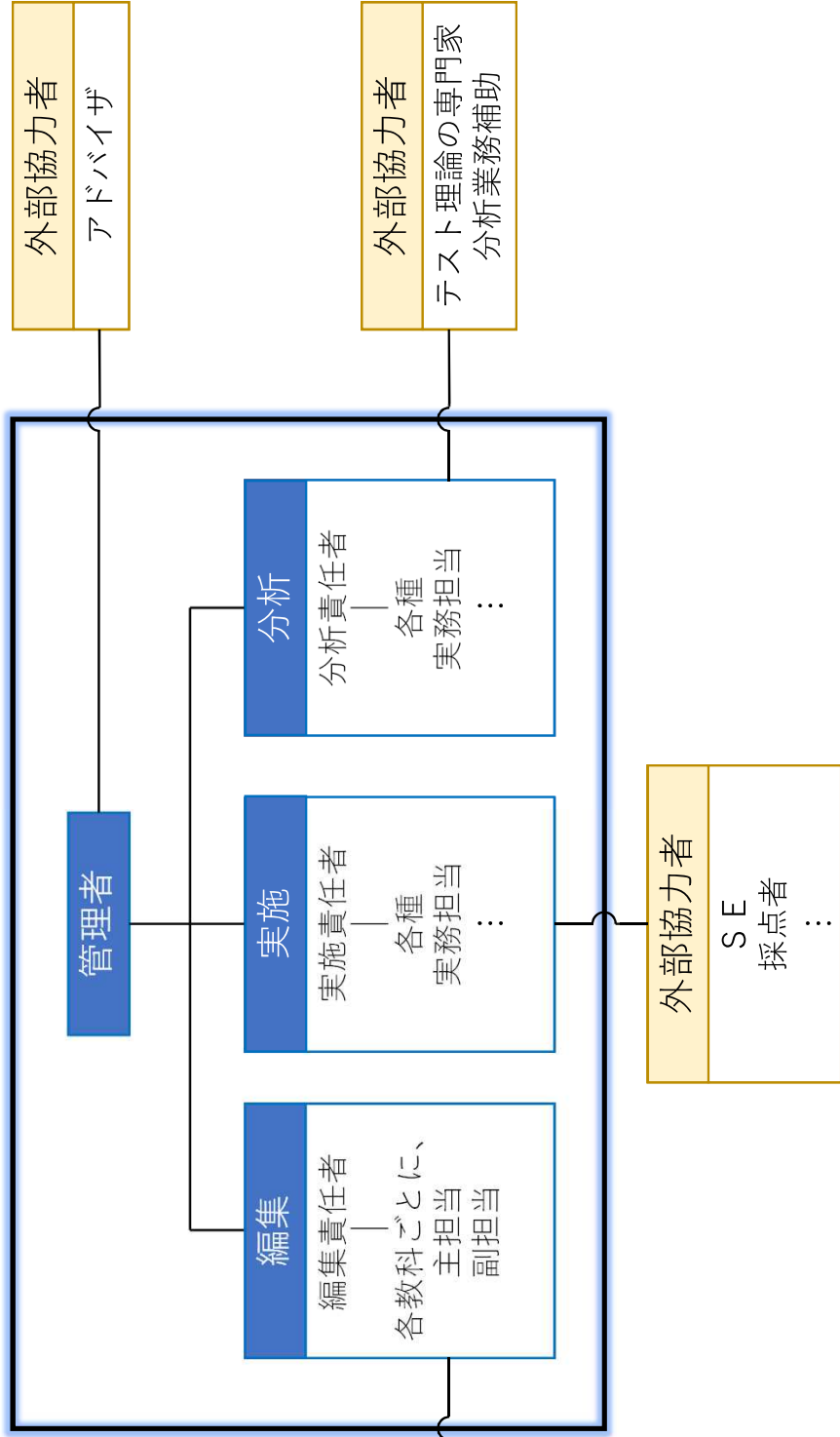
◎ 強く求められる ○ 求められる

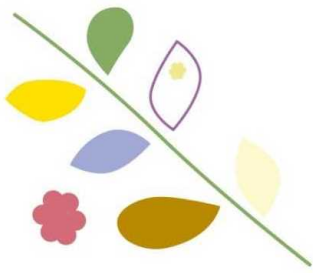
△ 最低限求められる — なくとも良い



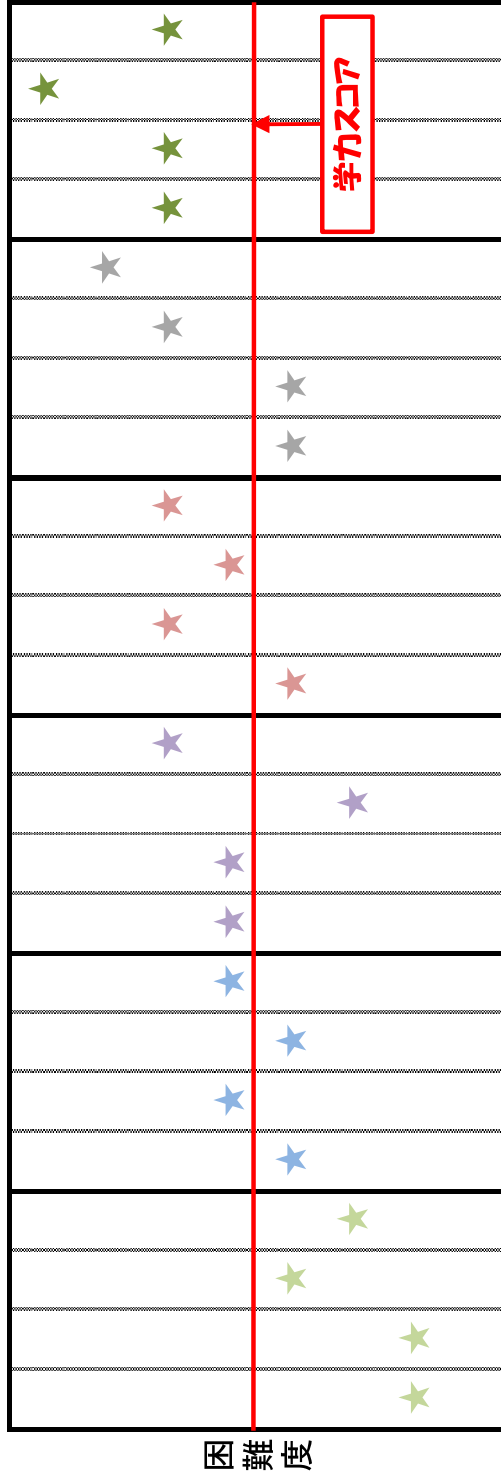


調査の実務に必要な体制





目々の学習での活用例



| | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| ○ | ○ | ○ | × | ○ | × | ○ | × | ○ | ○ | × | × | ○ | × | ○ | × | ○ | ○ | ○ | ○ | ○ | ○ | × | ○ |

| | | | | | | | | | | | | | | | | | | | | | | | |
|-----------------------------------|--|--|--|-----|--|--|--|-----|--|--|--|-----|--|--|--|-----|--|--|--|-----|--|--|--|
| 分野1 | | | | 分野2 | | | | 分野3 | | | | 分野4 | | | | 分野5 | | | | 分野6 | | | |
| × 現在の学力で解けるはずなのに間違えた問題 ⇒ この問題から復習 | | | | | | | | | | | | | | | | | | | | | | | |
| ○ 現在の学力以上の力が発揮できた問題 | | | | | | | | | | | | | | | | | | | | | | | |



※問題の困難度に関する情報と児童生徒の回答結果から得られる学力スコアを照らして、より効果的な復習や宿題を項目プールから提示することが可能になる

テスト理論は測定技術でもある

例：Test Fairness (公平性)と Test Equating(等化) の出現頻度の変遷(1950-2008)



◆テストの目的◆

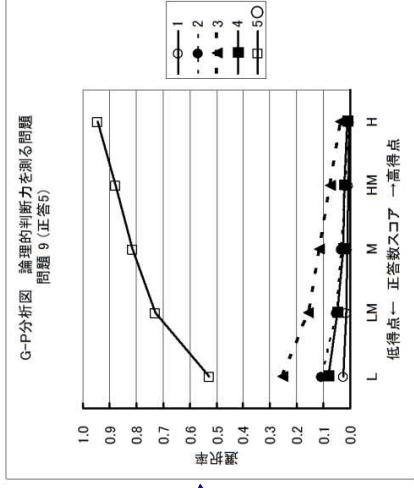
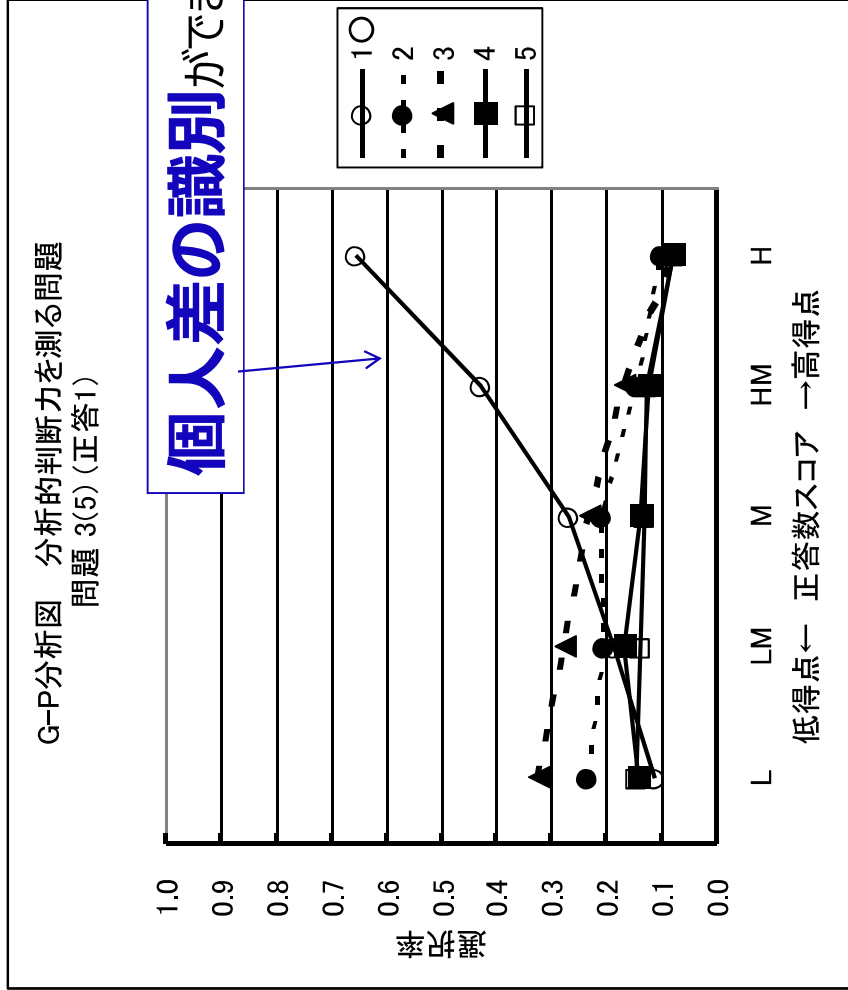
によって技術の使い方が異なる

- 1) 個人の処遇：個人スコア
 (◆**総合的評価**◆のための測定)
 ・選抜(対訴訟：極めて厳格な測定)
 ・医療系大学間共用試験CBT
 ・**全国学調_本体調査**
- 2) 集団の実態：集団スコア
 (◆**EBPM**◆のための測定)
 ・**全国学調_経年変化分析調査**
 ・PISA・TIMSS・NAEP等
- 3) 個人の進捗：個人スコア
 (◆**「学び」**◆のための測定)
 ・GIGA個別最適な学習
 ・形成的アセスメント
- 4) 個人の成長：個人スコア
 (◆**追跡**◆のための測定)
 ・学力発達のサポート
 ・埼玉県学力調査

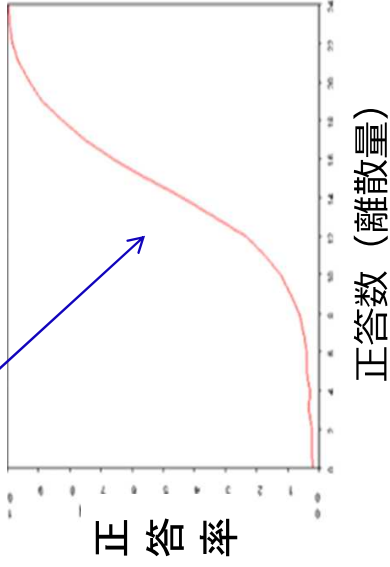
選択肢から見た良い項目

(法科大学院統一適性試験：日弁連法務研究財団より掲載許諾済)

https://www.jlf.or.jp/jlsat/touitsu_kakokekka/



良い問題の項目特性曲線：DNC数学Ⅰ (柴山が復元)



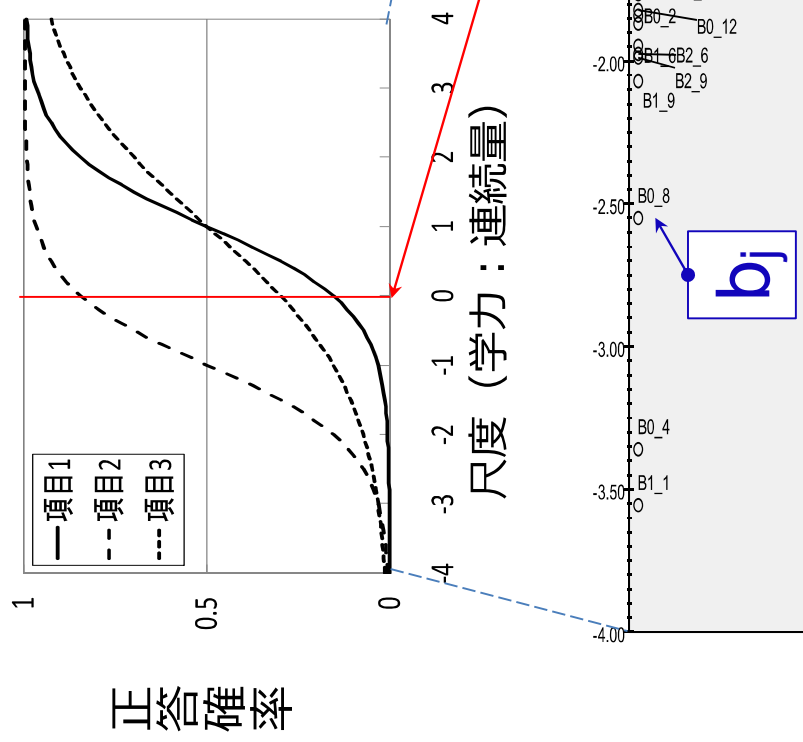
IRTモデルは統計モデル

思考過程・認知過程を記述するプロセスモデルではない

2-Parameter-Logistic model:2PLモデル

$$P(X_j = 1|\theta) = \frac{1}{1 + \exp\{-1.7a_j(\theta - b_j)\}}$$

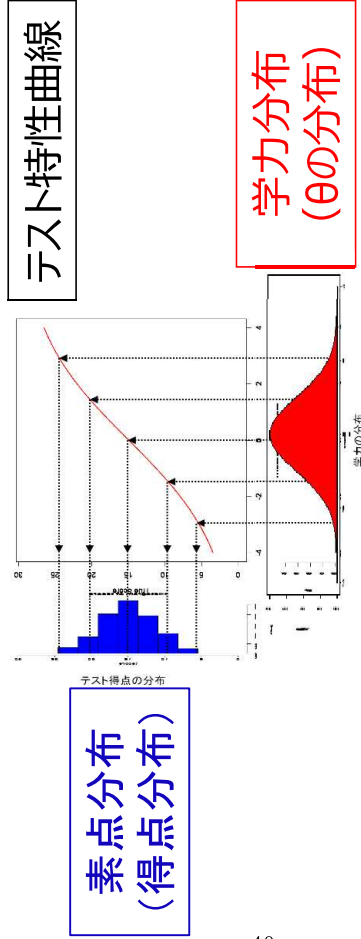
- 1) 学力 θ と項目困難度 b_j を分離
- 2) θ と b_j を同じ数直線 (尺度) 上で表現
- 3) b_j の組み合わせと正誤情報から θ を推定



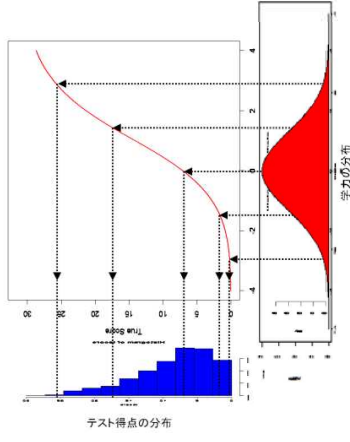
IRTモデルによるテストの得点分布の予測

平成22年度文部科学省委託調査研究報告書p.5
https://www.mext.go.jp/b_menu/shingi/chousa/shotou/085/shiryo/attach/1312362.htm

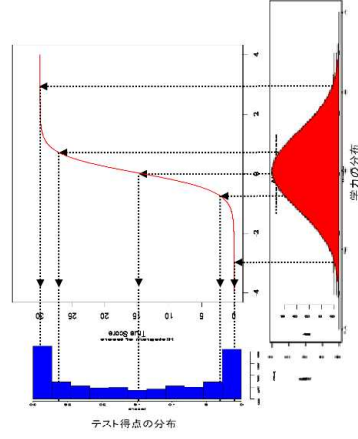
1) 一般的なテスト



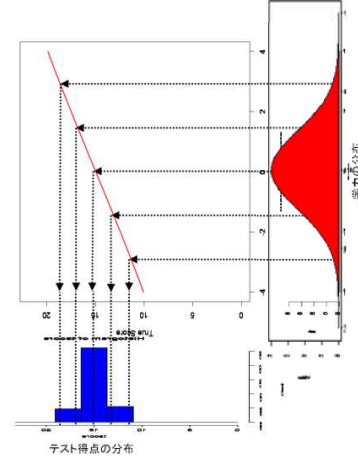
2) 高い学力層を選抜する場合



3) 資格の認定に使う場合



4) 個人差を小さく見せる場合

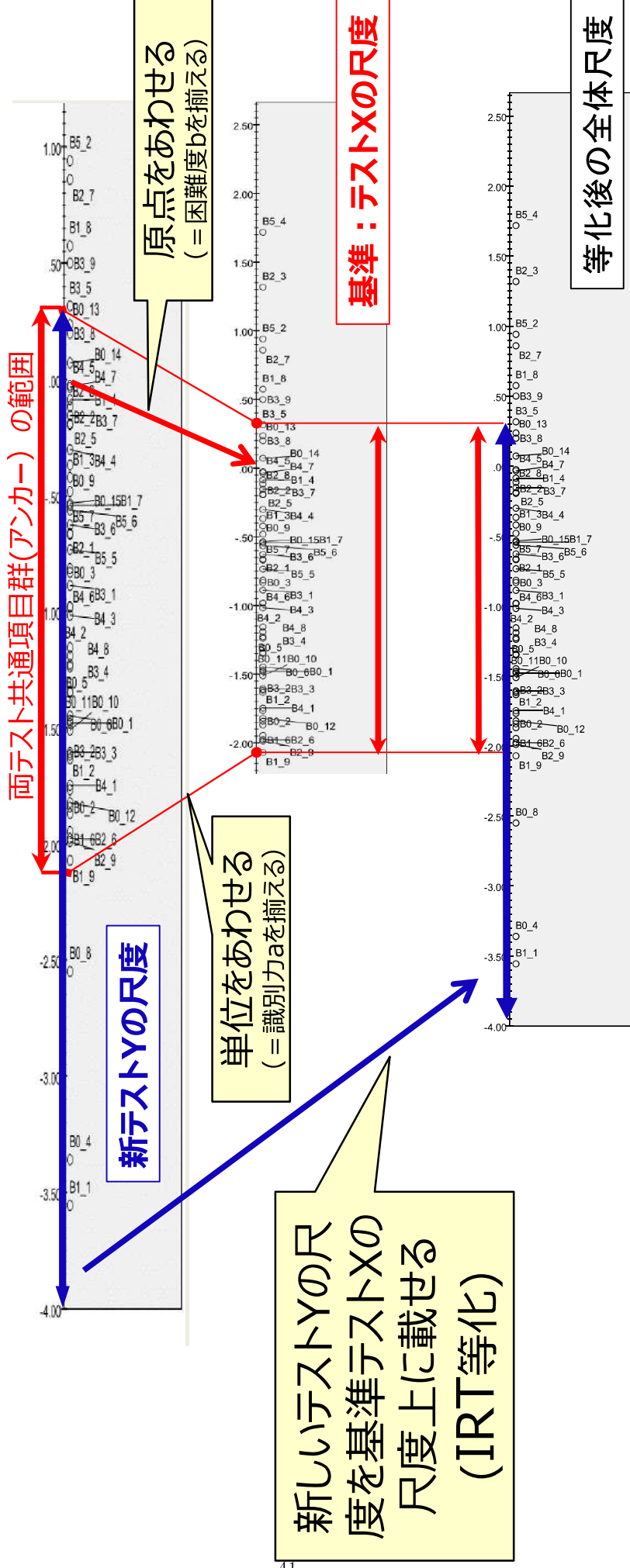


重要:

学力分布は同じでも,
 テストの特性が変われば,
 素点分布 / 得点分布の形
 は変化する

IRTによる尺度等化 (IRT等化) の基本

線形変換により原点と単位を合わせる



新テストYの尺度を基準テストXの尺度へ等化し構成した全体尺度

IRT等化のためのデータ収集デザイン

(等化デザイン：受検者または項目/問題に**共通する部分**がある)

- 等価グループデザイン

| グループ | テスト X | テスト Y |
|----------------|-------|-------|
| P ₁ | ○ | |
| P ₂ | | ○ |

- 単一グループデザイン

| グループ | テスト X | テスト Y |
|------|-------|-------|
| P | ○ | ○ |

- カウンターバランス デザイン

| グループ | テスト X | テスト Y |
|----------------|-------|-------|
| P ₁ | 1 | 2 |
| P ₂ | 2 | 1 |

- アンカーテストを伴う不等価グループデザイン (NEAT)

| グループ | テスト X | テスト A | テスト Y |
|------|-------|-------|-------|
| P | ○ | ● | |
| Q | | ● | ○ |

※P₁, P₂は同じ母集団Pからの異なる標本を、また、P₁Qはそれぞれ互いに異なる母集団からの標本（児童・生徒）を表す。
※カウンターバランスデザインの表中の数字は実施順を示す。

CBT実施方式のパターン分類

第2回CBT化検討WG
寺尾委員説明資料より
抜粋

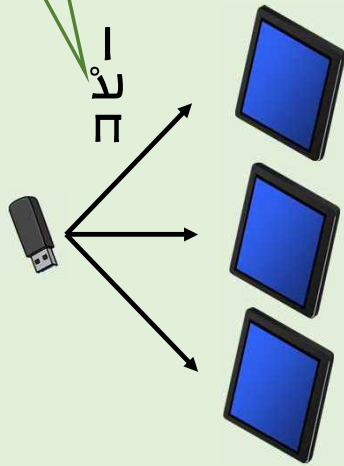
- **USB方式**
 - USBに調査アプリを入れておき，ネットワークを利用せずに実施。
(CD-Rを利用する方法もこれに分類される)
- **中間サーバ+LAN接続方式**
 - 児童・生徒の受験端末は，調査実施中は校内LANだけに接続し，外部ネットワークには接続しない。
 - 解答は一時的に中間サーバに送信しておき，解答終了後に時間差でメインサーバに送信する。
- **WAN接続方式**
 - 直接メインサーバに接続する。
(モバイルルーターを利用する方法もこれに分類される)

USB方式

・解答前

パターン1

各小・中学校

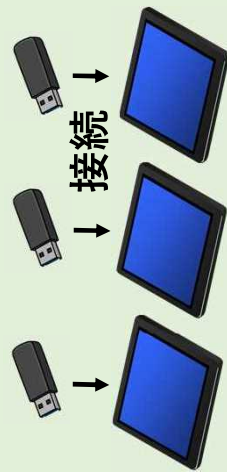


現場教員による作業

コピー

パターン2

各小・中学校



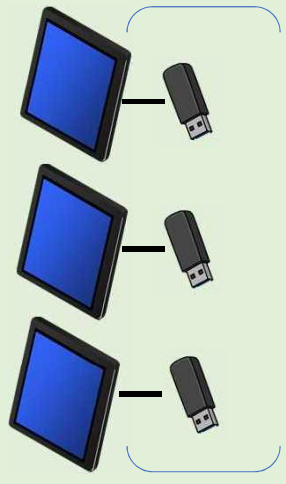
接続

・解答中



現場教員による作業

各小・中学校



・解答後

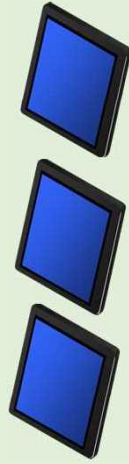
パターン1

各小・中学校

現場教員による作業



データ抽出



パターン2

各小・中学校



回収

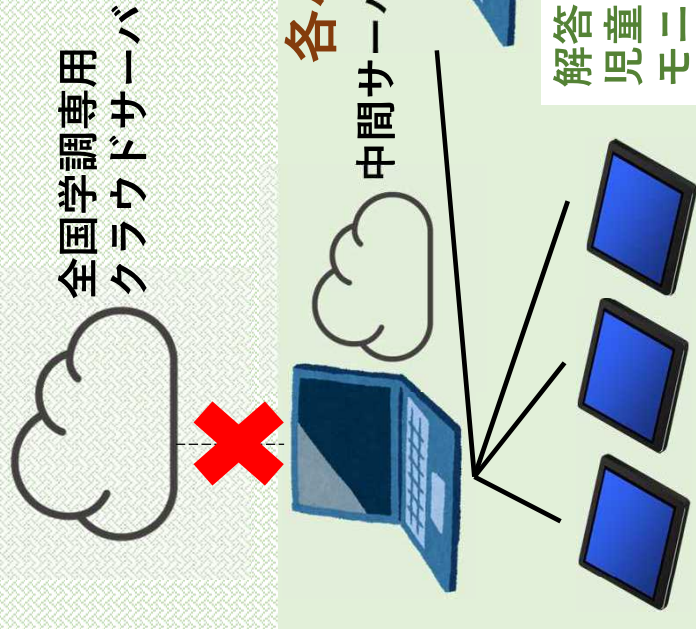


USB方式

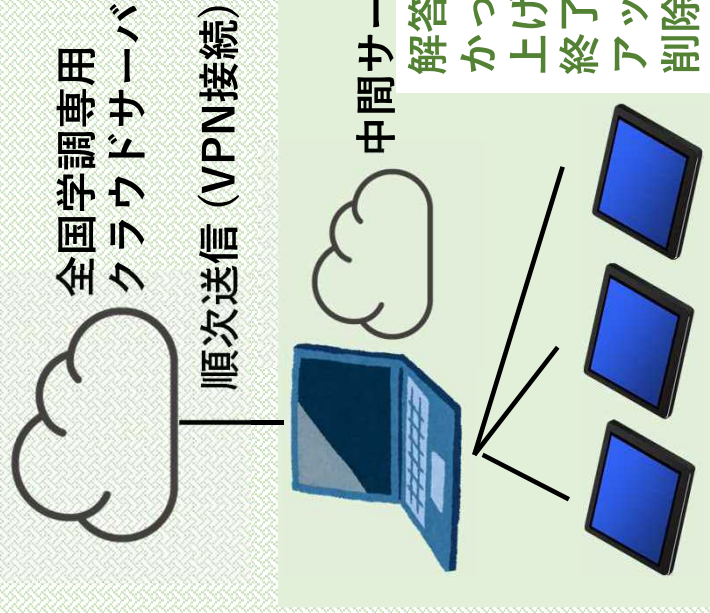
- **メリット**
 - 解答前～解答後まで一貫して、**ネットワークを用いない。**
⇒ ネットワーク関連のリスクや不安定さを解消できる。
- **デメリット (検討課題)**
 - **現場の先生方の負担が多大で、人海戦術 (特にパターン1)。**
 - 学校のPC管理ポリシーによっては、**USBポートを使用できない、環境復元システムなどの設定を解除をする手間がかかるなど、個別対応が求められる。**
 - 児童・生徒の中には**USBメモリの正しい抜き取り手順を守らずに抜き取る子もあり、データ損失のリスクが大きい。**
 - USBの中に入れた問題／問題セットしか出せない。

中間サーバ+LAN方式

• 解答中



• 解答終了後



学校ごとに異なる実施日を定め、ある程度大人数が解答する場合に想定される方式。

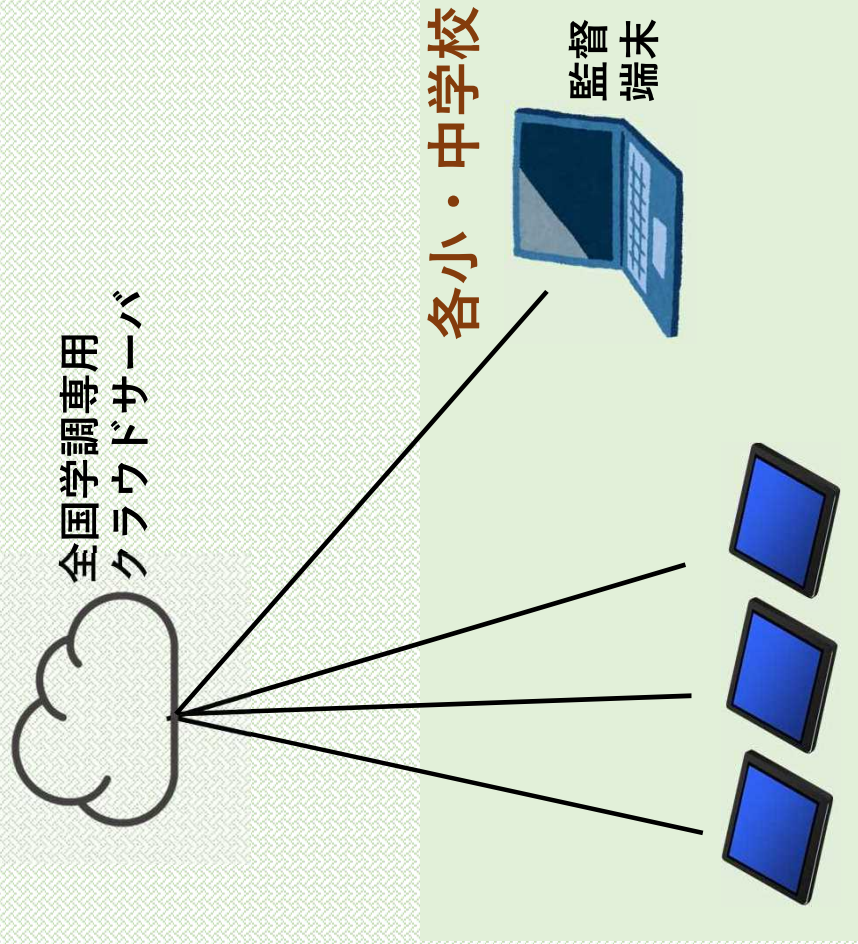
※ ある程度の大人数が1コマで調査を実施するときに、有効になり得る。

※ 解答の送信の失敗に備え、児童・生徒の端末にも解答がバックアップされるとよい。

中間サーバ+LAN方式

- **メリット**
 - 解答中は中間サーバだけとやり取りする。
⇒ WANに接続した場合と比較して、ネットワークの不安定さやその他のリスクを一定程度解消できる。
 - 解答が中間サーバに逐次送信される。
⇒ **トラブルが発生して中断されても、途中から解答を再開できる。**
- **デメリット（検討課題）**
 - **校内LANの状況に依存する。**
無線APを利用する場合、1台のAPが接続できる端末台数を確認
 - 中間サーバの設置方法や調査終了後にメインサーバにデータを送信する手順を明確に定める必要がある。

WAN方式（いわゆるオンライン方式）



・メリット

- ・手軽に実施できる。
 - ・日常的に用いられる場合は，費用に比してサーバ稼働率が高い。
- ## ・デメリット（検討課題）
- ・ネットワークが不安定な状況ではリスクが高い。
 - ・ロードバランサーの性能や校内LANの整備状況，帯域幅次第。

※ 解答の送信の失敗に備え，ローカルにも解答がバックアップされるとよい。

CBT実施方式のパターン分類

| | USB方式 | 中間サーバ+LAN方式 | WAN方式 |
|---------------|-----------------------------|---------------------------|--------------------------|
| 人的資源 | たくさんの人員が必要，負担大きい | 調査日を分散すれば，少数人員で実施可能。 | 調査日を分散すれば，少数人員で実施可能。 |
| 問題セット構成の柔軟性 | ×～△ USBに入っている問題セットのみの実施。 | △ 適応型やマルチステージは×かもしれない。 | ○ |
| 解答データの送信 | データの欠損が多数発生しうる。 | 欠損が生じにくい。 | 調査日を分散すれば，欠損が生じにくい。 |
| ネットワークラブルへの耐性 | ○ | △ (校内LAN次第) | 調査日が集中するなら×。 分散でも△。 |
| サーバへの負荷 | ○ (なし) | ○ | 調査日が集中するなら△。 分散なら○～△。 |
| 学校PCとの相性 | △ USBポートが利用できない学校もある。 | ○ 汎用webブラウザがあれば実施可能。 | ○ 汎用webブラウザがあれば実施可能。 |
| 解答状況管理 | × | ○ | ○ (taoの場合) |

児童・生徒が端末でできること

- 画面タッチ，タッチペン，トラックパッド
 - 児童・生徒の好みに応じて，画面の操作方法を選択できるように。
 - スマホやタブレットに慣れていると，画面タッチの操作は非常に親しみのある操作方法。
 - PCに慣れていると，トラックパッドを利用する場合もある。
- キーボード入力・手書き入力
 - キーボード操作に慣れていない児童・生徒のために，どちらでも解答できるようにする。（⇒議論が必要な点でもある）
 - キーボード操作と手書き入力の違いによって，問題の正誤状況（正答率や項目特性曲線）に影響を及ぼさないか，解答時間に差が生じないかどうかは**実証的に確認**する必要がある。

監督機能の重要性

- **問題セットが正しく配信されたことを確認**
 - 問題セットを複数バージョンを使用する場合には、割り当て予定の問題セットが確かに配信されていることを確認する必要がある。
- **解答が確実に送信されたことを確認**
 - トラブル等で解答が途中で中断されても、再ログインや再起動・端末交換などすれば、解答したところまでは復元されている必要がある。
 - 各児童・生徒が問題セットのうち「いま何割程度まで解答が終了したか」を監督する必要がある。

問題セットが正しく割り当てられているか、解答が中間サーバ/メインサーバに送信されたかを解答時間中に確認するために、**調査監督機能は必須**になる。

出題形式に関する5つのプラン

| | プラン1 | プラン2 | プラン3 | プラン4 | プラン5 |
|--------------|---------------|------------------|---------------|---------------------------|---------------------------|
| 実施方式 (概要) | 1人1回 分割日実施 | 1人1回 分割日実施 | 1人1回 分割日実施 | 1人複数回 | 1人複数回 |
| 出題方式 | 1フォーム | 分冊方式 | LOFT方式 | LOFT方式 | 適応型 (multi-stage含) |
| 問題バンク | 不要 | 必須でない (あると便利) | 必須 | 必須 | 必須 |
| 実施時期※ | 4月～5月 | 4月～5月 | 4月～5月 | 随時可 | 随時可 |
| 調査問題の 公表 | 全日程終了後に 可 | 全日程終了後に 可 | 原則不可 | 原則不可 | 原則不可 |
| 調査実施 方式 | 中間サーバ+ LAN | 中間サーバ+ LAN | 中間サーバ+ LAN | 中間サーバ+ LAN, または WAN | 中間サーバ+ LAN, または WAN |

※ COVID-19の影響により調査実施時期に変更が生じた場合は、この限りではない。

分冊方式

多数の小問群（ブロック）を複数の問題セットに割り付ける（重複テスト分冊法）。

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|-----|---|----|----|----|---|----|----|---|----|----|----|----|----|
| 第1問 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 第2問 | 2 | 3 | 4 | 10 | 6 | 13 | 12 | 9 | 1 | 11 | 5 | 8 | 7 |
| 第3問 | 4 | 10 | 11 | 5 | 7 | 12 | 9 | 2 | 3 | 6 | 13 | 1 | 8 |
| 第4問 | 7 | 12 | 8 | 9 | 3 | 4 | 11 | 6 | 13 | 1 | 2 | 5 | 10 |

↓ 4/10 4/10
 1校時 2校時

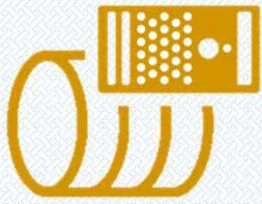
↓ 4/28
 6校時

※ 表中の1～13は、ブロックの番号を表している。

- **調査実施コマ1つにつき、問題セット1つを実施。**
- **調査の実施コマがすべて終了するまでは問題を公開できない。**
（調査全日程が終了した後に公開する）
- 学力調査でよく用いられるデザインだが、調査実施コマが柔軟に設定されうる CBT向きかどうかは議論の余地あり。

LOFT (Linear On-the-Fly Testing) 方式

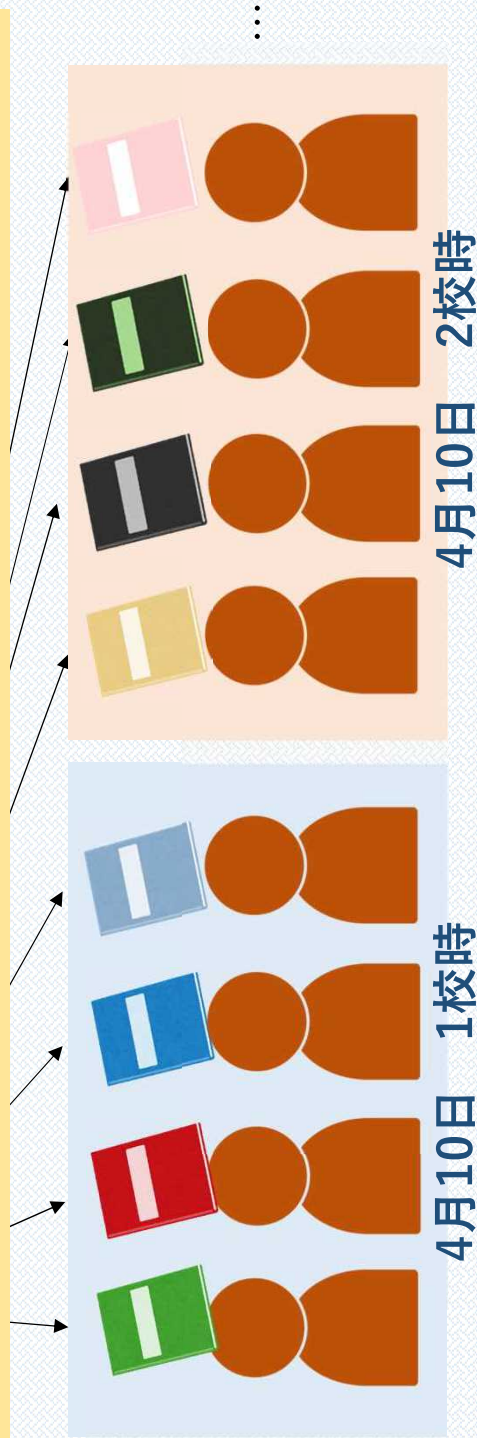
数千～数万問



問題バンク

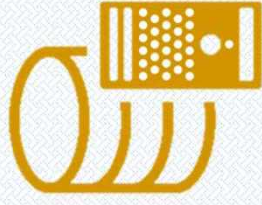
- 調査問題本体
- 調査問題の難易度をはじめとする統計量
- 出題領域
- 過去に出題された回数
- 解答に必要な時間 など

出題領域や問題の難易度、解答時間などが同じぐらいになるように問題セットを自動編集して配布



適応型とマルチステージ方式

数千～数万問

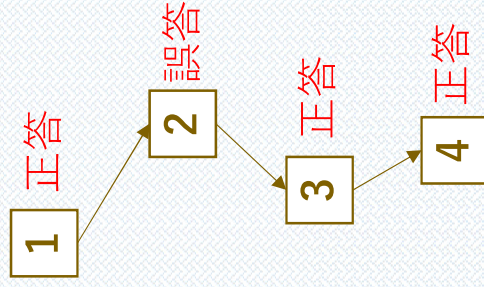


問題バンク

- 調査問題本体
- 調査問題の難易度をはじめとする統計量
- 出題領域
- 過去に出題された回数
- 解答に必要な時間 など

適応型 (アダプティブ)

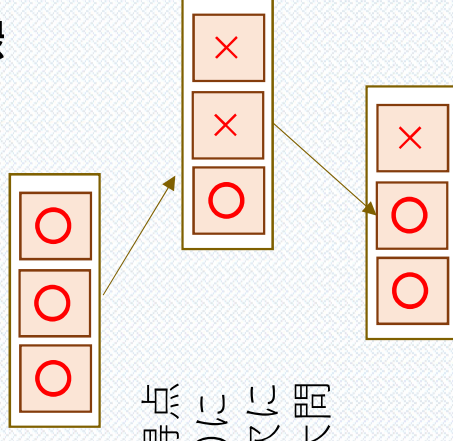
易 ← 問題の難易度 → 難



小問の正誤に応じて、次に
出題する問題
を変える。

マルチステージ方式

易 ← 問題の難易度 → 難



大問内の得点
(推定値合)に
応じて、次に
出題する大問
を変える。

第8回CBT化検討WG
乾先生説明資料より抜粋

AIP 理研AIPセンター RIKEN 自然言語理解チーム

応用テーマ：**教育アセスメント** × **言語処理**



教育における言語情報

- 思考力・判断力・表現力の
訓練・実践の媒体
- 記述式答案、作文・エッセイ
- 学習者のメモ、振り返りログ
- 教員の解説・フィードバック



自然言語処理

- 採点・診断・添削
- 解説・分析・対話

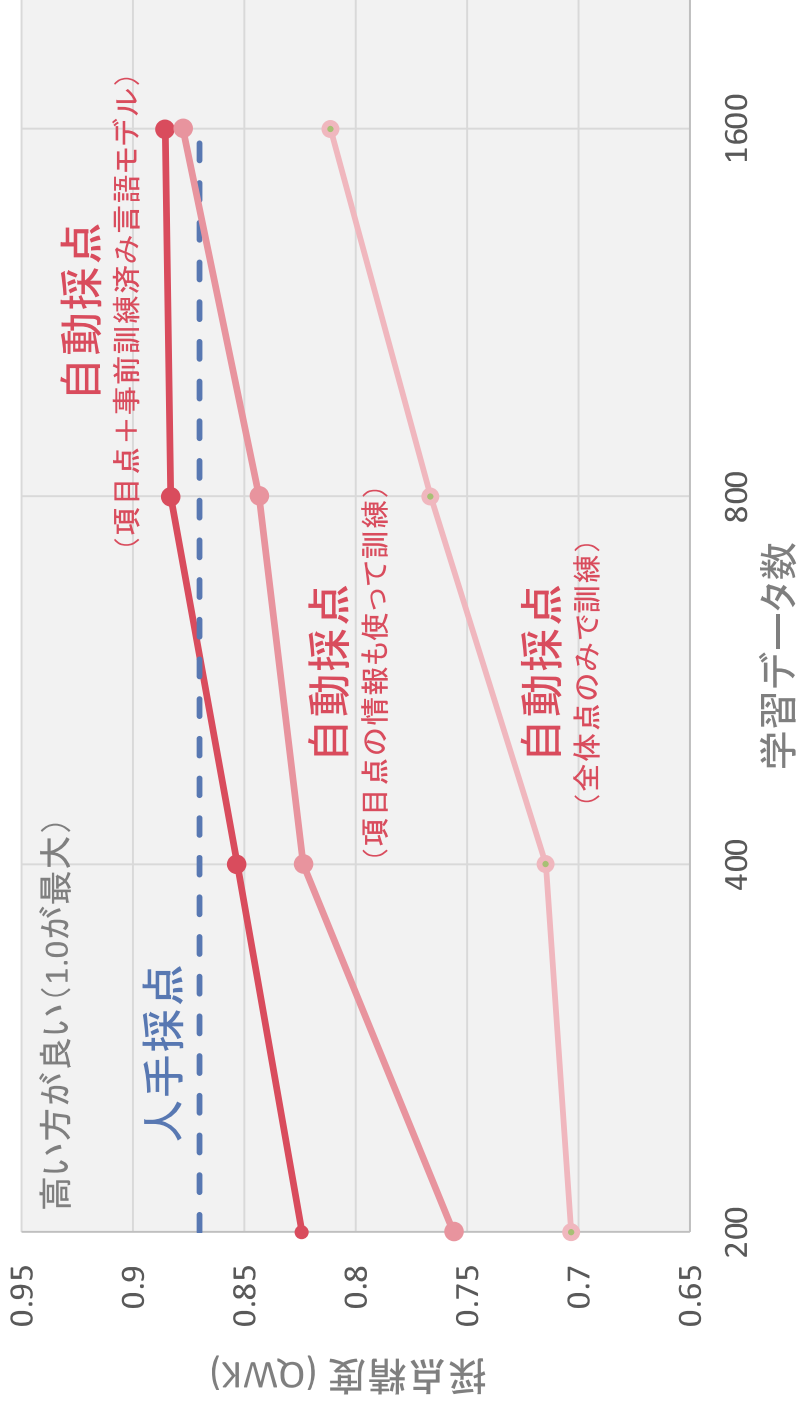
ねらい

- 先端技術を用いた**新しい学習/教育方法**の開発
 - 学習者のパフォーマンスの即時的な評価
 - 適応的なフィードバックによる対話的学習環境
 - 学習者のパフォーマンスの集成的な分析
- 自然言語処理の**未解決重要課題**（骨のある問題）
 - 不完全な文章の「**頑健**」な解析
 - 採点/評価結果の「**説明**」
 - 「(自分の判断を)説明できるAI」はAIの重要課題
 - 書き手の「**意図**」の推測
 - 「(文脈からの)意図の理解」は言語処理の重要課題

現在の到達点

1問当たり1000答案程度の訓練データがあれば
 (平均的には) **人間と同等**の採点精度が得られる

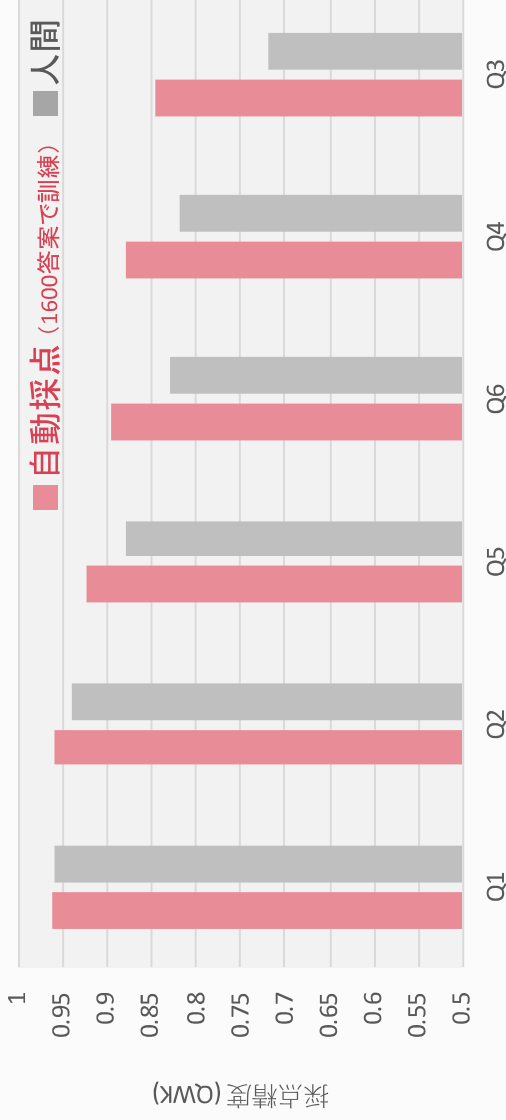
高校「国語」模試データ(代々木ゼミナール提供)での結果



(Mizumoto+ BEA 2019, Funayama+ ACL-SRW 2020)

採点が容易な問題と困難な問題

人間にとって採点が難しい問題は機械にとっても難しい



Q1 採点基準より一部抜粋

- ① 「他人は自分と違う」という説明 3点
- ② 「違い」の中身が具体化できていること 5点
(「他人と異質な考え方」)

Q3 採点基準より一部抜粋

- ① 「未練が断ち切れていない」ことを表現 4点
- ② 困惑した心情の表現が入っていれば 2点
(ネガティブなニュアンスの表現を広く認める)

含むべき内容の**基準が具体的**

=情報包含判別タスク

得点になる**表現が多様**、境界が**主観的**

→ 人の判断もゆれる、自動採点も困難

確信度推定の導入

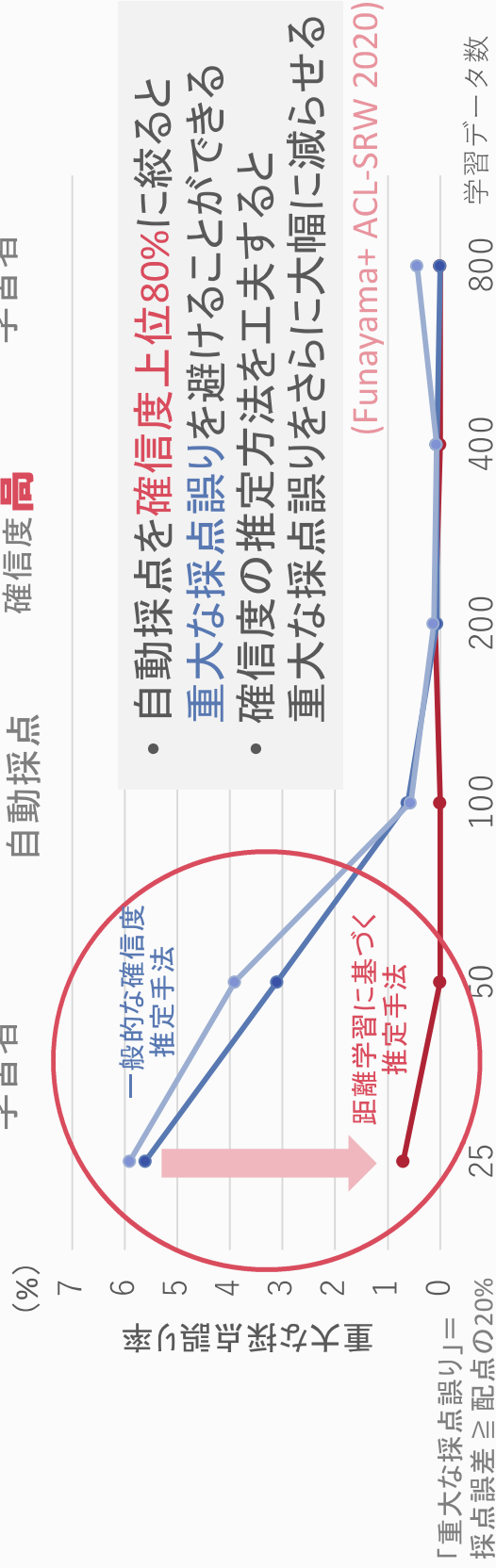
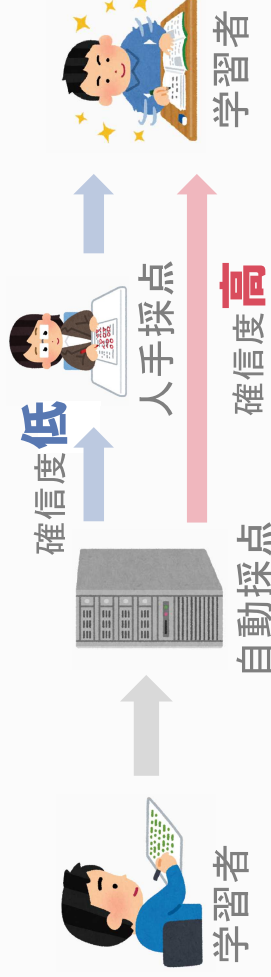
自動採点の個々の結果には**確信度 (confidence)**を付与できる

- ユーザ(学習者・教員)に**確信度を開示** → **納得感・信頼の向上**

– 何も説明がないと、「なんでこの答案でこの点数なのか？」
→ せめて**確信度情報**があれば、

「モデルも『自信がない』と言っているので、採点が間違っているのかも」

- 「**採点確信度が低い**答案は**人手採点に回す**」という使い方も可能



- 自動採点を**確信度上位80%**に絞ると**重大な採点誤りを避ける**ことができる
- 確信度の推定方法を工夫すると**重大な採点誤りをさらに大幅に減らせる**

AIは言葉の意味がわかっているわけではない

例えば自動翻訳はすいぶん流暢になったけれど、
 問題・答案を「理解」して採点・添削するAIが出現する訳ではない。
 人間の教師の代わりができるような技術にももちろんならない。



The screenshot shows the DeepL translation interface. On the left, the original Japanese text is displayed: 「教育アセスメント」と「自然言語処理」の接点をテーマとしたシンポジウムを開催します。理化学研究所では、記述式答案や論述文など、学習者の記述・論述を自動解析し、「どこがどう良いか/不足か、どう直すか」といかに評価・説明・助言する新しいAI技術の研究開発を進めています。本シンポジウムでは、下記のような方々に向けて、我々のチームが進めている研究開発を紹介するとともに、教育事業者をお招きしてパネル討論を行い、教育アセスメント×言語処理の可能性を参加者の皆さんと一緒に考えます。

On the right, the translated English text is shown: A symposium on the interface between "educational assessment" and "natural language processing" will be held. Our research and development at RIKEN is focused on the development of new AI technology that automatically analyzes learners' written answers and essays and evaluates, explains, and advises on what is good, what is lacking, and how to fix it. In this symposium, we will introduce our team's research and development to the following people and invite educators to participate in a panel discussion to explore the possibilities of educational assessment and language processing together with the participants.

Navigation elements include: 原文 日本語 (自動検出) | 訳文 英語 (us) | 用語集 | DeepL | 翻訳ツール | DeepL Pro | プランと価格 | アプリ | ログイン | Macにダウンロード 無料!

1. 「話すこと」調査の実施方法 (報告書 p2～)

＜経緯＞

- ◆平成27年6月：
「生徒の英語力向上推進プラン」にて、
全国学力・学習状況調査で英語4技能調
査の実施を決定。
- ◆平成27年7月～平成29年3月：
「全国的な学力調査に関する専門家会
議」(英語調査検討WG)にて、具体的な
実施方法を検討。教員の負担等を考慮
し、教員による面接調査ではなくコン
ピュータやタブレット等による 音声録音
方式が適切との提案。
- ◆平成30年5月：
予備調査実施 (特別支援学校を含む
公立中学校135校で実施完了)。

＜31年度調査の実施方法＞

◆実施方法

- ・予備調査と同様、Windows対応の学校PCやUSBを活用。
- ・1学級一斉実施、1授業単位時間に3学級実施。

◆調査プログラムの構成(本調査の委託事業者が開発)

- ①「話すこと」調査プログラム: CBTプラットフォーム「TAO」を採用
(PISAと同様)
- ②回収ツール: PCから音声データを回収し、データ数を確認

◆特例的な措置

- ・学校ICT環境の状況を十分踏まえた上で、設置管理者の判断により、調査を実施しないこととする。ことがで
きる。こととする。
- ・各学校のICT環境が非常に多様であり、その整備状況によって準備や負担が様々であることを踏まえ、今回の「話すこと」調査に
限っての措置。

◆実施手順

- ①平成31年 1～3月: 各学校における事前検証
調査で使用するPC1台で、検証用プログラムを展開
→調査実施→データ回収→音声データの確認
- ②3～4月: 本番用の調査プログラムのPC全展開
- ③4月: ヘッドセットの接続確認
- ④調査前日: 調査プログラムのPWロック解除
- ⑤4月18日: 調査実施
- ⑥調査後同日中: 学校にて音声データをUSBに回収
- ⑦データ回収後: PC上の調査プログラムを削除

2. 「話すこと」調査の実施状況(報告書 p9～)

- ◆中学校英語調査実施生徒数・学校数: 982,944人・9,988校
- ◆設置管理者の判断で実施しなかった学校(特例的な措置): 52,022人・434校
- ◆「話すこと」調査実施生徒数・学校数:
(英語調査実施生徒数の94.3%・学校数95.0%)
- ◆音声データ欠損等のあった生徒数・学校数(実施後に判明): 15,298人・1,658校
(「話すこと」調査実施生徒数の1.6%、学校数17.5%)

3. 検証方法 (報告書 p11～)

- ◆特例的な措置の適用について
→設置管理者等への意見聴取により全学校の状況を確認
- ◆音声データ欠損等の状況について
→欠損等が集中発生した6教育委員会の学校PCで事前検証ツール及び調査プ
ログラムの作動状況を再現、意見聴取など
- ◆調査プログラムの第三者による確認
→東京大学の協力により、PC1台あたり3回の動作及び音声データの録音保存
状況を検証し、正常動作を確認

4. 発生した主な事象の整理 (報告書 p12～)

(1) 実施要領に基づく特例的な措置の適用に関して (適用した理由)

- (a)推奨環境に満たないOSの使用、PC台数不足。
- (b)推奨環境は満たしていたが、事前検証ツールが正常に作動しなかった。
(OSのバージョンが古い、CPU、メモリ、HDDの容量に余裕がない、ヘッドセットで録音できない等)
- (c)環境復元機能の一時解除、シンククライアント方式のため必要となった高性能USBメモリの購入、PC整備の予算措置が間に合わなかった。
(特例措置適用全体の10%)
(同30%)
(同50%)
(同5%)
(同5%)
- (d)整備している学校PCがWindowsOSではなく、調査実施のためのWindows機を確保できなかった。
- (e)新設・大規模改修等により、前年度に事前検証ができず、年度当初の準備が間に合わなかった。

(2) 音声データ欠損等の発生に関して

① 音声データ欠損等の発生の原因と考えられるPCの事象

- ヘッドセットと内蔵マイクとのハウリングの可能性
- PCの動作スペースに余裕がなく、プログラムが正常に作動しなかった可能性

例えば、

- ・他のソフトウェアが調査実施時に作動しPCに負荷がかかり、調査プログラムが正常動作せず。
- ・前日に環境復元機能を解除したところ、長時間OSアップデート等がなされ、調査中もPCに負荷。
- ・シンクライアント等において、端末内での処理が間に合わず、迅速な処理が求められる短い設問において正常動作せず。

- 個別のPCの調査当日の偶発的な不具合の可能性

② 調査設計上の課題

- 今回の調査プログラムには、PCスペースの余裕との関係上、調査時に録音・保存が正常になされているかを確認する機能がなかったこと。
- 生徒数と回収データ数が一致しているかの確認方法が十分には機能しなかったこと。
- 音声データ欠損等の有無の確認は調査期間終了後となるが、採点システムの制約上、音声データ欠損を理由とする調査のやり直しは当日限りという取扱いになっていたため、やり直しができなかったこと。

(3) 生徒の実施状況等に関して

- 学校や設置管理者からは、英語4技能の育成の重要性と「話すこと」の意義を再認識したとの意見があった。
- 生徒からは、調査プログラムの操作方法に戸惑い調査開始が遅れた、近隣の席の生徒の声が届かなくて調査ができなかった等の意見があった。
- (4) 調査実施担当教職員の作業に関して
 - 設置管理者や学校は、「話すこと」調査の実施に向けて、最大限の準備と対応をしていただいた。(独自マニュアル作成、研修会の開催、ICT支援員の派遣など)
 - 他方、調査が4月であり、準備から実施までに教職員の人事異動があることや、作業量の多さについての時間的・心理的負担感があったとの意見があった。

5. 課題の整理・分析 (報告書 p17～)

(1) ICT活用の利点を生かした調査方法の設計

- 用いるハード・ソフトウェアの種類を簡潔にすること。
- 学校のICT環境を利用する場合は、多様なICT環境に対応できる工夫を講じること(多様なOSへの対応など)。
- 学校PC等とWebブラウザを活用したオンライン調査の導入が大いに期待される。
 - この場合、100万人規模での一斉調査であることから、サーバーの構築、通信環境の確保等が不可欠。
- (2) 各生徒の調査実施(録音・保存)と解答データの回収の有無の確認
 - PC等には不測の事態が生じうることを前提に、当日やむを得ず実施できない場合やデータ欠損等が生じた場合の調査やり直し等の期間・手順等の設定。
 - 全設問が確実に録音できているかどうかを確認できる仕組みの設定。

(3) 近接する生徒からの影響の抑制

- 生徒が自分の解答に集中できるよう、外部音の遮断や間隔をあげた座席配置等の配慮。
- 調査を2日間にわけて実施すること(同一問題の一再実施にこだわらない在り方の検討も論点のひとつ)。

(4) 生徒が調査方法を体験する機会の確保

- 事前に、調査プログラムと同型プログラムを提供する等、生徒が調査方法を事前に体験できる機会を設定。

(5) 調査環境を整備するための期間の確保等

- 設置管理者が調査に向けて必要なICT環境を整備することが可能となるように、遅くとも調査実施年度の2年前には、実施方法、求められるICT環境について提示。
- ICT環境の整備にあたっては、動作や更新等のタイミングをPC管理者が制御できるよう、学校PCとして一層使いやすい環境整備を期す。

5 埼玉県学調のCBT化に向けて：(1) 埼玉県学力・学習状況調査CBT導入推進事業

現状・課題

【埼玉県学力・学習状況調査】

- 対象：公立小中学校(さいたま市を除く)の小4～中3(毎年度約30万人)
- 実施形態：紙媒体(Paper-Based-Testing)で実施(以下、PBT)
- 特長：一人一人の「学力の伸び(経年変化)」や「非認知能力」などを継続して把握
- 今後の展開：調査結果とAIを活用した実証研究を実施。個に応じた指導の実現を目指す。

【国の動き】

- 新時代の学びを支える先端技術活用推進方策(R1.6)
ICTを基盤とした先端技術や教育ビッグデータを効果的に活用し、誰一人取り残すことのない、公正に個別最適化した学びの実現が求められる。
- GIGAスクール構想：児童生徒1人1台端末、高速大容量通信環境を整備(R2～R4)
- 全国学調：CBT(Computer-Based-Testing)化検討WGにて検討(R2.5月～)
CBT化に向けた取組をR3概算要求。小規模試行とオンライン学習システムの検証

【CBT導入のメリット】

- より精緻なデータの取得・蓄積・分析が可能
 - ・解答までの時間などのログや思考過程を取得することが可能
 - ・将来的に個別最適化学習のツールになり得る。
- 業務工程の効率化(コスト軽減にも寄与)
 - ・調査資料の印刷、配送・回収が不要
 - ・選択式問題の自動採点化などによる調査結果返却のスピードアップ

【CBT導入に向けた課題】

- 全小中学校で安定的・継続的に実施できる方式の検討
 - ※大量の問題・解答データの送受信方法(オンラインor外部媒体 etc)
 - ※個人情報取扱い(外部サーバーとオンラインで回答情報をやりとりするなど)
- CBTの利点を活かした問題作成、出題・採点方式・標準規格等の検討
- CBTの実施に必要な環境の整備(問題移行(PBT→CBT)、結果比較(PBT⇄CBT))
- 児童生徒(特に小学生)がICT機器を使用した学習に慣れる必要
 - ※授業におけるデジタル機器の利用時間がOECD加盟国中最下位

取組の方向性

試行・実証を重ねながら、県学力・学習状況調査を紙媒体での調査からCBT調査へ段階的に移行する。

事業内容

【令和3年度：フェーズ1】

- ・調査スキームの検討(問題設計、実施体制 etc)
- ・モデル校(複数市町村の小中各4～5校)でCBT調査を試行(問題作成、調査実施、採点まで)
- ・CBTの実施に必要な環境の確認(問題移行、結果比較)
- ・課題の検証、対応策の検討、スキームの修正

【令和4年度：フェーズ2】

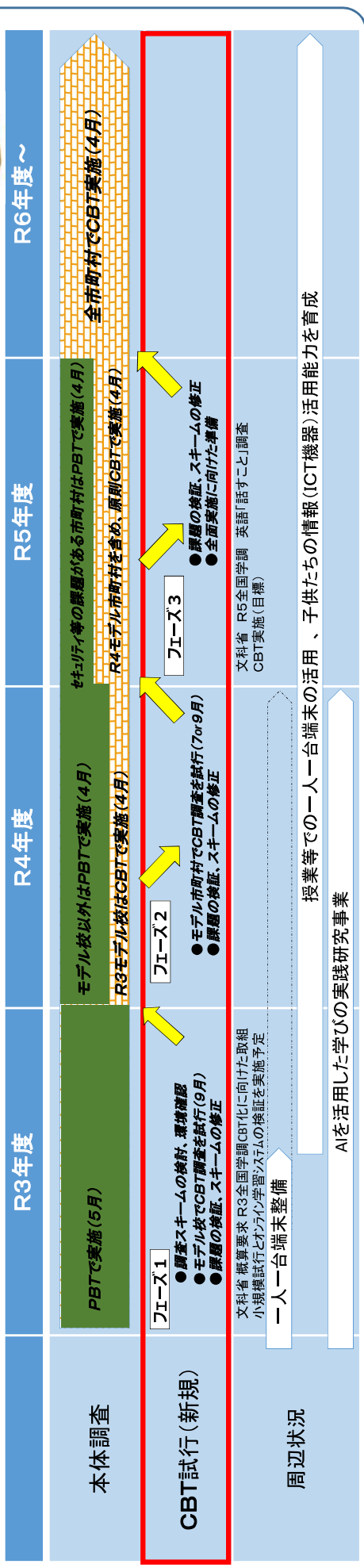
- ・R3モデル校は本体調査でCBT実施
- ・モデル市町村全体(R3モデル校のある4～5市町村)でCBT調査の試行(問題作成、調査実施、採点、分析)
- ・課題の検証、対応策の検討、スキームの修正
- ・市町村・学校への説明、実施に向けたアフォー

【令和5年度：フェーズ3】

- ・本体調査を原則CBTで実施
- ・課題の検証、対応策の検討、スキームの修正
- ・市町村・学校における全面实施に向けた準備
- ・個人結果票の電子化にあたり、「個別アドバイ

【令和6年度～】

CBT全面実施



精緻なデータの蓄積・分析を実現し、個に応じた指導により、学力や非認知能力など子供達の成長につなげる！

5 埼玉県学調のC B T化に向けて：（4）全国学調との関係等

（全国学調と地方学調（埼玉県学調）の役割分担）

○埼玉県では、全国学調と埼玉県学調のそれぞれの特徴を踏まえて、以下のように活用を進めている。

- ・全国学調は、調査問題や解説資料等を通して、新学習指導要領で示された「児童生徒に身に付けさせたい資質・能力」を具体的なメッセージとして示している。（県としても、対象学年・教科以外の教員も含め教員全体で全国学調を解くなどして、「児童生徒に身に付けさせたい資質・能力」を組織的に把握するよう指導している）
- ・埼玉県学調は、児童生徒の学力の伸びを測定できるという独自の特長を有する。
- 全国学調と県学調のそれぞれの機能を踏まえた活用が重要。

仮に、全国学調を I R T で実施し問題を非公開とするとしても、教育指導上の役割を担保する仕組みが必要ではないか。

（全国学調 C B T 化への期待・提案）

① 実施準備について

- ・令和元年度の「話すこと」調査では、U S B 方式による実施で学校現場に大きな負担があった。次回（令和 5 年度）の実施に当たっては、事前の動作確認等も含めて、学校現場への負担に配慮いただきたい。
- ・特に、事前のフィードバックも大規模一斉に行うのではなく、小規模での実証・改善を段階的に重ねることが必要ではないか。
- ・仮に G I G A スクール端末を活用して次回の「話すこと」調査を行う場合は、ヘッドセット等の機器準備も必要ではないか。

65

② 実施方式について

- ・規模を生かして、複数回実施や個に応じて問題を変えるような実施方式も検討いただきたい。
- ・特別な教育的支援を要する児童生徒に関して、学校に負担がかからない調査の仕組が必要（日本語指導を要する児童生徒については、振り仮名が画面上に表示されるなど）。

③ 調査結果について

- ・仮に全国学調が I R T 方式を取っても、自治体別の平均正答率で実施結果を公表すれば、その数値ばかりが着目されることになる。I R T に適した調査結果の公表の仕組が必要。
- ・自治体職員や教員にデータ分析の知見は乏しいので、分析結果を分かり易く指導主事や教員に伝える仕組みや、分析結果から課題を読み解くためのデータスキルの育成が不可欠。

④ その他

- ・キーボードなど P C 操作面の影響が少なくなるよう、並行して児童生徒の P C 操作スキルの積極的な育成が重要。教育指導面での対応も必要ではないか。

国と自治体との役割分担・協働の方向性について（提案）

- 文部科学省や他府県教育委員会等との密接な連携・協働を実現すべく、これらの主体がIRT・CBTについて情報共有や意見交換ができるコンソーシアムのような「枠組み」を構築することが必要ではないか。
そのような枠組みでの交流・議論を通じ、公立学校でのIRT・CBT活用の意義やメリットの優れた「見える化」も促進できるのではないか。
- 既に示した全国学調の実績と意義を踏まえつつ、全国学調と地方学調との関係が協動的に整理されるべきではないか。
その中で、今後、既述の「枠組み」の活用も通じ、文部科学省による地方学調の実態把握や、地方学調におけるCBT化・IRT導入・パネルデータ化の支援が行われるとともに、高度な知見を必要とするIRT・CBTの活用をはじめ、今後、地方自治体において指導主事等（ひいては学校現場）のデータリテラシーの育成が急務となることから、文部科学省がその育成の支援を実施することが重要ではないか。
- 当面、京都府教育委員会としても、IRT・CBT実証研究を通じ、学校現場での学力向上や指導改善に注力することになるが、中期的には、全国学調を含め、こうした学力・学習状況調査で集まるデータの利活用、すなわち、教育施策・活動（インプット）の効果の把握・分析結果について、どのように教育施策・活動の改善に繋がっていくのかについて、国・地方が協調した仕組みの検討が必要ではないか。