

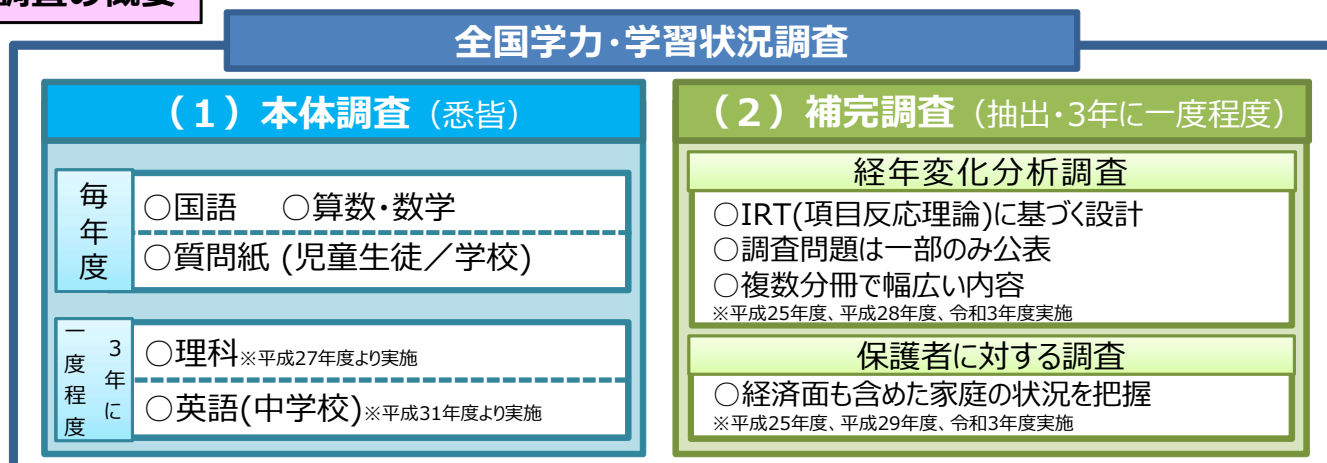
全国学力・学習状況調査の概要

調査の目的

義務教育の機会均等とその水準の維持向上の観点から、

- 全国的な児童生徒の学力や学習状況を把握・分析することによって、
国や全ての教育委員会における教育施策の成果と課題を分析し、その改善を図る
- 学校における個々の児童生徒への教育指導や学習状況の改善・充実に役立つ
- そのような取組を通じて、教育に関する継続的な検証改善サイクルを確立する

調査の概要



(1) 本体調査

【対象】 国・公・私立学校の小学校第6学年、中学校第3学年 全児童生徒
小学校約2万校、中学校約1万校 児童生徒約200万人

【実施日程】 毎年度 全国同日一斉 (調査日は、例年4月中旬)

【内容】

○教科に関する調査

・国語、算数・数学 / (3年に一度程度) 理科、中学校英語
(出題内容)

- ①身に付けておかなければ後の学年等の学習内容に影響を及ぼす内容や、実生活において不可欠であり常に活用できるようになっていることが望ましい知識・技能等
- ②知識・技能を実生活の様々な場面に活用する力や、様々な課題解決のための構想を立て実践し評価・改善する力等

○質問紙調査

- ・(児童生徒に対する調査) 学習意欲、学習方法、学習環境、生活の諸側面等
- ・(学校に対する調査) 指導方法に関する取組や人的・物的な教育条件の整備状況等

(2) 補完調査 (経年変化分析調査・保護者に対する調査)

○経年変化分析調査

【対象】 統計的手法に基づき抽出した学校の小学校第6学年、中学校第3学年の児童生徒
小学校約600校、中学校約750校 児童生徒約11万人 ※令和3年度

【実施日程】 3年に一度程度 約1か月の期間中、対象学校が実施可能な日

【内容】

- ・国語、算数・数学、中学校英語のうち、いずれか1教科を実施
- ・IRT (項目反応理論) に基づく調査設計 (複数分冊で幅広い内容を出題)

○保護者に対する調査

【対象】 本体調査及び経年変化分析調査を実施した児童生徒の保護者

【内容】 児童生徒の家庭における状況、保護者の教育に関する考え方等に関する質問紙調査

GIGAスクール構想の実現とは

Society 5.0時代を生きる全ての子供たちの可能性を引き出す、個別最適な学びと協働的な学びを実現するため、児童生徒の「1人1台端末」等のICT環境を整備

→ 令和元年度から令和5年度までの計画として、令和元年度補正予算において、学校における児童生徒「1人1台端末」と、高速大容量の通信ネットワークを一体的に整備するための予算（2,318億円）を計上。

→ 令和2年度第1次補正予算において、「1人1台端末」整備の前倒しや、家庭でも繋がる通信環境の整備など、災害や感染症の発生等による学校の臨時休業等の緊急時においても、ICTの活用により全ての子供たちの学びを保障できる環境の整備に必要な予算（2,292億円）を計上。

→ 上記に加え、「国民の命と暮らしを守る安心と希望のための総合経済対策（令和2年12月8日閣議決定）」を踏まえ、令和2年度第3次補正予算、令和3年度予算へ「GIGAスクール構想の拡充」等、ICT環境の整備や、活用に必要な経費を計上。
これらを通じて、GIGAスクール構想の実現をさらに加速。

～1人1台端末・高速大容量ネットワークが広げる学びの可能性～

GIGAスクール構想が目指す学びのDX

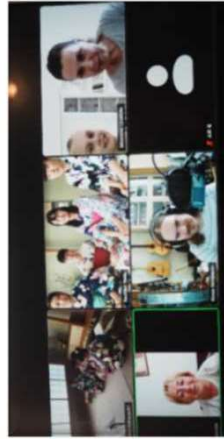
中山間地域の学校における
遠隔授業の活用



端末を「文房具」としてフル活用した
学校教育活動の展開

- 学習の基盤となる情報活用能力の育成
- 動画や音声も活用し、児童生徒の興味を喚起、理解促進
- 情報の収集・分析、まとめ・表現などによる探究的な学習の効果的な推進
- 障害のある児童生徒の障害の特性に応じたきめ細かな指導・支援の充実など多様なニーズへの対応
- 板書や採点・集計の効率化等を通じた学校の働き方改革

海外の学校との交流学習



大学や企業等と連携した学習



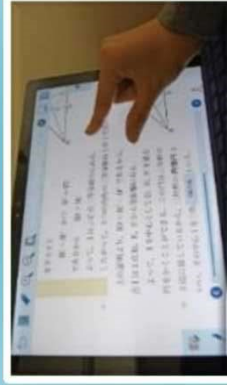
地域の機関や外部人材と
連携した学習



端末を「文房具」としてフル活用した
学校教育活動の展開

- 学習の基盤となる情報活用能力の育成
- 動画や音声も活用し、児童生徒の興味を喚起、理解促進
- 情報の収集・分析、まとめ・表現などによる探究的な学習の効果的な推進
- 障害のある児童生徒の障害の特性に応じたきめ細かな指導・支援の充実など多様なニーズへの対応
- 板書や採点・集計の効率化等を通じた学校の働き方改革

学習者用デジタル教科書の活用



様々なデジタル教材の活用



学習履歴等を活用したきめ細かい
指導の充実や学習の改善



不登校児童生徒に対する
学習指導



病気療養児に対する学習指導



臨時休業時における
オンラインを含む家庭学習



発達段階に応じて遠隔・オンライン教育も積極的に活用

全ての子供たちの可能性を引き出す、
個別最適な学びと、協働的な学びを実現

端末の調達に関する状況

確定値

○ 調査の概要

- 公立の小学校、中学校、義務教育学校、中等教育学校(前期課程)及び特別支援学校(小学部・中学部)の端末の整備状況(令和3年3月末時点)
- 提出自治体等数:1,812自治体等 ※「自治体等」とは都道府県、市区町村、一部事務組合を含む公立学校情報機器整備費補助金の対象である公立の義務教育段階の学校設置者

○ 納品完了時期

全自治体等のうち **1,748自治体等(96.5%)** が令和2年度内に納品を完了する見込み

※「納品完了」とは児童生徒の手元に端末が渡り、インターネットの整備を含めて学校での利用が可能となる状態を指す。

(自治体等数)

令和元年度 までに整備済	令和2年度内に納品済み (389自治体等・21.5%)				令和2年度内に納品 (1,337自治体等・73.8%)			4月以降納品	
	4～8月まで に納品済	9月	10月	11月	12月	1月	2月		3月
22 (1.2%)	29 (1.6%)	28 (1.5%)	56 (3.1%)	67 (3.7%)	209 (11.5%)	138 (7.6%)	323 (17.8%)	876 (48.3%)	64 (3.5%)

令和2年度内に議会の承認済み(1,807自治体等・99.7%)				令和2年度内に議会の承認(5自治体等・0.3%)			
令和2年度8月ま でに承認済み	9月	10月	11月	12月	1月	2月	3月
1,440 (79.5%)	306 (16.9%)	20 (1.1%)	10 (0.6%)	31 (1.7%)	2 (0.1%)	0 (0.0%)	3 (0.2%)
令和2年度内に入札の公示済み(1,776自治体等・98.0%)				令和2年度内に入札の公示(28自治体等・1.5%)			
令和2年度8月ま でに公示済み	9月	10月	11月	12月	1月	2月	3月
1,166 (64.3%)	221 (12.2%)	223 (12.3%)	111 (6.1%)	55 (3.0%)	13 (0.7%)	10 (0.6%)	5 (0.3%)
※8自治体等が令和3年度以降に公示予定と回答							
令和2年度内に事業者の選定済み(1,763自治体等・97.3%)				令和2年度内に事業者の選定(40自治体等・2.2%)			
令和2年度8月ま でに選定済み	9月	10月	11月	12月	1月	2月	3月
900 (49.7%)	344 (19.0%)	238 (13.1%)	154 (8.5%)	127 (7.0%)	27 (1.5%)	12 (0.7%)	1 (0.1%)
※9自治体等が令和3年度以降に選定予定と回答							

※ 端数処理の都合上、合計が100%に一致しない。
 ※ 公立学校情報機器整備費補助金(以下「補助金」という。)によって整備する端末の状況を示しており、補助金を活用せず整備している自治体等については補助金の措置分(2/3)に相当する台数についての状況を示している。

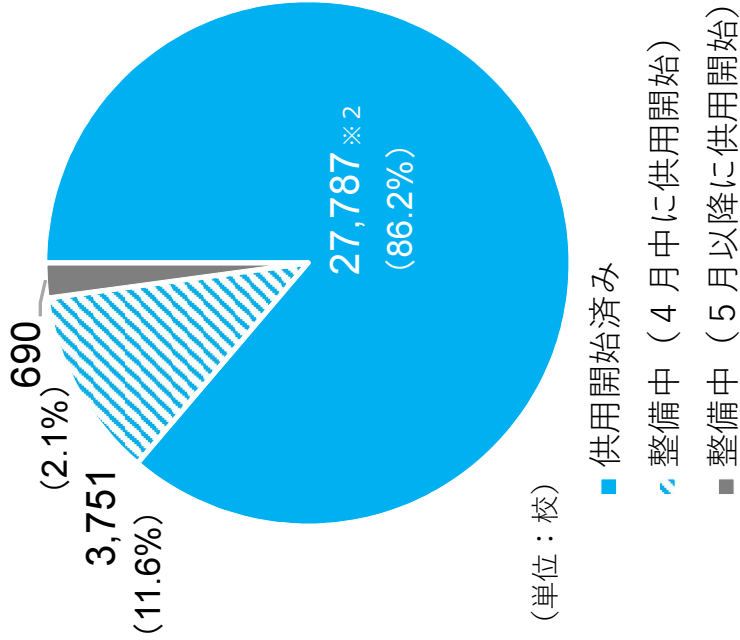
校内通信ネットワーク環境整備等の状況（令和3年3月末時点）

調査の概要

- ・ 令和3年3月末時点の公立の小学校、中学校、義務教育学校、高等学校、中等教育学校及び特別支援学校の校内通信ネットワーク環境整備等の現状（令和3年2月時点での予定）
- ・ 提出自治体等数：1,815自治体等（学校数：32,787校）

校内ネットワーク環境の現状（整備に取り組んでいる学校数：32,228校 ※1）

86.2%の学校が令和2年度内に、97.9%の学校はほぼ新学期から供用開始の見込み



・ 本年度内に供用開始	27,787校 (86.2%)
・ 本年4月末までに供用開始	31,538校 (97.9%)

※1 整備しない学校559校を除く。整備しない理由は、統廃合予定、校舎の改築予定、未光地域、小規模校のため既存で対応、休校中等。

※2 供用開始済み 27,787校の内訳

3月までに整備を完了し供用開始	24,701校
GIGAスクール構想以前に整備済み	1,934校
LTE端末で対応	1,152校

⇒ 各自治体等に対し、学校におけるネットワーク環境の安定的な確保について確認を行い、必要な施策を講じることについて通知（インターネット環境の詳細（接続速度、同時利用率等）については、多くの自治体等で校内ネットワーク整備が完了する5月以降に改めて調査予定）

学びの保障オンライン学習システム (MEXCBT) について

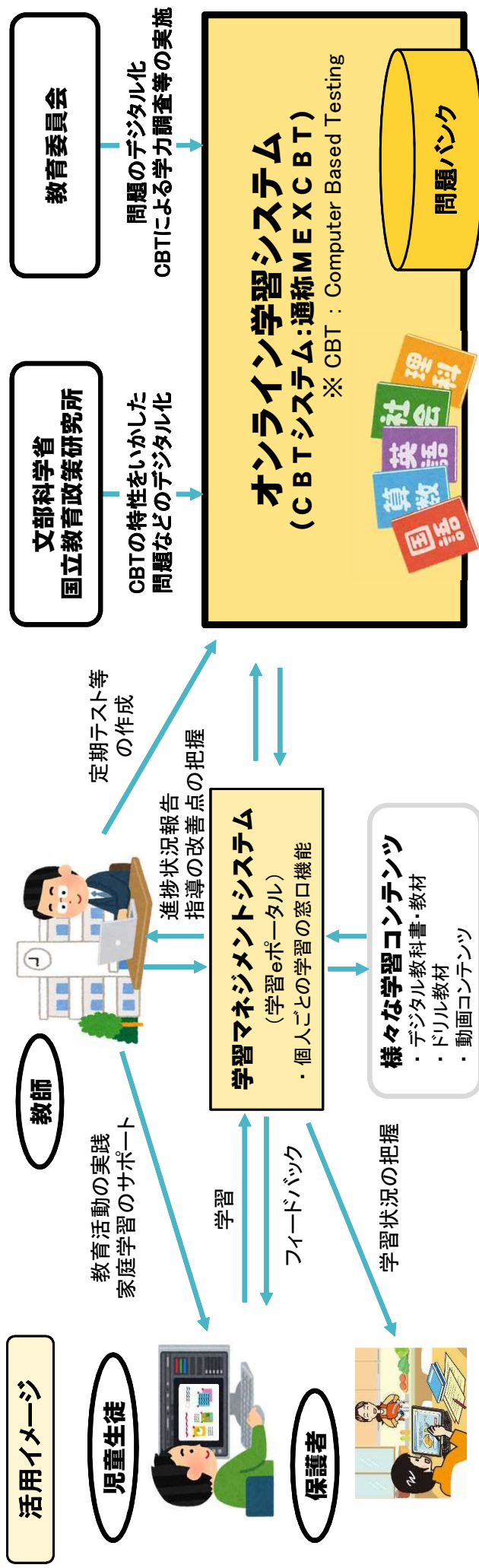
MEXCBTについて

- 緊急時における、子供たちの学びの保障の観点から、国や地方自治体等の公的機関等が作成した問題を活用し、**児童生徒が学校や家庭において、学習やアセスメントができるCBTシステム (MEXCBT:メクビット)** を文部科学省で開発 (様々な知見を総合してシステム開発を行うため、事業者連体のコンソーシアムに委託)。
- 希望する全国の小・中・高等学校等で活用可能にし、「GIGAスクール構想」により実現する「**1人1台端末**」を活用した「**デジタル**」ならではの**学び**を実現。

スケジュール

	対象学校数	搭載する問題	システム
令和2年度 (約1億円)	約300校の小・中・高校	国が作成した既存の学力調査等の問題 (全国学力・学習状況調査の問題等、約2000問)	プロトタイプ
令和3年度 (約28億円)	希望する全国の小・中・高校等で活用可能	上記に加え、地方自治体等が作成した学力調査等の問題を搭載	実証を踏まえた機能改善・拡充※ 解答結果の分析・フィードバック

※監督機能や制御機能など、大規模調査を実施する際に必要となる機能についても試行・検証する予定



学びの保障オンライン学習システム (MEXCBT) の概要と活用の流れ

システム概要

【総論】

- 児童生徒が学習端末を用いてオンラインで問題演習等ができるシステム(問題やデータの相互運用が可能な国際標準規格に基づく汎用的なシステム※)を開発

- ※MEXCBTは、国際標準規格QTI/LTIに準拠したwebベースのCBTシステムであるTAOをベースとして開発
- ※パブリッククラウド(サーバは日本国内に設置)を活用

【活用方法】

- 通常活用している学習端末を用いて、家庭からでも学校からでもアクセスが可能
- 2通りの活用方法が可能
- 選択式問題や一部短答式問題は自動採点

①一問一答形式

学年・教科を選び、一問一答形式で解答後に解説等が表示され学習する方式

②複数問題解答形式

学年等を選び、何問かの束で解答する方式



【具体的な問題】

- 国や地方自治体等の公的機関等が作成した問題を活用
(例) 全国学力・学習状況調査問題、
高等学校卒業程度認定試験問題、
自治体独自の学力調査問題など

活用の流れ

① 問題を選ぶ

学習 e ポータル



教員

② 問題を解いて学習する

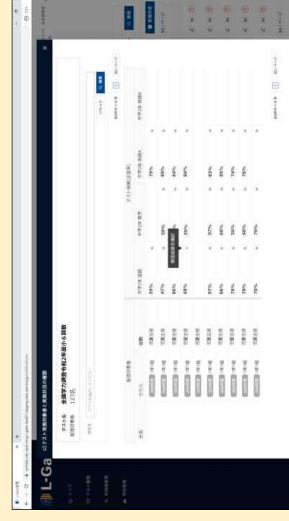
MEXCBT



児童生徒

③ 結果を確認する

クラス内の学習結果を確認



自分の学習結果を確認



学習 e ポータル

学びの保障オンライン学習システム（MEXCBT）のプロトタイプの実証（令和2年度）

プロトタイプの実証

- ✓ 将来的な全国の小中高生の活用を視野に、令和2年度はCBTシステムのプロトタイプを開発。
- ✓ 令和2年度は国が作成した既存の問題（全国学力・学習状況調査問題や高等学校卒業程度認定試験問題など）をデジタル化して、約2000問を搭載。
- ✓ 令和2年12月にプロトタイプを開発し、令和3年1～2月に、小中高約300校で実証を実施。

実証の様子



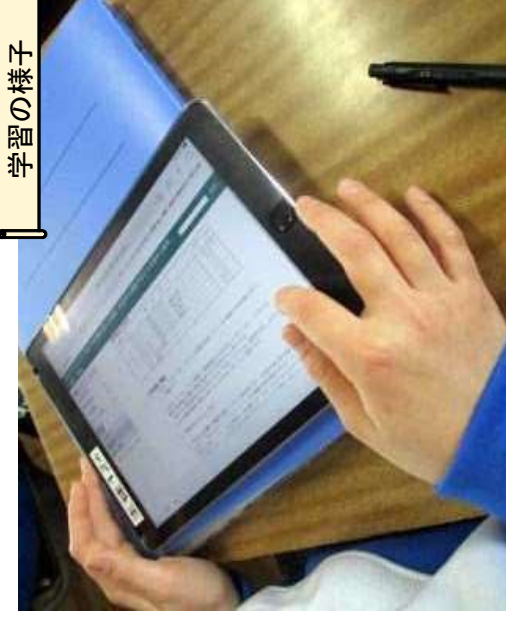
授業での活用

七尾市立朝日小学校HPより抜粋



朝学習での活用

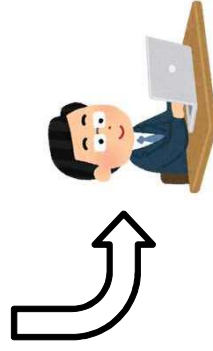
梶原町立梶原学園HPより抜粋



学習の様子

福島大学附属中学校HPより抜粋

MEXCBTを活用した現場からの声（一部抜粋）



MEXCBTは、授業中や放課後に活用したり、家庭学習（宿題）の際に活用したりした。

児童生徒は問題を解けば正答率が出るため、楽しそうに使っていた。今後も利用したい。

教員は配信するだけでテストを利用できるため、印刷や採点の手間が省け、業務効率が向上した。

全国学力・学習状況調査のCBT化に向けた令和3年度試行・検証について（案）

令和3年度予算額：50百万円

背景

- GIGAスクール構想やPISA等の国際的な学力調査のCBT（コンピュータ等を使用した調査）による実施の流れを踏まえ、全国学力・学習状況調査のCBT化について、「全国的な学力調査のCBT化検討ワーキンググループ」において、専門的・技術的な観点から検討を実施中。
- 同ワーキンググループの中間まとめ「論点整理」※（令和2年8月28日）を踏まえ、全国学力・学習状況調査のCBT化に向けて、小規模から試行・検証に取り組む。

（※）「全国学力・学習状況調査のCBT化に向けて、まずは小規模から試行・検証に取り組み、課題の解決を図りつつ、確実に段階的に規模・内容を拡張・充実させていくことが早期の進展、実現につながる」

試行・検証内容

○実施規模

- ・小学校第6学年、中学校第3学年の児童生徒 約1万人
（ネットワーク環境等を考慮し、小学校50校程度、中学校50校程度で実施）

○主な検証事項

初期段階の実証研究で確認すべきと考えられる事項を中心として試行・検証を実施

（1）ネットワーク・システムの検証

①事前のネットワーク環境

- ・試行実施前に、実証校の端末、ネットワーク等について測定し、ネットワーク環境を事前検証

②ネットワークの負荷

- ・各学校におけるネットワーク接続形式やネットワーク環境が異なることを踏まえ、2つのパターン※1で、通信負荷をかけ、ネットワークや機器に関する詳細な測定を実施
（※1）学校から直接インターネットへ接続する形式
センター等に集約して接続する形式

③CBTシステム※2のサーバの負荷

- ・調査実施時のCBTシステムのサーバの負荷状況の測定
（※2）学びの保障オンライン学習システム（MEXCBT）を活用

（2）実施体制の検証

- ・学校での実施体制や必要となるサポート体制
実証校におけるCBT実施手順、実施時の技術的トラブル等を検証
（必要に応じて実施支援員を派遣）

（3）問題の検証

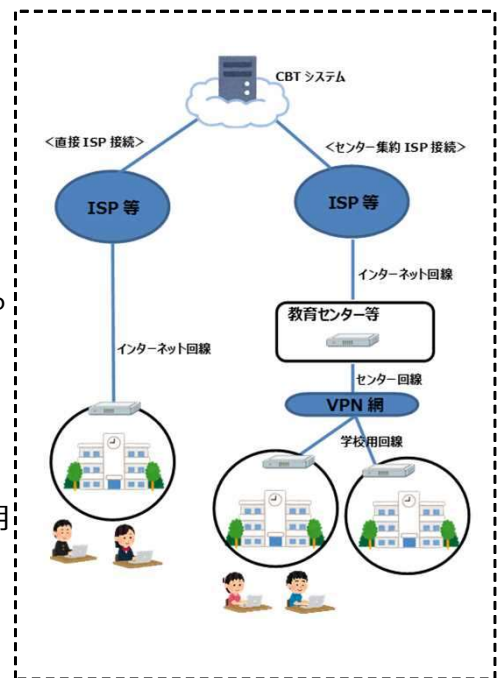
①問題表示形式等による差異

様々な問題表示形式等を使用し、CBTにおける問題表示形式や特性を活かした問題への解答等の違いを児童生徒へのアンケート等も実施しつつ、検証

②児童生徒の文字入力

児童生徒のキーボード操作等による文字入力能力と解答状況等と組み合わせて分析

＜ネットワーク・システム図＞



スケジュール（予定）

- ・**令和3年10月～11月** 実証校が実施可能な日時で試行・検証を実施
- ・**令和4年** 試行・検証の結果を踏まえ、ネットワークやシステム等、技術的な要件等について検討明らかになった課題等を踏まえつつ、着実な実施に向けて、さらに検証等を実施

「GIGAスクール構想」の実現ロードマップ（イメージ）

		2020年度 (令和2年度)	2021年度 (令和3年度)	2022年度 (令和4年度)	2023年度 (令和5年度)	2024年度～ (令和6年度～)	
ハード	ネットワーク	校内ネットワーク 学習系ネットワーク円滑化	引き続き整備を推進	建て替えや全面的な改修工事の際に併せて更新			
	学習者用端末	3人に1台分	公立小・中・高・特支等（環境整備5か年計画）	次期環境整備計画			
		3人に2台分 (義務教育段階)	小5・6 中1	小1～4 中2・3	将来的に保護者負担（BYOD）への移行を見据えつつ、「デバイス」の考え方や支援方策の在り方を整理		
		低所得世帯等 (高校学校段階)	高等学校等	BYODの推進も含めた高等学校等の設置者の取組を支援しつつ、各設置者に対して1人1台端末環境の整備を働きかけ			
人材	ICT活用教育アドバイザー		自治体への支援				
	GIGAスクールサポーター	4校に2人	4校に1人				
	ICT支援員		4校に1人				
ソフト	1人1台端末環境の利活用		「GIGA StudX 推進チーム」による支援				
	学習者用デジタル教科書 デジタル教材	デジタル教科書の在り方検討	デジタル教科書の普及促進、効果・影響の検証 デジタル教材等との連携			教科書改訂（小：R6）に合わせ、実証等を踏まえた新たな導入方策を実施	
	教育データの標準化	教育データ標準第1版 (学習指導要領コード)	教育データ標準第2版	随時、教育データ標準の改訂を実施			
	オンライン学習システム (CBTシステム)	プロトタイプ開発		システムの全国展開、CBT化の推進			

国際学力調査における実施状況

	PISA2018 生徒の学習到達度調査	TIMSS2019 国際数学・理科教育動向調査
概要	実施主体	国際教育到達度評価学会 (IEA)
	対象	小学校 4 年生、中学校 2 年生
	対象 抽出	抽出
	頻度	1995 年からは 4 年ごとに実施。(1964 年から実施)
	実施期間	2019 年 2～3 月実施に実施 (日本)
	調査項目	算数・数学、理科
結果のレポート	測定指標	参加国の国際平均値を 500 点、標準偏差を 100 点の分布モデルにおける推定値として算出する。各国の児童生徒の得点分布を調べるために、625 点、550 点、475 点、400 点という 75 点刻みの国際標準水準が設定され、各国ともその得点以上に何%の児童生徒が含まれるかが算出される。
	学生単位	X
	学校単位	X
	国単位	● (国レベル)
	結果の比較	国家間の比較
	主な目的	中等教育段階における児童・生徒の算数・数学及び理科の教育到達度を国際的な尺度によって測定し、児童・生徒の学習環境条件等の諸要因との関係を分析すること
デジタルベースの評価の概要	CBT の範囲	2019 年調査より、コンピュータ使用型調査(eTIMSS)での受験が導入された。 ※TIMSS2019 の参加国 (64 国・地域) の半数以上が、コンピュータを使用して受験
	研究開発期間	国際本部では、2014 年に、eTIMSS の開発計画を公にし、開発及び検証のサイクルを回し始めた。 以降、我が国では、eTIMSS プレパイロット調査を 2017 年 3 月に実施。eTIMSS パイロット調査/項目同等性調査を 2017 年 5-6 月に実施 (紙とデジタルベースの両方の形式によりモジュールフェーズを検証)。予備調査を 2018 年 3 月に実施。

出典) 文部科学省平成31年度委託 学力調査を活用した専門的な課題分析に関する調査研究,
「学校のICT環境を活用したCBTに求められる諸条件等の調査研究 (市場調査) 」, アビームコンサルティング株式会社より文科省作成

諸外国の学力調査の実施状況（CBT形式での実施国）

	アメリカ	オーストラリア	フランス	オランダ	スウェーデン
学力調査等の名称	全米学力調査 (National Assessment of Education Progress : NAEP) : 主要評価 (main assessments : MA) と 長期的傾向評価 (long-term trend (LTT) assessments)	全国評価プログラム-読み書きと計算能力- (National Assessment Program - Literacy and Numeracy : NAPLAN)	6年生学生評価 (L'évaluation des acquis des élèves de sixième)	ファイナルテスト (eindtoets)	義務教育における全国試験 (Nationella prov i grundskolan)
対象	【MA】初等教育：4年生、中等教育：8年生、12年生 【LTA】9歳、13歳、17歳	3年、5年、7年、9年生	6年生	8年生	3、6、9年生
悉皆/抽出	抽出	悉皆	悉皆	悉皆	悉皆
頻度	MA:毎年、LTA:4年ごと	毎年	毎年	毎年	毎年
実施期間	3カ月	2週間	2週間程度	2～3日	任意
教科	主にリーディング、ライティング、数学、科学	リーディング、ライティング、言語慣習 (例：スペリング、文法、発音)、計算能力	フランス語：読解力、口頭理解、言語知識 算数：数字の知識、計算・問題解決力、幾何学・測定	必須科目：算数と言語 業者により、追加科目が設けられている	言語：読解力、リスニング、綴り、言語ケア、語彙 数学：数学、演算、計画診断 社会的感情機能
実施主体	全米教育統計センター (NCES)	オーストラリア・カリキュラム評価報告機構 (ACARA)	—	オランダ政府 ※テストの選択は学校	オランダ政府 ※テストの選択は学校
開始年	1969年	2008年	1989年	—	—
受験者数	サンプルは評価の目的によって異なる。国全体の結果が必要となる評価の場合、サンプルとなる生徒数は10,000-20,000人。州やTUDA (Trial Urban District Assessment)の結果のみ必要となる評価の場合は、各管轄区域の役100校から約3,000人の生徒が含まれる。	3年生：301,378人 5年生：304,231人 7年生：284,723人 9年生：260,851人 ※2018年の結果。各教科により若干異なるため、リーディングの数を記載	2017年は7,100校の6年生830,000名が対象であり、受験率は98%。	—	—

概要

	アメリカ	オーストラリア	フランス	オランダ	スウェーデン
CBTの範囲	NEAPの数学、読解、科学をCBT版で実施(2019年1～3月)。少数の学生は紙ベースで実施。	一部の学校のみ NAPLAN Online に参加。	全部	CBT版の全国試験を提供する事業者を選択した場合、全科目が CBT となる	9年生、高等学校および高等学校レベルの成人教育の第二言語としての英語、スウェーデン語、スウェーデン語の試験におけるEッセイ
コンピュータ対応型テスト CAT	×	●	●	●	×
機器の準備	キーボード、スタイラス、イヤホン、管理者用タブレット、およびデバイスが通信するためのクロードワイヤレスネットワークを提供するルーターを備えた学生用タブレットを含む、必要なすべての機器を提供する。	NAPLAN Online への移行は、学校に適切なハードウェアとインターネットへの信頼できるアクセスがあることを前提としている。	義務教育(16歳まで)におけるコンピュータ導入率は、2019年度時点で、未だ少数派ではあるが、一人一台が公立中学、高校で進められている。校内のコンピュータ・ルームに1クラスの人数分(およそ30台)設置されたコンピュータに、1クラスずつ入れ替わりでテストを受ける例も多いものと考えられる。	—	—
委託事業者(試験の運営等)	ピアソン社 ※ピアソン社が採点資料の作成と採点の他、すべての評価資料の作成、梱包、配布を実施。 2001年にデジタルベースの評価(DBA)への移行をサポートするための調査を開始。2008年から2015年にかけて、様々な評価にテクノロジを組み込み、新しいフレームワークを開発。2015年には移行を開始。2016年に数学と読解を試験的に実施。その後、全教科の移行に向けた開発を実施。	ピアソン社 ※クイーンズランド州は、ピアソン以外の事業者が受託。	—	4事業者 (CVTE, VISION, Diataal, AMN)	DigiExam 社
研究開発期間		オーストラリア・カリキュラム評価報告機構(ACARA)が2012年から2015年間に、NAPLAN Online に向けた開発および設計研究を実施。オーストラリア政府の教育訓練省によって資金提供された。オープンベース(選択制)で2018年よりオンライン形式に移行。	2014年2月17日制定法で、国民教育省の中に教育デジタル局が設けられ、国の義務・高等教育課程でのデジタル・トランスフォーメーション事業を推進する管轄部署が明確化された。また、全国一斉学力テストのデジタル化も決定された。	同左	デジタルベースの試験は、2018年から2021年の間に、100校を抽出してパイロットプロジェクトとして実施。100校には国から助成金を支給。

デジタルベースの評価の概要

	アメリカ	オーストラリア	フランス	オランダ	スウェーデン
測定指標	各試験受験者の平均試験スコア、特定の達成レベル以上の生徒の割合。	NAPLANの結果は、5つの国家達成スケールを使用して報告される。	個人ごとの成績は数値ではなく、5段階評価のみが与えられる	試験により、各生徒のスコア、達成パーセンテージ、レベルが提供される	-
学生単位	×	●	●	●	●
学校単位	×	●	×	●	×
地域/国単位	●	●	●	×	●
結果の比較	州間および大都市間の比較	生徒、学校、州/地域の比較。メディアは、結果に基づいて学校をランク付けした学校の「リーグテーブル」を編集する。	大学区および全国レベルの習熟度が公表される	試験により、学校の全国平均との考察が提供される	スウェーデン統計局(SCB)が情報を収集し、国のデータとして公表
主な目的	米国の公私立学校の生徒の学力の継続的な評価	政府、教育当局、学校、地域社会が、オーストラリアの若者が重要な教育成果を達成しているかどうかを判断する	教員が自身のクラスを評価すること、地域の政策を評価すること、国全体の取組を評価すること	生徒に適した中等教育の検討	全国レベルでの公平な評価と評定のサポート
結果のレポート					

OECD 生徒の学習到達度調査 (PISA2018) ICT活用調査

ICT活用調査

生徒に、携帯電話、デスクトップ/タブレット型コンピュータ、スマートフォン、ゲーム機など、様々なデジタル機器の利用状況について尋ねた調査。

学校外のインターネットの利用について

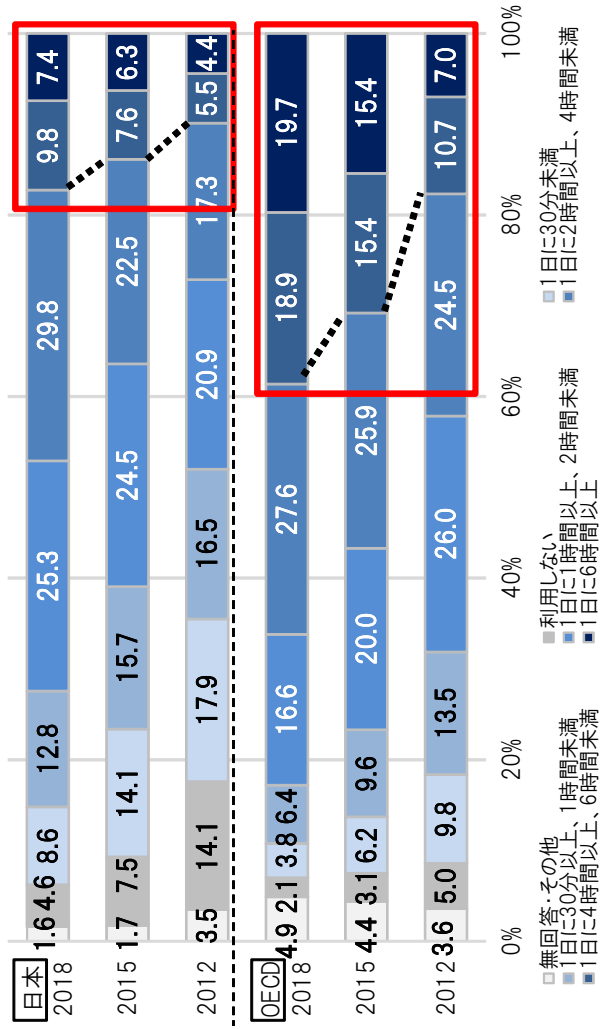
◆利用時間の推移

- 日本、OECD平均ともに、平日、学校外でインターネットを4時間以上利用する生徒が増えている。
- なお、4時間以上利用する生徒の割合を比較すると、日本は、OECD平均より少ない。

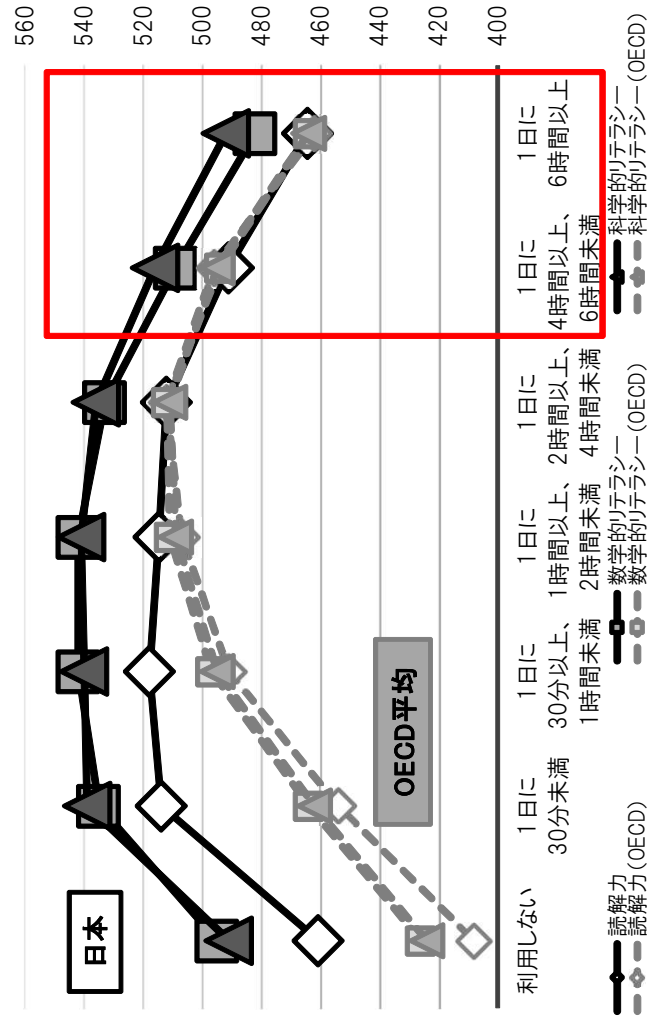
◆利用時間と3分野の平均得点の関係

- 日本、OECD平均ともに、学校外でのインターネットの利用時間が4時間以上になると、3分野ともに平均得点が低下。
- 一方、4時間未満の利用について見ると、日本は30分以上4時間未満利用する生徒の3分野の平均得点はほとんど差がないが、OECD平均は利用する時間が長いほど平均得点は高くなる傾向がある。

● 学校外での平日のインターネットの利用時間(経年変化)



● 学校外での平日のインターネットの利用時間別の3分野の平均得点



学校・学校外でのデジタル機器の利用状況

◆日本は学校の授業(国語、数学、理科)におけるデジタル機器の利用時間が短く、OECD加盟国中最下位。

「利用しない」と答えた生徒の割合は約80%に及び、OECD加盟国中で最も多い。

◆日本は、他のOECD加盟国と同様、学校外で多様な用途にデジタル機器を利用している。

○他国と比較して、ネット上でのチャットやゲーム(1人用ゲーム・多人数オンラインゲーム)を利用する頻度の高い生徒の割合が高く、かつその増加の程度が著しい。

・「毎日」「ほぼ毎日」利用すると回答した生徒の割合の増加の程度(2012年調査との比較)

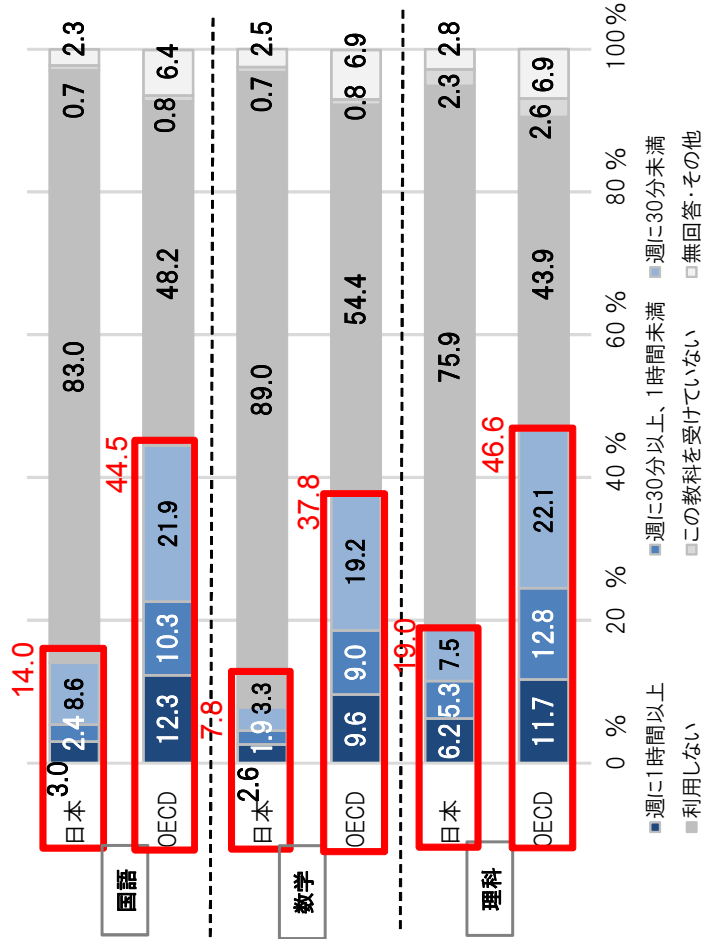
・「ネット上でチャットをする」：日本60.5ポイント増、OECD平均15.4ポイント増

・「1人用ゲームで遊ぶ」：日本21.3ポイント増、OECD平均7.1ポイント増

・「多人数オンラインゲームで遊ぶ」：日本19.4ポイント増、OECD平均7.9ポイント増

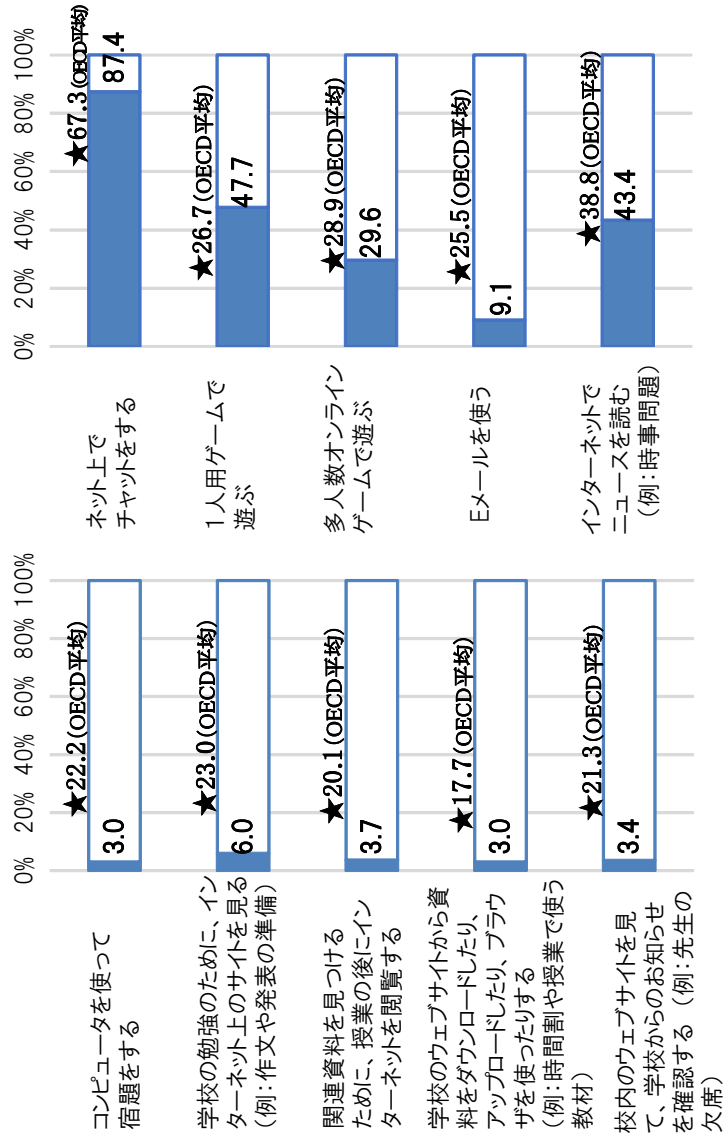
○コンピュータを使って宿題をする頻度がOECD加盟国中最下位。

● 1週間のうち、教室の授業でデジタル機器を利用する時間



● 学校外での平日のデジタル機器の利用状況

(青色帯は日本の、★はOECD平均の「毎日」「ほぼ毎日」の合計)



情報活用能力調査（小・中・高等学校）

調査概要

【趣旨】 児童生徒の情報活用能力の実態の把握，情報活用能力育成に向けた施策の展開，学習指導の改善，教育課程の検討のための基礎資料を得る。

【調査方法】 児童生徒の情報活用能力の実現状況に関する調査を、コンピュータを使って実施。

調査結果概要

	対象学年・人数	調査時期	調査時間
小学校	第5学年（116校 3,343人）	H25.10～H26.1	45分×2
中学校	第2学年（104校 3,338人）		50分×2
高等学校	第2学年（135学科 4,552人）	H27.12～H28.3	50分×2

	できたこと	課題	キーボードによる文字入力数
小学校	○ 整理された情報を読み取ること	▲ 複数のウェブページから目的に応じて、特定の情報を見つけ出し、関連付けること ▲ 情報を整理し、解釈すること ▲ 受け手の状況に応じて情報発信すること	5.9文字/分 ※ 小学校は、中・高と入力文字及び実施時間が異なるため、参考値
中学校	○ 整理された情報を読み取ること ○ 一覧表示された情報を整理・解釈すること	▲ 複数のウェブページから目的に応じて、特定の情報を見つけ出し、関連付けること ▲ 複数のウェブページの情報を整理・解釈すること ▲ 受け手の状況等に応じて情報発信すること	15.6文字/分
高等学校	○ 整理された情報を読み取ること ○ 少ない階層からなるウェブページの情報を整理・解釈すること	▲ 複数の情報がある多くの階層からなるウェブページから、目的に応じて特定の情報を見つけ出し、関連付けること ▲ 複数の統計情報を条件に合わせて整理し、それらを根拠として意見を表現すること ▲ ある事象の原因や傾向を推測するために、どのような情報が必要であるかを明確にすること ▲ 多項目かつ桁数の多い数値のある表で示された統計情報を、表計算アプリケーションを使って数的な処理をすること	24.7文字/分

上位の学校群の特徴（小・中学校調査）

- ① 上位の学校群の教員は、下位の学校群と比べ、次のような授業の実施頻度が高い傾向にある。
- ・ 児童生徒に自分の考えを表現させること
 - ・ 児童生徒に情報を整理させること
 - ・ 児童生徒に情報手段の特性に応じた伝達及び円滑なコミュニケーションを行わせること など

② 上位の学校群の児童生徒は、下位の学校群と比べ、学校で次のようなICT活用をしている頻度が高い傾向にある。

- ・ 情報を収集すること
- ・ 表やグラフを作成すること
- ・ 発表するためのスライドや資料を作成すること

生徒質問紙調査から見える傾向（高等学校調査）

課題や問題点を解決しようとする場合に、「関連付け」、「取捨選択」、「優先順位付け」、「振り返り」といったメタ認知的方略（※）を取る生徒ほど得点が高い。

※ 「メタ認知的方略」自己の認知活動を意識的にモニターしたりコントロールしたりする方略

全国学力・学習状況調査の再編の方向性について

2021年3月22日

耳塚寛明 全国的な学力調査に関する専門家会議・座長（青山学院大学・コミュニティ人間科学部・学部特任教授（お茶の水女子大学名誉教授））

（文部科学省 第8回全国的な学力調査に関する専門家会議 配付資料）

（0）はじめに

- 1 （参考資料2を参照）「全国的な学力調査のC B T化ワーキンググループ」（以下、C B T化WG）の中間まとめ「論点整理」は、「全国学力・学習状況調査（以下、全国学調）において、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、などの調査の目的と、引き続き一斉実施で行うか、調査対象は悉皆か抽出か（中略）などの調査の実施方法などは、相互に他のあり方に大きな影響を及ぼす面がある」ため、「国として何のために調査を行うかという調査目的の整理が不可欠」であるとし、親会議である本専門家会議において議論を行うことを求めた。そこでは、併せて、国と地方自治体を実施している学力調査の役割分担等も含めて、総合的に全国学調のあり方を検討することを求めた。
- 2 私の発表は、直接的には、このC B T化WGからの要請に基づいた議論を、この専門家会議で行うために、全国学調の再編の方向性について整理したものである。
- 3 調査の目的や、国と自治体の役割分担を明確にすることは、C B T化を進めるためだけに必要になったというわけではない。そうではなくて、全国学調が、設計され、導入された当初から内包されていた課題であった。（注1）
- 4 全国学調の目的は次のようなものとされている。
 - ・義務教育の機会均等とその水準の維持向上の観点から、全国的な児童生徒の学力や学習状況を把握・分析し、教育施策の成果と課題を検証し、その改善を図る。
 - ・学校における児童生徒への教育指導の充実や学習状況の改善等に役立てる。
 - ・そのような取組を通じて、教育に関する継続的な検証改善サイクルを確立する。
- 5 国としてどのような目的の調査が必要であるのかを明確にし、あわせて国と地方自治体を実施する学力調査の役割分担のあり方についても整理を行うことにする。調査の目的が明確にされた後に、その目的を達成するためにどんな調査実施方法等が設計されるべきか（方法等の最適化）が今後さらに検討される必要がある。

（1）全国学調の成果

- 1 全国学調の十余年におよぶ実施は、主に次のような成果を生んできたと考えられる。
 - ・P D C Aサイクルの浸透 教育の成果をデータで確認し、なぜそのような結果が生じたのか分析を行い、改善を図ろうとする。指導改善面に関して、学校や教育委員会で一定程度浸透していることが質問紙調査の結果からも明らかになっている。国がこの調査を教育施策の検証のために十分に活用していくことが必要。
 - ・B問題のメッセージ性 B問題を中心に、問題が提示されたことを通じて、新しい学力に関するメッセージが現場に届けられた。質問紙調査の結果からも、問題が提示されたことで指導を改善する動きが進んでいることが明らかになっている。

- ・ **都道府県間の相対的な学力差の縮小** ほとんどの都道府県が±5ポイントの範囲内で平均正答率との相対的な差が縮まってきている。
- ・ **本体調査とは別に設計・実施された補完調査** IRT（項目反応理論）を採用している経年変化分析調査は、現状、「補完調査」という位置づけであるが、将来的に国の学力調査事業の柱の一つになるもの。保護者調査は、社会経済的背景等による学力の格差をモニターするために重要。経年変化分析調査と保護者調査を組み合わせ、今後3年に1度程度実施していくことが決定していることは、評価されてよい。

（2）全国学調（本体調査）の課題

- 1 （国にとって）全国的な学力の水準と格差について、時系列的な変化をモニターすることができない。
- 2 教科が限られ、また問題数が少なすぎて、教育課程の全体をカバーできない。
- 3 （指導改善）個々の子どもの学力の変化を観察し、結果返却に一定期間かかっていることなど、結果を指導に速やかに活用したり、また継続的な指導改善を行うために、最適の仕組みにはなっていない。
- 4 （CBT化）同日一斉実施など、現在のままの調査設計では、技術的観点からCBT化のハードルが高い。
- 5 一部の自治体で調査の直前にテスト対策が過熱しているという報道がある。そのようなテスト対策が意味を持ちにくい調査の実施方法の模索の必要性。

（3）学力調査事業の再編の方向性

- 1 国として、どのような目的の調査が必要であるのかを整理し、課題を本体調査だけで対応しようとするのではなく、その調査ごとに適切に役割分担をし、最適な方法を設計することが肝要である。
- 2 この観点から、全国学調のこれまでの成果や課題、全国学調への地方からの期待を踏まえると、全国学調は、以下のような目的を持った「二本柱」の調査へと再編される必要があるのではないか。
 - ① 第一に、時系列的な学力の水準とばらつきを、国レベルでモニターするための調査
 現在は、本体調査を補完するための調査として実施されている経年変化分析調査（保護者調査を含む）を、国として実施すべき主要な学力調査として位置づけ直す。IRT（項目反応理論）を採用し国際水準の調査として、国レベルで時系列的な学力の変化を正確に観察することを目的とする。
 この目的からは、標本抽出調査で十分であろう。調査規模の拡大にともなって非標本誤差も大きくなるので、むしろ抽出調査のほうが正確である。保護者調査を合わせて実施することにより、社会経済的背景による学力の格差もモニターできる。毎年実施する必要はなく、3年に一度程度でよい。分冊方式により、教育課程の幅広い領域をカバー

することが可能である。

調査の規模から考えて、CBT化を先行して実施可能であろう。また、CBT化に伴いIRTの手法については、国際学力調査の動向も踏まえつつ、より効率的な測定の方法について検討を進めてはどうか。

②第二に、地方や学校に対して、学習指導要領の理念や目標、内容等を具体的に示し、児童生徒の学習指導の改善や、地方の教育施策の検証に資することを目的とした悉皆調査

この目的に沿った国による悉皆調査へのニーズは、「質の高い問題による悉皆調査が必要」（京都府教育委員会）との指摘や、「引き続き教育現場に対し、身に付けるべき力についてのメッセージを出し続けていただきたい」（戸田市教育委員会）という声もあり相当程度大きいものと考えられる。

現行の全国学調の本体調査を基礎に、CBT化を進めることとし、こうした期待に応える観点から、調査問題の公表の可否など調査設計について、さらに専門的に検討を行っていく必要がある。その際、先進的な学力調査を進めている地方自治体など、地方も参画しつつ、パネルデータの蒐集など学力調査全体のグランドデザインを意識し、調査の設計に当たることが望ましい。

3 なお、上記の二本柱のほかに、③国際学力調査（PISAやTIMSS）には継続して参加することが望ましい。それぞれの調査における評価の枠組みに基づいて、諸外国と比較することは、日本の学力の水準と質的特徴を知る上で、有用である。また④政策課題の出来に備えて、特定の課題に特化したトピック調査を実施する余力を、国は担保しておくことが望ましい。

4 以上は国として、国レベルでの学力の水準とばらつき、その変容を把握するために実施すべき学力調査であるが、これとは別に、地方自治体が、自らの教育施策の検証に活用することを目的とし、あるいは個々の児童生徒の学力の発達等を継続的に測定して指導改善に活用することを目的とした、地方独自の調査の実施も考えられる（国として必要な支援を行うことも考えられる。）。

（4）CBT化WGへの期待

1 CBT化WGには、専門的・技術的観点から、上記①②の調査の役割・機能分担の方向性を踏まえつつ、CBT化に向けた具体的な論点整理と、今後のCBT化の進め方について検討を進めることを期待したい。

2 経年変化分析調査のCBT化はIRTや分冊方式の導入とセットで、それ自身、学力の水準とばらつきを国レベルで正確に測定するために必要なことであるが、一方で②悉皆調査のCBT化を実現するための工程の一部でもある。このことを視野に入れて、CBT化WGにおける検討を進めるよう期待したい。

（注1）（0）-3について 全国学調が始まった平成19年の前年、私は、『内外教育』「ひとこと」欄（2006年5月16日号、時事通信社）に、その趣旨の文章を寄稿している。

TIMSS 2019と全国学力・学習状況調査の比較

	TIMSS2019 国際調査 (International Study)	TIMSS2019 国内調査 (National Study)	(参考) 平成30年度 全国学力・学習状況調査
実施形態・方式	筆記型 (従来と同様)	筆記型 (従来と同様)	筆記型
調査対象 (学校)	国公立の小・中学校 290校 (小学校 148校、中学校 142校 合計約9,000名) ※層別して学校と学級を抽出	国立大学附属の中学校 (うち66校 合計約2,500名)	国公立の小・中学校 (小学校 約2万校、中学校 約1万校)
調査対象 (学年)	小学校4年生および中学校2年生 (うち1学級を抽出)	中学校2年生 (うち1学級を抽出)	小学校6年生および中学校3年生
実施教科等	算数・数学および理科 (ただし日本の学習指導要領と 完全には一致しない) 質問紙調査も実施	数学および理科 (ただし日本の学習指導要領と 完全には一致しない) 質問紙調査も実施	国語、算数・数学、および理科 (理科は3 年に一度) 質問紙調査も実施
実施期間	平成31年(2019年) 2月～3月(4週間の調査期間) のうち、対象校が希望する1日	平成31年(2019年) 2月～3月(4週間の調査期間) のうち、対象校が希望する1日	平成30年4月17日 原則として全国同日実施
調査結果の公表	令和2年(2020年)12月を予定 (調査報告書)	令和3年(2021年)頃を予定 (今後IEAと調整)	平成30年7月31日

調査対象の児童生徒の抽出は、国際的に決められた
ガイドラインに従って、参加各国の児童生徒の状況
が最もうまく描けるように行われます。

調査計画やサンプリング
等を国際本部と個別協議
して、承認されました。
(日本独自の取組です)



NIER

国立教育政策研究所

National Institute for Educational Policy Research

TIMSS 2019でのコンピュータ使用型調査

- (1) 基本的には、筆記型の調査問題と同じです。（2部構成、選択式・短答式・記述式 etc…）
- (2) コンピュータの特性を活かして、解決過程の可視化により児童生徒の思考に対する洞察を深めるとともに、**PSI (Problem Solving and Inquiry ; 問題解決と探究)** という、新類型の問題が加わりました。

PSIの具体は次のスライド
- (3) 当初の計画では、①タブレットを使用 ②手書き入力可能でしたが、最終的に、①PC・ハードウェアキーボードの使用も許容 ②手書き入力は不可という方式でTIMSS 2019の国際的な枠組みが固まりました。なお、調査設計の初期段階から教科調査のオンライン実施の予定はありませんでした。

主な原因は、国際本部におけるシステムの開発が予定どおりに運ばなかったことと推測されます。
- (4) 上記の国際的な枠組みのもと、日本では、学校備品**PC**とTIMSS事務局が手配した**タブレット**の双方で実施しました。（対象校に対し、調査問題のアプリケーションに必要なメモリ、解像度などの水準を示し、学校備品**PC**が水準を満たさない場合、TIMSS事務局がタブレットを手配。）
PCの場合は、調査問題のアプリケーションおよび解答データはUSBで配布・回収し、**タブレット**の場合は、調査問題のアプリケーションを搭載したタブレットを学校に配送し、調査実施後に解答データが保存されたタブレットを回収しました。

国研としては、レンタル機材の手配 ↑ 及び 実施支援員の確保 ↓ が、調査経費のうち、かなりの部分を占めることになりました。
- (5) 調査当日には、実施状況の調査と機材トラブル対応を目的として、対象校1校につき2名の**実施支援員**を派遣しました。（TIMSS事務局が手配）



NIER

国立教育政策研究所

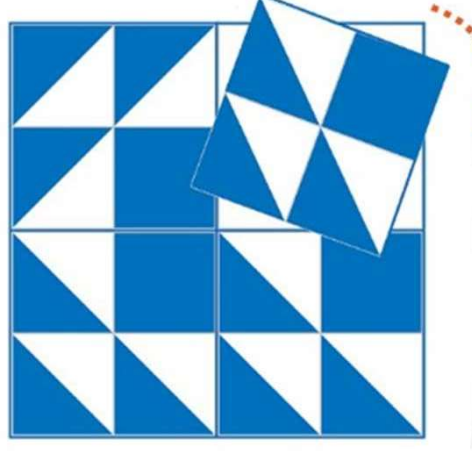
National Institute for Educational Policy Research

PSI (Problem Solving and Inquiry ; 問題解決と探究)

(1) 研究室での実験や日常生活の文脈を模した
問題状況

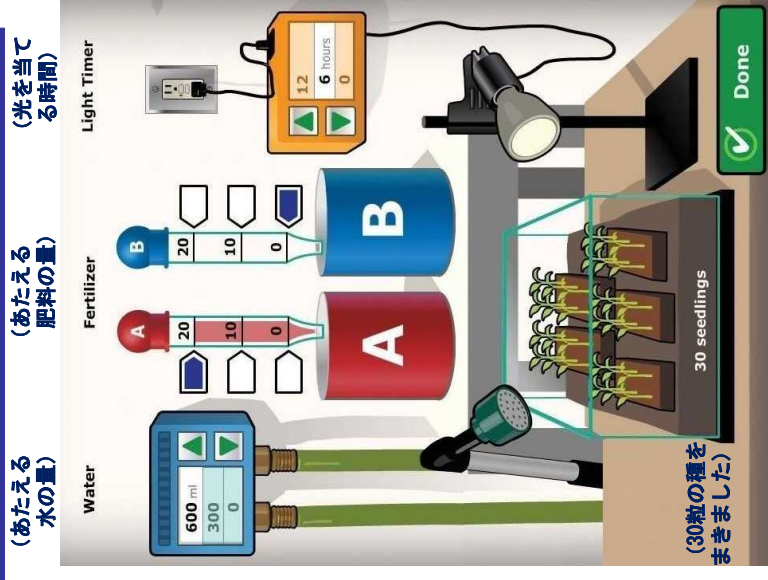
(2) 児童生徒の解答に応じて、次の問いが変化
(従来の筆記型調査では困難な方式)

(3) カラフルで動画を含み、
視覚的にも魅力的



パーツを回転させて図形
の対称性について考える

(※) 画像は、TIMSS 2019の
英文パンフレットより抜粋



植物の成長実験を計画し、
結果について考察する

PISA調査の実施方法

- 義務教育修了段階である15歳の生徒が対象
- 抽出調査。日本の縮図となるように生徒をサンプリング
- 学習到達度テスト(読解リテラシー・数学的リテラシー・科学的リテラシー)と質問調査を実施
- IRTを活用した調査設計
- 2015年～CBT調査に移行

学習到達度テストの結果

- 日本の15歳の生徒(母集団)から一部の生徒を標本として選ぶ
- その生徒が、問題に取り組んだ正誤等の解答情報を用いて、統計モデルを使い各生徒の得点を推定
→単なる得点ではない。

IRTモデルに基づき推定された能力の値

- 推定された生徒の能力値から日本の15歳児の生徒全体の学力を推定
- テストの結果は、各国の15歳の生徒全体の平均得点として示す

IRTを用いた調査設計

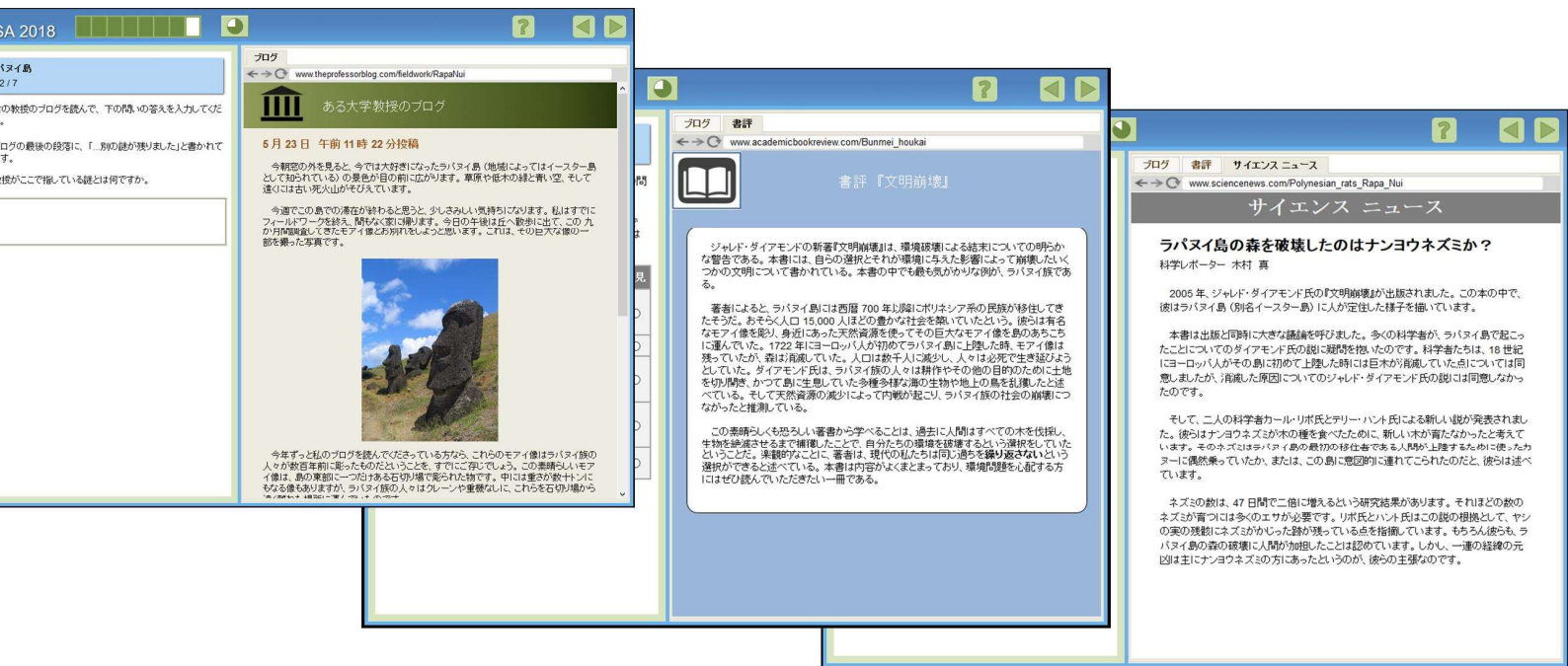
• 重複テスト分冊方式

- 四つ以上の大問（課題文と小問）をまとめて問題群を作成。
- 問題群の解答時間は30分を想定。
- 問題群の組み合わせを変えることにより異なるテスト（例：2018年36種類）を作成
- 生徒は予め割り当てられた1種類のテストに2時間かけて取り組む

調査問題数の増加

- 多様な課題文
- 筆記型調査時に比べ、測定したい内容をより広く、そして細かくカバーすることを目指す
- 高い/低い習熟度レベルの小問数の増加→測定の精度が高まる
- 筆記型調査時に比べ中心分野とマイナー分野との測定精度の格差が減少

2018年「読解リテラシー」公開問題 (ラパヌイ島)



PISA 2018
?

ラパヌイ島
問 1 / 7

右の教授のブログを読んで、下の問いの答えを一つクリックしてください。

ブログによると、教授がフィールドワークを始めたのはいつですか。

1990 年代
 九か月前
 一年前
 五月の始め

ブログ


www.theprofessorblog.com/fieldwork/RapaNui

ある大学教授のブログ

5月23日 午前11時22分投稿

今朝窓の外を見ると、今では大好きになったラパヌイ島(地域によってはイースター島として知られている)の景色が目の前に広がります。草原や低木の緑と青い空、そして遠くには古い死火山がそびえています。

今週でこの島での滞在が終わると思うと、少しさみしい気持ちになります。私はすでにフィールドワークを終え、間もなく家に帰ります。今日の午後は丘へ散歩に出て、この九か月間調査してきたモアイ像とお別れをしようと思います。これは、その巨大な像の一部を撮った写真です。



今年ずっと私のブログを読んでくださっている方なら、これらのモアイ像はラパヌイ族の人々が数百年前に彫ったものだということを、すでにご存じでしょう。この素晴らしいモアイ像は、島の東部一帯にだけある石切り場で彫られた物です。中には重さが数十トンにもなる像もありますが、ラパヌイ族の人々はクレーンや重機なしに、これらを石切り場から

読解プロセス「情報を探し出す」
 出題形式「多肢選択」
 課題文の特徴「連続した叙述。複数のテキスト」
 難易度「559点(習熟度レベル4)」
 正答「九か月前」
 正答率「日本42.4%(OECD 48.4%)」

PISA 2018

ラバヌイ島 問3/7

右の『文明崩壊』の書評を読んで、下の表の中から、次の問いの答えをクリックしてください。

下の表のそれぞれの文は、書評『文明崩壊』からの抜粋です。これらは事実または意見のどちらですか。「事実」または「意見」のどちらかをクリックしてください。

下の文は事実または意見のどちらですか。	事実	意見
本書には、自らの選択とそれが環境に与えた影響によって崩壊したいくつかの文明について書かれている。	<input type="radio"/>	<input type="radio"/>
中でも最も気がかりな例が、ラバヌイ族である。	<input type="radio"/>	<input type="radio"/>
彼らは有名なモアイ像を彫り、身近にあった天然資源を使ってその巨大なモアイ像を島のあちこちに運んでいた。	<input type="radio"/>	<input type="radio"/>
1722年にヨーロッパ人が初めてラバヌイ島に上陸した時、モアイ像は残っていたが、森は消滅していた。	<input type="radio"/>	<input type="radio"/>
本書は内容がよくまとまっており、環境問題を心配する方にはぜひ読んでいただきたい一冊である。	<input type="radio"/>	<input type="radio"/>

ブログ 書評

www.academicbookreview.com/Bunmei_houkai

書評『文明崩壊』

ジャレド・ダイヤモンドの新著『文明崩壊』は、環境破壊による結末についての明らかな警告である。本書には、自らの選択とそれが環境に与えた影響によって崩壊したいくつかの文明について書かれている。本書の中でも最も気がかりな例が、ラバヌイ族である。

著者によると、ラバヌイ島には西暦700年以降にポリネシア系の民族が移住してきたそうだ。おそらく人口15,000人ほどの豊かな社会を築いていたという。彼らは有名なモアイ像を彫り、身近にあった天然資源を使ってその巨大なモアイ像を島のあちこちに運んでいた。1722年にヨーロッパ人が初めてラバヌイ島に上陸した時、モアイ像は残っていたが、森は消滅していた。人口は数千人に減少し、人々は必死で生き延びようとしていた。ダイヤモンド氏は、ラバヌイ族の人々は耕作やその他の目的のために土地を切り開き、かつて島に生息していた多種多様な海の生物や地上の鳥を乱獲したと述べている。そして天然資源の減少によって内戦が起こり、ラバヌイ族の社会の崩壊につながったと推測している。

この素晴らしい恐ろしい著書から学ぶことは、過去に人間はすべての木を伐採し、生物を絶滅させるまで捕獲したこと、自分たちの環境を破壊するという選択をしていたことだ。楽観的なことに、著者は、現代の私たちは同じ過ちを繰り返さないという選択ができるかと述べている。本書は内容がよくまとまっており、環境問題を心配する方にはぜひ読んでいただきたい一冊である。

読解プロセス「評価し、熟考する」
 出題形式「複合的選択肢」
 課題文の特徴「連続した叙述。複数のテキスト」
 難易度「完全正答654点（習熟度レベル5）部分正答528点（習熟度レベル3）」
 正答「上から、事実、意見、事実、事実、意見」
 正答率「日本44.5%（OECD 47.4%）」

PISA 2018

ラバヌイ島 問6/7

右のタブをクリックすると、それぞれの資料を読むことができます。

二つの説に関して、それぞれの原因とそれらに共通する結果を正しい位置にドラッグ & ドロップして、下の表を完成させてください。

二つの説

原因	結果	提唱者
		ジャレド・ダイヤモンド
		カール・リボとテリー・ハント

モアイ像は同じ石切り場で彫られた。

ラバヌイ島に同じ石切り場があった。大木が消滅した。

ナンヨウネズミが木の種を食べ、その結果新しい木が育たなかった。

ラバヌイ島の住人は、モアイ像を運ぶために天然資源が必要だった。

移住者はカヌーを使ってネズミをラバヌイ島に運んできた。

人間は耕作やその他の理由のために木を切って土地を切り開いた。

ブログ 書評 サイエンス ニュース

www.sciencenews.com/Polynesian_rats_Rapa_Nui

サイエンス ニュース

ラバヌイ島の森を破壊したのはナンヨウネズミ

科学レポーター 木村 真

2005年、ジャレド・ダイヤモンド氏の『文明崩壊』が出版されました。彼はラバヌイ島（別名イースター島）に人が定住した様子を描いていました。

本書は出版と同時に大きな議論を呼びました。多くの科学者が、ラバヌイ島の森の消失についてダイヤモンド氏の説に疑問を抱いたのです。科学者たちは、18世紀にヨーロッパ人がその島に初めて上陸した時には巨木が消滅していた点については同意しましたが、消滅した原因についてのジャレド・ダイヤモンド氏の説には同意しなかったのです。

そして、二人の科学者カール・リボ氏とテリー・ハント氏による新しい説が発表されました。彼らはナンヨウネズミが木の種を食べたために、新しい木が育たなかったと考えています。そのネズミはラバヌイ島の最初の移住者である人間が上陸するために使ったカヌーに偶然乗っていたか、または、この島に意図的に連れてこられたのだと、彼らは述べています。

ネズミの数は、47日間で二倍に増えるという研究結果があります。それほど数のネズミが育つには多くのエサが必要です。リボ氏とハント氏はこの説の根拠として、ヤシの実の残骸にネズミがかじった跡が残っている点を指摘しています。もちろん彼らも、ラバヌイ島の森の破壊に人間が加担したことは認めています。しかし、一連の経緯の元凶は主にナンヨウネズミの方であったというのが、彼らの主張なのです。

読解プロセス「情報を探し出す」
 出題形式「多肢選択」
 課題文の特徴「連続した叙述。複数のテキスト」
 難易度「559点（習熟度レベル4）」
 正答「九か月前」
 正答率「日本42.4%（OECD 48.4%）」

調査実施形態(筆記からCBTへ)

筆記型

2000年～2012年



CBT

2015年～



CBT化に向けて、
10年以上かけて
開発

主要三分野の調査をしながら、オプション調査でCBT化の準備。
小規模(2006)から始まり、少しずつ開発。現在も開発は続く。

PISAにおけるCBT化の過程

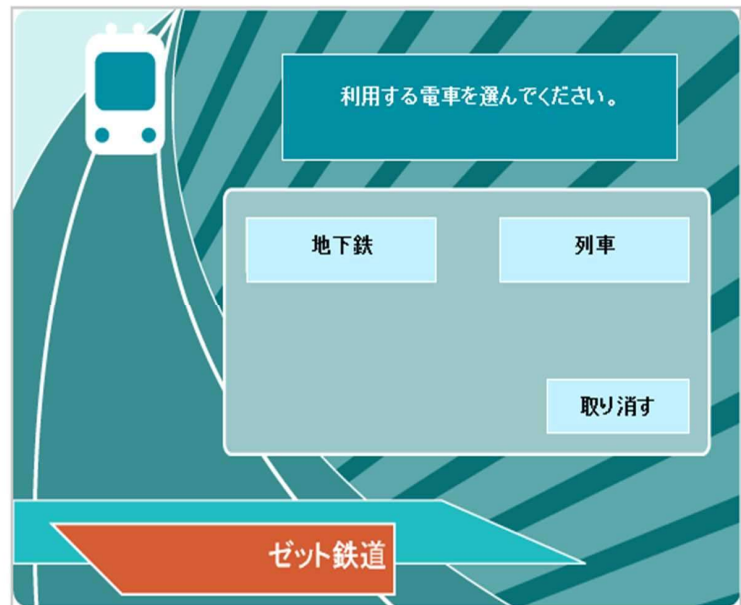
- 2003年～「CBT化の計画」
- 2006年予備調査「CBT科学」
- 2009年調査「CBT読解力」
- 2012年調査「CBT問題解決能力」「CBT読解力」「CBT数学」
- 2015年調査CBT全面実施
- 2018年調査「適応型テスト」導入

2012年「問題解決能力」問題例 (切符)


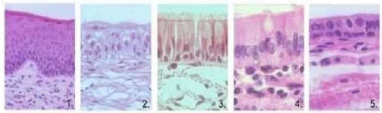
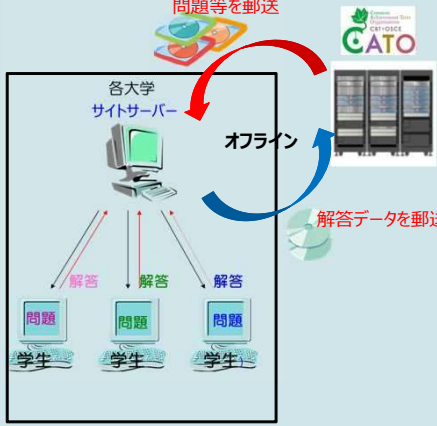
駅に自動券売機があります。切符を買うためには、右の図のタッチパネルを使って次の3つの操作を行わなければなりません。

- 利用する電車(「地下鉄」または「列車」)を選びます。
- 運賃の種類(「普通運賃」または「割引運賃」)を選びます。
- 切符の種類(「一日乗車券」または「普通乗車券」)を選びます。一日乗車券は、購入日に限り一日乗り放題になります。普通乗車券(複数枚購入できる)を買った場合は、別の日に使うこともできます。

3つの操作が完了すると「購入する」ボタンが表示されます。「購入する」ボタンを押す前であれば、いつでも「取り消す」ボタンを押すことができます。



医療系大学間共用試験 (CATO) -CBTを中心に-

実施主体	<ul style="list-style-type: none"> ・公益社団法人医療系大学間共用試験実施評価機構 (CATO = Common Achievement Tests Organization) ・医学系82大学、歯学系29大学が参加 
実施目的	<ul style="list-style-type: none"> ・臨床実習において患者の診療に参加する学生の知識、態度及び技能が臨床実習開始可能と判断できる水準に到達していることを評価する共通の標準評価試験 (令和3年5月の医師法・歯科医師法改正により令和5年度から公的化) ※知識の部分をCBTで評価
実施方式	<ul style="list-style-type: none"> ・大学ごとに個別に試験実施日を設定 (8月の第3月曜日から翌年の3月31日までの期間) し、各大学それぞれで一斉に実施。追再試験制度あり。 ・試験時間は合計6時間+受験者アンケート30分
実施規模	<ul style="list-style-type: none"> ・109医学系・歯学系大学、医学系約1万人、歯学系約3千人 (2018年実績) ・医学系1大学あたり100~140人程度、歯学系1大学あたり50~120人程度
問題セット (医学系)	<ul style="list-style-type: none"> ・IRTを利用し、受験生ごとに異なる問題セット (セット間の難易度を調整) (アダプティブではない) ・試験問題は、合計320問 (五選択肢/多選択肢/順次解答4連問五選択肢択一形式) <div style="border: 1px solid black; padding: 5px; margin: 5px 0;"> <p>240問が難易度と識別力が明らかな問題 + 80問が新規・試行問題 (採点対象外) で 項目特性を評価・判断し、問題バンクに蓄積</p> </div> <div style="margin-top: 10px;"> <p>問題例 (五選択肢択一形式)</p> <p>膀胱や尿管の内面をおおう上皮はどれか (写真はカラー表示)。</p>  <p>A. 1 B. 2 C. 3 D. 4 E. 5</p> <p>正解: B</p> </div>
作問プロセス (医学系)	<ul style="list-style-type: none"> ・問題作成マニュアルに基づき各大学が作成 →各大学から3名ずつ選出した教員でブラッシュアップ →CATOの17人の小委員会で試行問題を決定 ・問題バンクは27,000題※2018年時点
システム	<div style="display: flex; align-items: flex-start;"> <div style="flex: 1;">  </div> <div style="flex: 2; padding-left: 10px;"> <p>※実施にあたっては事前の環境確認・システム動作確認が必須。事前確認をクリアしないと実施不可。</p> <ul style="list-style-type: none"> ・CATOにセンターサーバを設置し、問題、試験実績・結果等を管理 ・CATOから実施キットや問題集 (暗号化済み) 等をCD等で郵送 ・各大学にサイトサーバを設置。試験端末への問題の配信、解答、アンケートの回収、試験実施状況の監視を実施、CATO派遣監督者が試験の公平性を確認。 ・受験生は出題された問題について審議を依頼することが可能 ・実施後に解答データ・各種ログデータ (暗号化済み) をCDに出かして、実施キットとともにCATOへ郵送 </div> </div>
結果提供	<ul style="list-style-type: none"> ・自動採点 ・解答データ受領後3日以内に各大学へ個人成績、大学成績、アンケート結果を返却
試行	<ul style="list-style-type: none"> ・H12年にシステムに関する会議体を設置。 ・H14年からトライアルを4回実施→H17年から正式実施 <div style="border: 1px solid black; padding: 5px; margin-top: 5px;"> <p>トライアル0: 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</p> <p>トライアル1: H14年 80大学7,915人 プール問題累計数: 2,305問 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</p> <p>トライアル2: H14~H15年 80大学7,827人 プール問題累計数: 5,208問 全受験生に同じ問題を出すアンカー問題を20%出題しIRTを精緻化</p> <p>トライアル3: H15~H16年 80大学8,109人 プール問題累計数: 9,083問 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</p> <p>トライアル4: H16~H17年 80大学7,836人 プール問題累計数: 10,834問以上 正式実施を想定した運用のチェック等</p> </div>



共用試験の特徴

- 国家試験を除いて全医学生・歯学生に対する全国共通の標準試験の実施は日本では初めて。
- 医学系、歯学系全大学が会員となる公益法人を設置し、試験問題の作成から試験の実施、改善までを行う。
- 試験日が大学によって異なる。
- 試験問題も受験生ごとに異なる。
- 医師法、歯科医師法とも関連。
- 進級要件、合格しないと臨床実習へ進めない。
- 到達すべき学修目標は「モデル・コア・カリキュラム」に示されている。
- 実施の準備と実施当日の運営は参加大学教職員が責任者となって行う。
- CBT, OSCEの公平性・客観性・透明性を確保するために、他大学の教員が公益社団法人医療系大学間共用試験実施評価機構から機構派遣監督者として派遣される。
- 各大学教職員の多大な協力のもと運営されている。

診療参加型臨床実習開始前の到達レベル 評価の仕組



全国の大学が共通で利用できる標準評価システム(共用試験)を用いて実施

- 知識の統合的理解・問題解決能力、臨床推論等の評価については、異なる時期・場所でも公平に実施・評価できる方式

→ コンピュータを用いた客観試験 CBT を選択

- プール問題によるランダム出題方式＝システム開発、適正な試験問題の蓄積、学生の成績と試験問題のセキュリティ確保

- 項目反応理論(IRT)の応用:問題特性の解析、出題、成績分析(能力値推測)
- 実施本部体制構築と機構試験監督の派遣

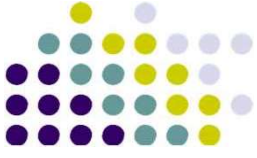
- 態度・技能については、客観的臨床能力試験OSCEで評価**

- 技能・態度に関する学習評価項目と実施課題作成
- 評価法・評価マニュアル、教育用DVD・評価用DVD作成、評価者講習会、データ等管理システム開発、模擬患者養成、全大学での実施準備
- 実施本部体制構築と機構試験監督・ST別外部評価者(6名)の派遣



試験の信頼性・公平性・客観性・透明性等を担保するために我が国の代表的な試験評価の専門家が参画する試験信頼性向上専門部会(※)を設置し、恒常的に活動 (※2020年4月から試験信頼性向上委員会)

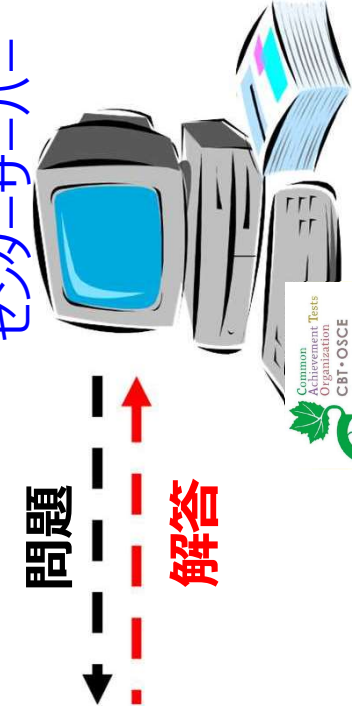
CBT (Computer Based Testing) の実際



各大学
サイトサーバー

公益社団法人
医療系大学間共用試験実施評価機構
センターサーバー

- 項目反応理論 (IRT: Item Response Theory) の利用
- 出題問題の事後評価 (項目特性曲線)
- 受験生毎に異なる出題セット間の難易度調整 (難易度の等質化)
- 受験生の能力評価 (IRT標準スコア)



問題プールから受験生ごとに異なる問題をランダムに出題。平均難易度に差がないように調整。

項目特性値を持った大量の問題をプール

2018時点
27,000題



Common Achievement Tests Organization
自動採点



試験端末(受験者毎に問題が違う)

出題数 = 320設問/学生/1日

全6ブロック: 1ブロック60分

ブロック1~4: 単純5肢択一形式60設問 1問1分

ブロック5: 多選択肢択一形式40設問 (鑑別診断、病態等) 1問1分半

ブロック6: 順次解答4連問5肢択一形式40設問 (臨床推論) 1問1分半

- 新規作成問題 (採点対象外) + プール問題 (採点対象) の組合せによる出題問題セット作成と調整作業
- CBT実施セット準備・配布
- 新規出題問題の事後評価作業
- 問題の特性評価、プール問題候補決定
- プール問題の特性変化検討
- 共用試験実施評価評価機構委員によるブラッシュアップ集中作業
- 新規出題 (試行) 問題決定
- 各大学へ問題作成依頼
- 各大学でのブラッシュアップ後に新規作成問題提出

参加大学へ成績送付

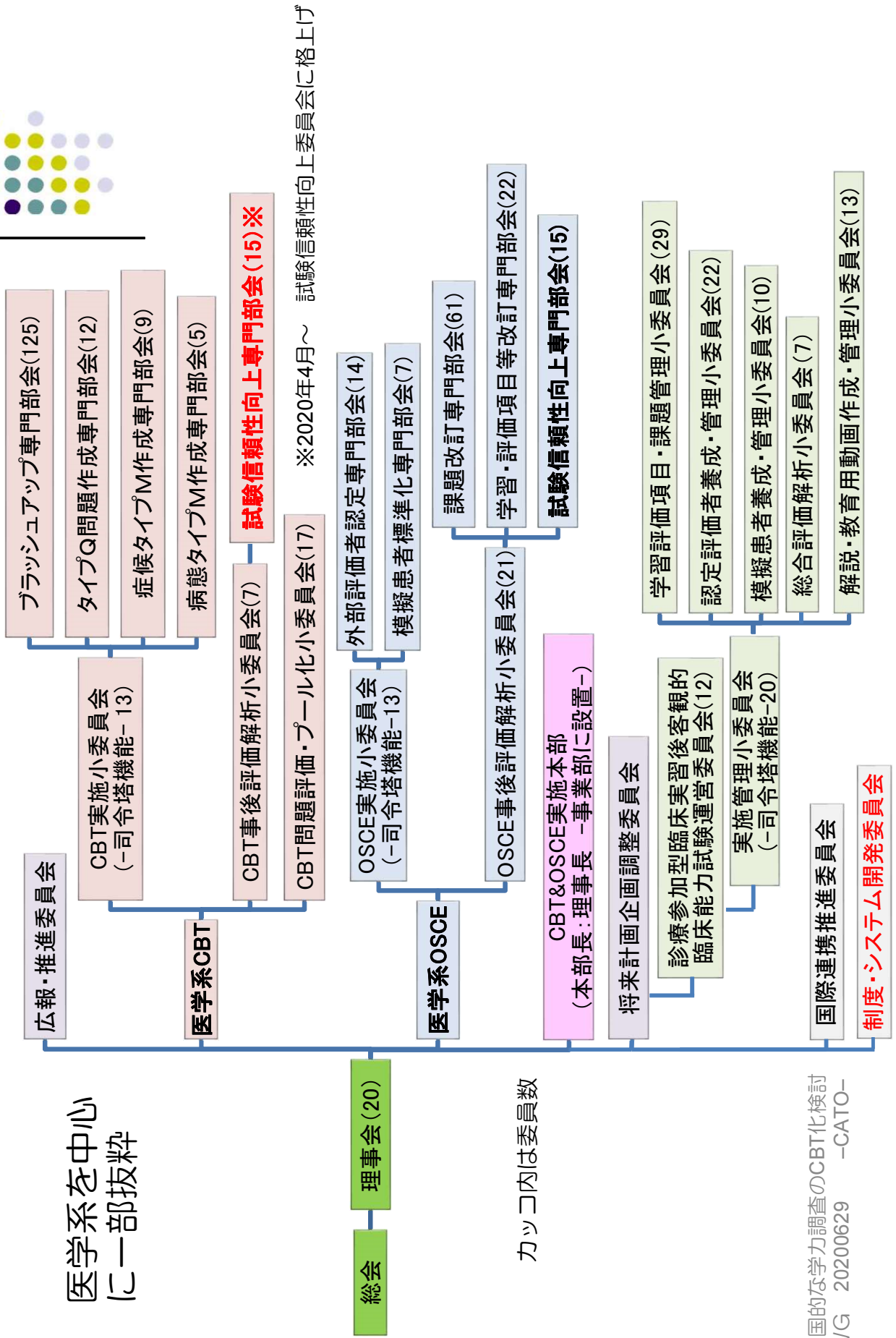
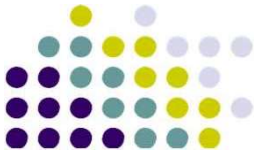
全国的な学力調査のCBT化検討WG
20200629 -CATO-



ブロックの構成

- ブロック1～4：各1時間：五選択肢択一形式で**60題**
⇒コア・カリキュラムA～D、FとE2、E3から出題
- ブロック5：1時間：多選択肢択一形式
⇒臨床症例問題：34題と基礎医学問題：6題の計**40題**
- ブロック6：1時間：順次解答4連問形式
⇒**10症例40題**
ブロック6については一度解答したら戻れない
- **全6時間で320題**

共用試験の運営を支える組織



CBT問題作成から成績返却まで

問題入力システム、問題作成テンプレート、問題作成マニュアル、ブラッシュアップチェックリスト等収載

- 各医科大学・医学部への作成依頼（計画的依頼、問題作成キット送付）
- 各医科大学・医学部でのブラッシュアップ

2018実績 221名参加
各大学に3名の推薦を依頼

- 各医科大学・医学部選出教員による中央ブラッシュアップ
- 問題評価・プール化小委員会による試行問題決定

中央ブラッシュアップの前後で機構委員によるPreブラッシュアップ、Postブラッシュアップを実施)

- 試行問題＋プール問題（採点対象）の組合せによる出題
- 問題セット作成と調整作業、総覧（実際に受験生が見るモード）
- CBT実施キット準備・配布

2018実績 75回開催

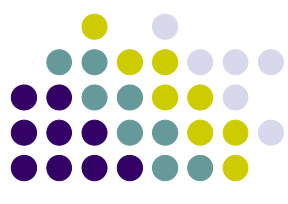
2018実績 本試験129名、
追再試験70名派遣

- 各医科大学・医学部でCBT実施（機構派遣監督者派遣）

- 各医科大学・医学部からCBT実施キット回収
- 機構内で自動採点（IRT標準スコア値推定も含む）

- 事後評価（試行問題＋プール問題）

- 各医科大学・医学部へ成績返却
- CBT解析結果の公表



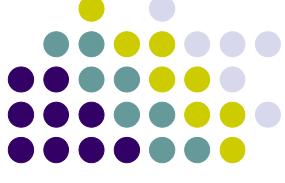
CBT問題の質の保証(1)

- 異なる時期に試験を行い、能力を評価する
 - ✓ 問題プールから受験生ごとに異なる問題をランダムに出題する
 - ✓ 出題セットの平均難易度に差がないようにする必要がある
 - ✓ **項目反応理論** (IRT: Item Response Theory)
 - ① 試行 (採点対象外) として出題した問題の評価 (**200名以上**の受験生からの項目特性曲線: 難易度と識別度)
 - ② 受験生ごとに異なる出題セット間の難易度調整 (難易度の等質化)
 - ③ 受験生の能力評価 (IRT標準スコア)

➤ 医師国家試験の場合

- ✓ 事前に問題の評価ができないために事後に評価する
- ✓ 採点除外問題
 - ① 正解が複数存在: 複数正解
 - ② 必修問題: 「問題としては適切であるが、必修問題としては不適切」
→ 正解者は正解として扱い、不正解者は削除
 - ③ 一般問題: 「選択肢の表現が不十分で正解を得ることができない」
「設問が不適切で正解が得られない」「選択肢に誤りがあり正解が得られない」
- ④ 採点除外問題: 4題 (113回)、3題 (112回)、7題 (111回)

CBT問題の質の保証(2)



▶ 共用試験CBTの質保証

- ✓ 良質な問題の作成→全医学部に問題作成依頼（説明会の開催）
- ✓ 良質な問題のプール
- ① 医学部委嘱教員による機構でのブラッシュアップ（5～6人のグループで2日間で50～60題）・・・2日間を3セット実施。各チームは必ず専門分野以外の教員も含める構成としている。
- ② 機構内「問題評価・プール化委員会」でのブラッシュアップ
- ③ 採点対象外として出題問題の事後評価（正答率、受験生からの審議依頼問題の評価）
- ✓ 採点問題のプールとセットとしての評価
- ① 出題問題の特性変化の検討→大幅に正答率が変動した問題の削除（復元本により問題の漏洩？）
- ② 問題の記載法の変更に対応→症例（外国人名、検査値など）の記載法
- ③ 受験生から審議依頼問題の検討→トライアルだけでなく採点評価問題として出題した問題にも実施
- ④ 出題セット作成前に出題可能問題の確認
- ⑤ セット組の問題の再度の確認→実施小委員による総覧を実施
- ✓ **試験信頼性向上委員会**
- ① CBTによる評価の妥当性を検討
- ② 2か月に1回開催（CBT・教育測定学・統計学・テスト理論の専門家による検討）

▶ 以上のステップにより良質で評価可能な問題が出題されている

CBT&IRTをセットで



1. CBTとIRTをセットで導入することを前提に改めて必要な要件を検討
 - 1) 試験システムの根幹
 - 非同時実施
 - 非同一問題
 - 追試験、再試験
 - セキュリティ
 - ブループリント（モデル・コア・カリキュラムに基づいた出題）
 - IRT適用（適切な露出回数確保、プール問題の確保）
 - 2) 学生の適切な能力評価のために必要な配慮
 - 問題形式（出題形式）
 - マルチメディア対応
 - アダプティブ対応
 - CBTに慣れていない学生への対応（公開問題、体験テスト、操作説明ツール）
 - 3) フィードバック（受験料を徴収して行う試験、ハイステークスな試験の観点から）
 - 受験生へのフィードバック
 - 実施大学へのフィードバック
 - 問題作成、事後評価、実施管理、プール問題管理部会等へのフィードバック
 - 行政へのフィードバック

CBT&IRTをセットで -2-



1. CBTとIRTをセットで導入することを前提に改めて必要な要件を検討

4) feasibility

- 実施大学の環境（ICTインフラ、ICT要員）
- 運用手順の確立、障害対応
- 各種マニュアル、手引き（問題作成キット、事前動作確認キット、試験実施キット）
- 受験票の設計
- 受験環境の確認
- 受験生、実施大学責任者からの実施後調査票
- 機構派遣監督者によるモニタリングシステム

必要な要件の検討を踏まえて、トリアル実施計画を策定
並行して、

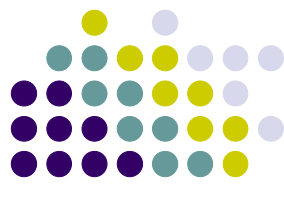
- ① トリアルサポート施設（※）の設置要請
- ② トリアルサ実施のための財政支援を要請
- ③ 正式実施後の運営母体となる社団法人設立の準備

を開始。

共用試験の意義を再確認と周知。



トライアルの経過



1. トライアル0（ゼロ）
 - 1) CBT導入可能性の検討のためのプロトタイプシステム開発
 - 100人に100問の問題をコンピュータ出題するシステムの構築（非ランダム出題）
2. トライアル1（80大学中80大学で実施 受験者数7,915人）
 - 1) 100人以上の受験生に100問コンピュータ出題するシステムの構築
 - 2ブロック制
 - タイプAのみ
 - ランダム出題（受験生毎に異なる出題）
 - 難易度調整なし
 - 素点評価
- 2) 問題作成システムの開発、各大学にシステムを配布、問題作成用PCを1大学1台配布)
- 3) 出題問題管理システムの開発
- 4) 試験実施システムの開発

トライアルの経過 -2-



3. トライアル2（80大学中80大学で実施 受験者数7,827人 追再試験30大学 518人）

1) 100人以上の受験生に300問コンピュータ出題するシステムの構築

- 6ブロック制
 - タイプA、W、Q（順次解答型連問形式）
 - ランダム出題（受験生毎に異なる出題）アンカー問題20%
 - 難易度調整なし
 - 素点評価
- 2) 追・再試験実施機能の追加
 - 3) 出題形式の多様化対応（順次回答2連問、4連問）
 - 4) 事前体験ツールの作成（体験テスト、操作説明ツール）
 - 5) 試験開始許可キーの導入
 - 6) フライング設定機能の付加
 - 7) パニックID機能の付加
 - 8) 事後評価システムの開発
 - 9) IRT適用準備

トライアルの経過 -3-



4. トライアル3（80大学中80大学で実施 受験者数8,109人 追再試験38大学 701人）

1) 100人以上の受験生に320問コンピュータ出題するシステムの構築

- 6ブロック制
- タイプA、W、Q（順次解答型連問形式）、R、L（多選択肢連問形式）
- ランダム出題（受験生毎に異なる出題）アンカー問題20%
- 素点による難易度調整
- 素点評価

2) プール問題と試行問題の区分出題機能の付加

3) 出題形式の多様化対応（多選択肢形式）

4) 重複問題排除機能の付加

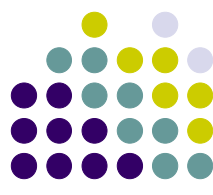
5) 視覚素材の出典管理機能付加

6) 高解像度ディスプレイ対応

7) 事後評価機能の強化（経年データ表示、累積統計）

8) 問題作成システム評価工ディシヨンの開発（修正履歴、管理情報、ブラッシュアップ機能強化）

トライアルの経過 -4-



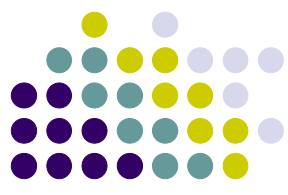
5. 最終トライアル -正式実施を想定した運用のチェック-
(80大学中80大学で実施 受験者数7,836人 追再試験41大学 612人)
 - 1) 100人以上の受験生に320問コンピュータ出題するシステムの構築
 - ・6ブロック制プラス1アンケートブロック
 - ・タイプA、Q（順次解答型連問形式）、L（多選択肢連問形式）、タイプW・R廃止
 - ・ランダム出題（受験生毎に異なる出題）アンカー問題20%
 - ・IRTによる難易度調整
 - ・素点評価、IRT評価
 - 2) 受験者固有番号の追加
 - 3) 各大学フィードバックデータの定型化（個人別成績表、結果総表、年度総表）
 - 4) アンケートブロックの導入
 - 5) サイトサーバにおけるリアルタイムモニタ機能追加
 - 6) 試験時障害対応機能の追加（スタンダードモード、予備端末移動、予備受験票対応障害時トレース機能）
 - 7) ブロック解答中、ブロック間休憩時のロック機能の追加
 - 8) 試験実施システム事前動作確認機能の追加

トライアル期間を通じたプール問題の蓄積状況



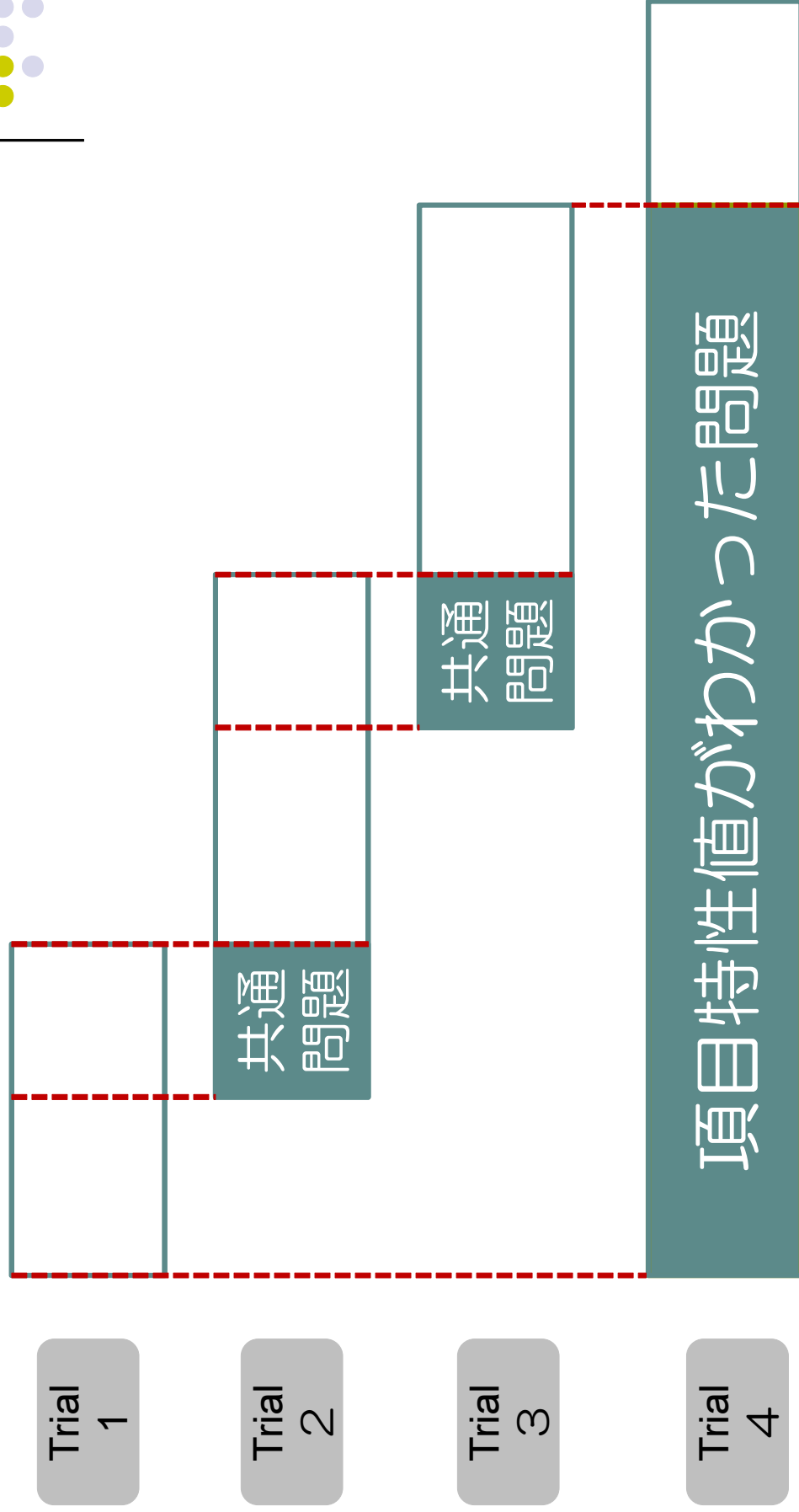
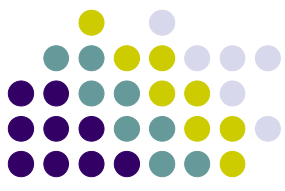
1. トライアル1
80大学からタイプA（5肢択一形式） 9,919設問提出、**トライアルを通じて、1万問を超える
プール問題（採点対象問題）を確保**
ブラッシュアップ委員会による精選作業で、2,791設問採択（採択率28.1%）
事後評価解析委員会による試験実施後の問題評価により 2,305設問をプール問題候補とした。
2. トライアル2
80大学からタイプA（5肢択一形式） 9,322設問提出
ブラッシュアップ委員会による精選作業で、3,723設問採択（採択率40.0%）
事後評価解析委員会による試験実施後の問題評価により 2,908設問をプール問題候補とした。
3. トライアル3
80大学からタイプA（5肢択一形式） 7,108設問提出
ブラッシュアップ委員会による精選作業で、3,803設問採択（採択率53.5%）
新形式問題作成チームにより、タイプW、R、Qについて合計449設問作成。
事後評価解析委員会による試験実施後の問題評価により 3,875設問をプール問題候補とした。
4. 最終トライアル
80大学からタイプA（5肢択一形式） 3,175設問提出
タイプL（多選択肢2連問形式・・・鑑別診断問題） 2,080設問
タイプQ（順次解答4連問形式・・・臨床推論問題） 1,964設問
ブラッシュアップ委員会による精選作業で、問題タイプ別にそれぞれ、
1,316設問（採択率41.4%）、712設問（採択率 34.2%）、412設問（採択率21.0%）採択
実施小委員会による新規作成、旧問題サルベージにより、タイプA259設問、タイプQ56設問作成
事後評価解析委員会による試験実施後の問題評価により 3,172設問をプール問題候補とした。

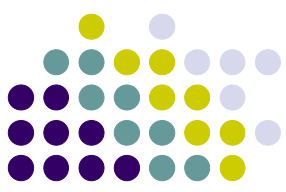
トライアル期間を通じたプール問題の蓄積



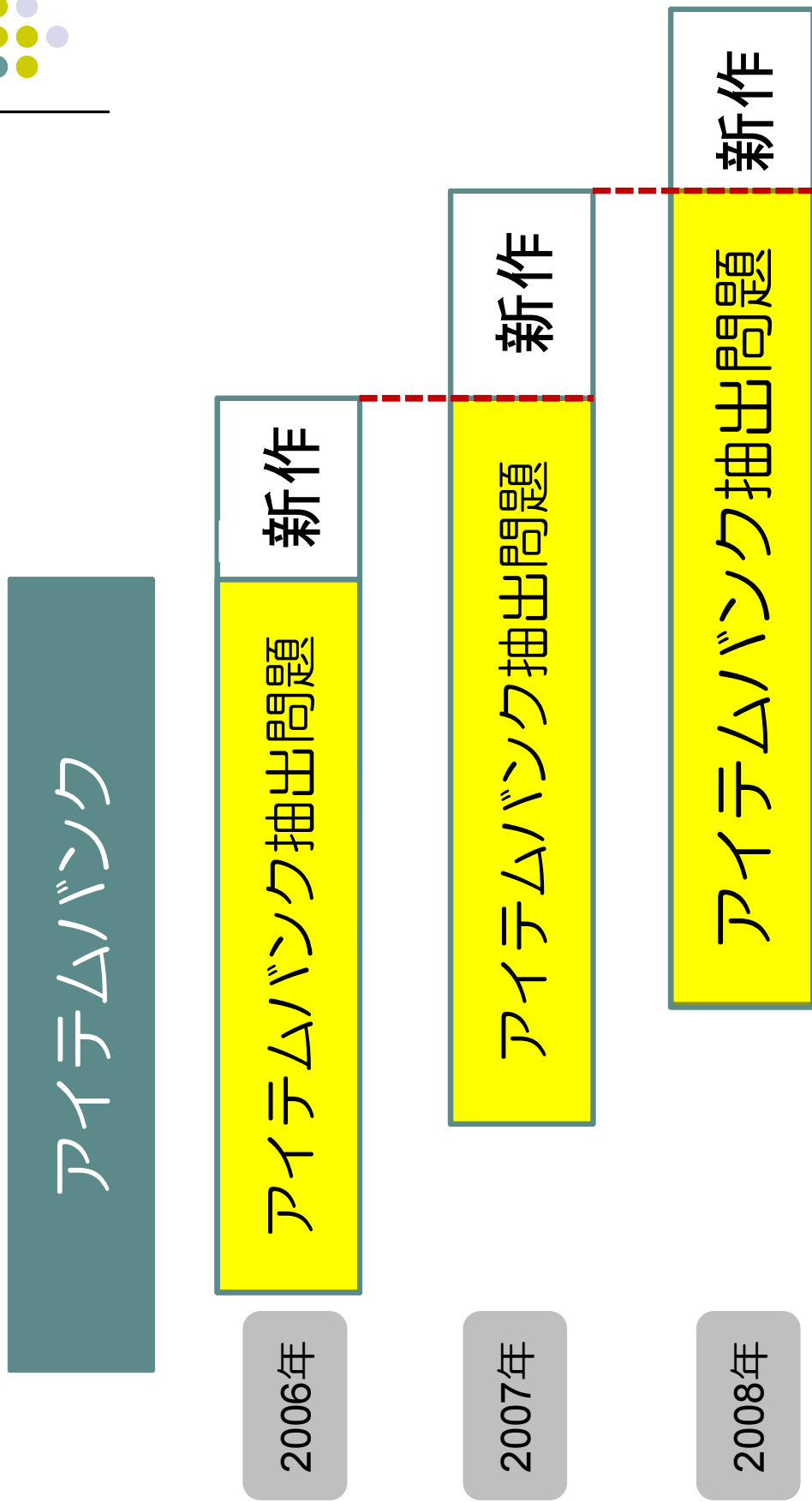
- アイテムバンクを作成し、それに基づき出題
- 毎年参加大学の教員が多くの試験問題を作成することから、他の試験に比べる
と新作の問題数が多く、良質のアイテムバンクの構築と多数の新作代謝が可能
- 新作問題の評価にはトライアルを4年間実施し、十分な問題数を確保
- 本試験ではアイテムバンクから抽出した問題と新作問題を同時に出题して新作
問題の項目特性値の評価を行う

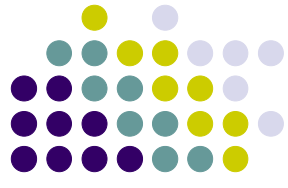
トライアルのテストデザイン





新作問題とアイテムバンク

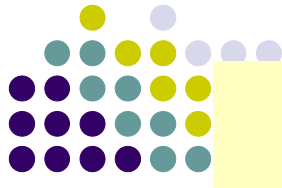




共用試験で項目反応理論が用いられる背景

- 共用試験CBTは、同一時期に一斉試験が困難なため、異なる時期に、異なる場所、異なる問題の試験を実施しても公平な評価が得られる試験方式を採用する必要がある。
- CBTでは、過去に出題した問題について、複数回の内容のチェックと解答状況の解析を行い、良質かつ適切と考えられる問題を継続的に蓄積した。
- CBT実施時には、このプールした問題バンクの中から、受験生ごとに異なった問題がランダムに抽出されて出題される。
- 異なる問題が出題されるための不公平感をなくすために、受験生一人一人に出題される問題セット間の難しさの差をできるだけ小さくするように設定する。
- この方法だけでは、一人一人のテストの難しさの差を完全になくすことはできないため、受験生に出題された項目の特性を考慮して問題セット間の難しさの差に影響されない成績評価の方法として項目反応理論を用いることとなった。

IRTの適用（CBTとセット）



1. 項目（出題問題）の事後評価

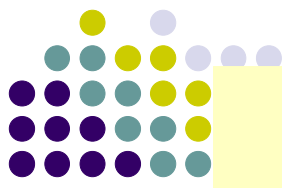
<目的>

良質かつ適切（臨床実習開始前の難易度として）と考えられる項目の継続的な蓄積

<IRTの適用>

- 項目特性曲線のパターン（急峻なパターン→識別力の視覚化）
- 項目困難度（難しさのレベルの定量化）
- 項目識別力（識別力の定量化）
- 項目特性値の標準誤差（項目特性値の推定精度の定量化）
- 基本となる項目特性値と直近の年度で推定された項目特性値の比較（項目特性値の変化状況の把握）

IRTの適用（CBTとセット）



2. 出題セットの作成

<目的>

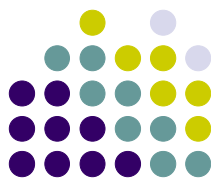
ランダム出題のため異なる問題セットが受験生に出題されることからセット間の難易度を小さくする。

<IRTの適用>

- 項目困難度によるセット間の難易度差の調整
- 情報関数を用いたセット間の難易度差と識別力差の調整



IRTの適用（CBTとセット）



3. 試験成績の評価とフィードバック

<目的>

受験生に出題された項目の特性を考慮して項目セット間の難しさの差に影響されない成績評価を行う。

<IRTの適用>

出題された問題セット間の難易度差に影響されない受験生の成績評価としてIRT標準スコアを用いる。

1) 全国的評価

- 結果表、冊子の提供：毎年の全体的成績、月別の推移、再試験受験者の本試験と再試験の比較、問題タイプ別比較、コアカリ大分類別比較など
- 講演会：経年変化の提供

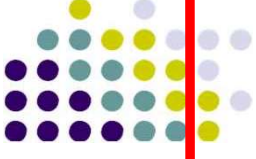
2) 個人の評価

- 個人別成績表：IRT標準スコアの個人別成績と学内分布における位置の提供、コアカリ大項目別問題タイプ別リーダーチャートによる学内平均との比較

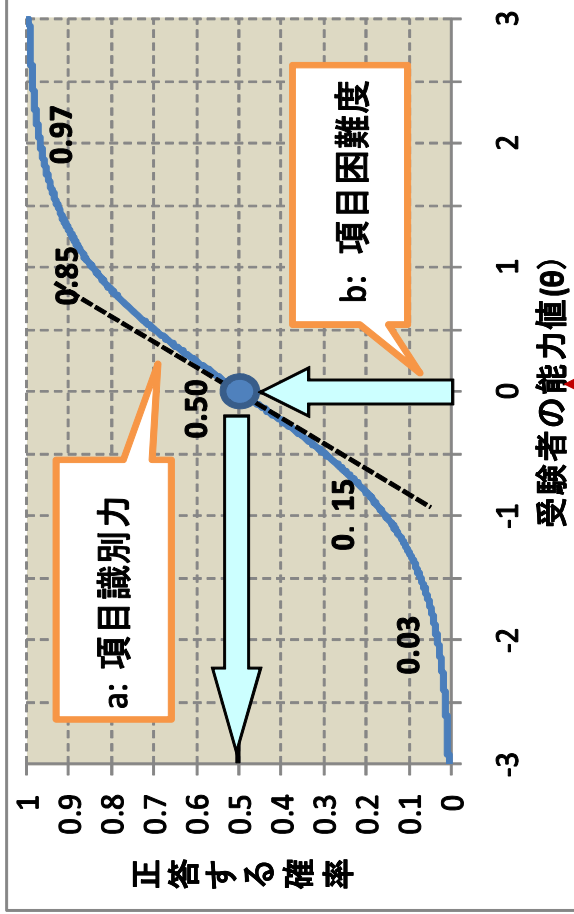
3) 学内評価

- 総表（個人別の成績一覧表）：各大学における受験生評価に利用可能な資料の提供

項目反応理論



項目特性曲線



学生の総合得点

各大学に問題作成

↓
中央のブラッシュアップ委員会にて修正等

↓
(近年の採択率≒80%)

↓
新規問題としてプレテストに出題 (採点対象外)

プレテストに出題された個々の試験問題 (項目) について、学生の能力 (総合得点) とその問題を正答する確率の関係を、2パラメータ成長曲線を用いて回帰分析し、パラメータ a , b を求める。

$$\text{正答する確率} = P_j(\theta) = \frac{1}{1 + \exp(-1.7a_j(\theta - b_j))}$$

↓
事後評価で、正答率、識別指数、点双列相関係数とともに両パラメータ値を基準として適正な問題を選んでプールの問題とする

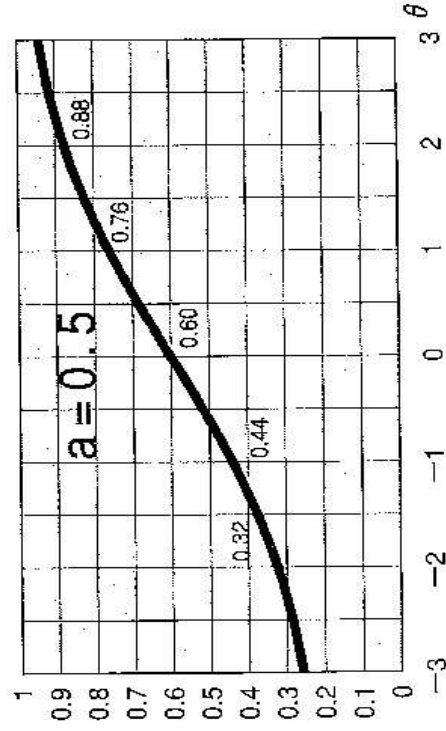
項目特性値 (項目識別力と項目難易度)



a 項目識別力

能力差のつく問題かどうかの指標

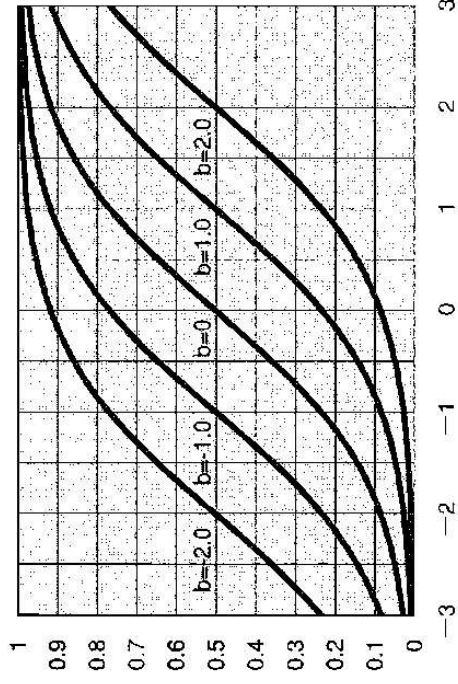
1. 特性曲線の最大の傾きを表す。
2. a が大きいほど識別力が大きい
3. 通常は0.3~2.0程度



b 項目難易度

問題の難しさを表す指標

1. その問題を正解できる確率が0.5の能力
2. b が大きいほど難しい問題
3. 多くは-2~+2の範囲程度



「従来のテスト」と「項目反応理論」の違い



	従来のテスト	項目反応理論
能力の表現方法	正答数, 正答率 など	問題の難易度・識別度からの 統計値で表現
項目の難易度	経験に基づく	事前テストのデータに基づく
事前テスト	不要	問題の難易度・識別力 推定に必要
異なるテスト間 での比較	不可能	可能



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

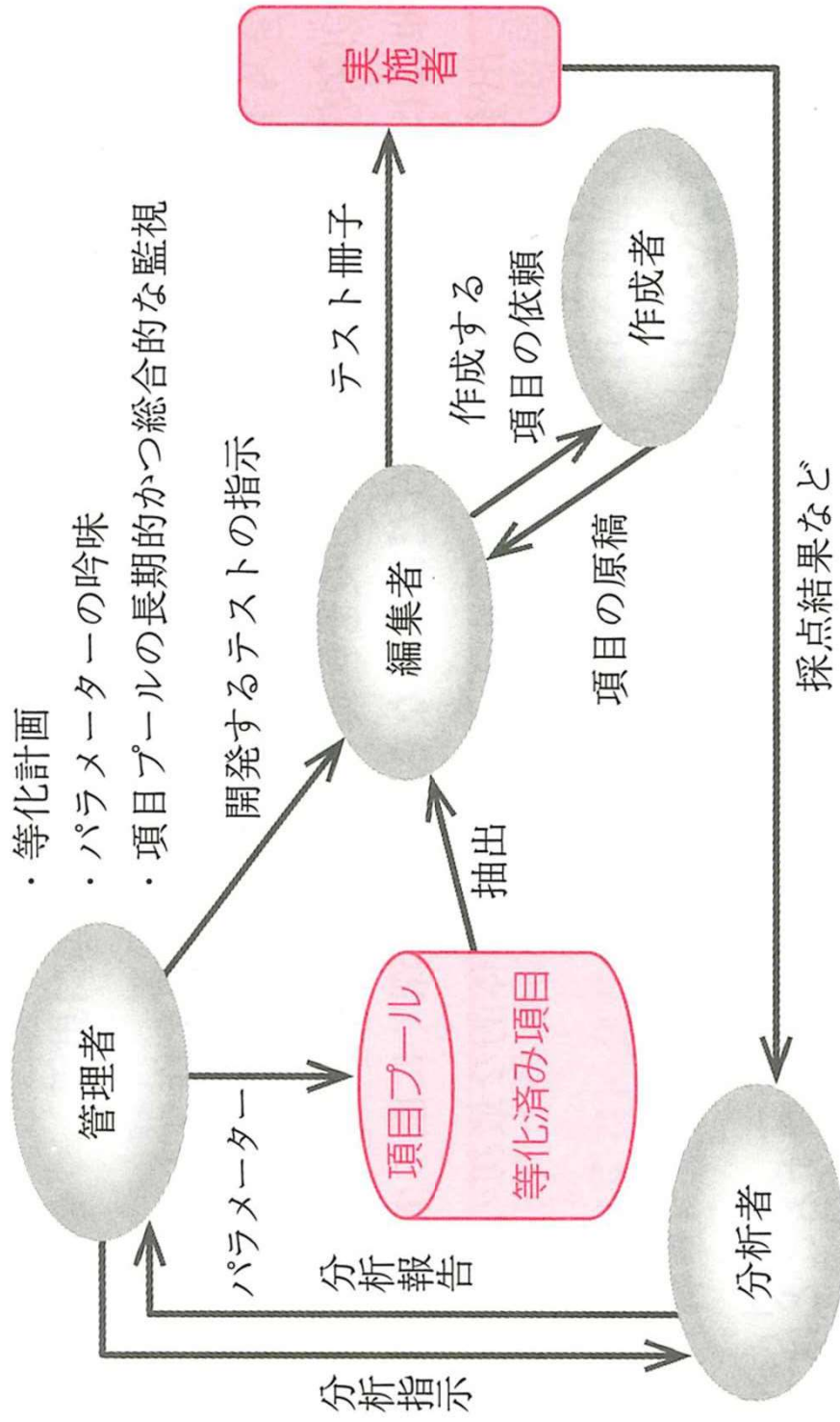
- 1) 調査を実施する目的の明確化
 - 2) 目的を達成するための計画立案*と体制整備
 - 3) 問題セットの作成
 - 4) 調査実施
 - 5) 分析・等化
 - 6) 結果の活用
 - 7) 検証
- ⇨ 3) ~ 5) を「調査の実務」と呼びます

*：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	コミュニケーション力 適度な柔軟性	・判断力 ・計画性
編集者	○	◎	△	△	○	○			・バランス感覚 ・調整力
作成者	—	—	◎	△	◎	○			・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。

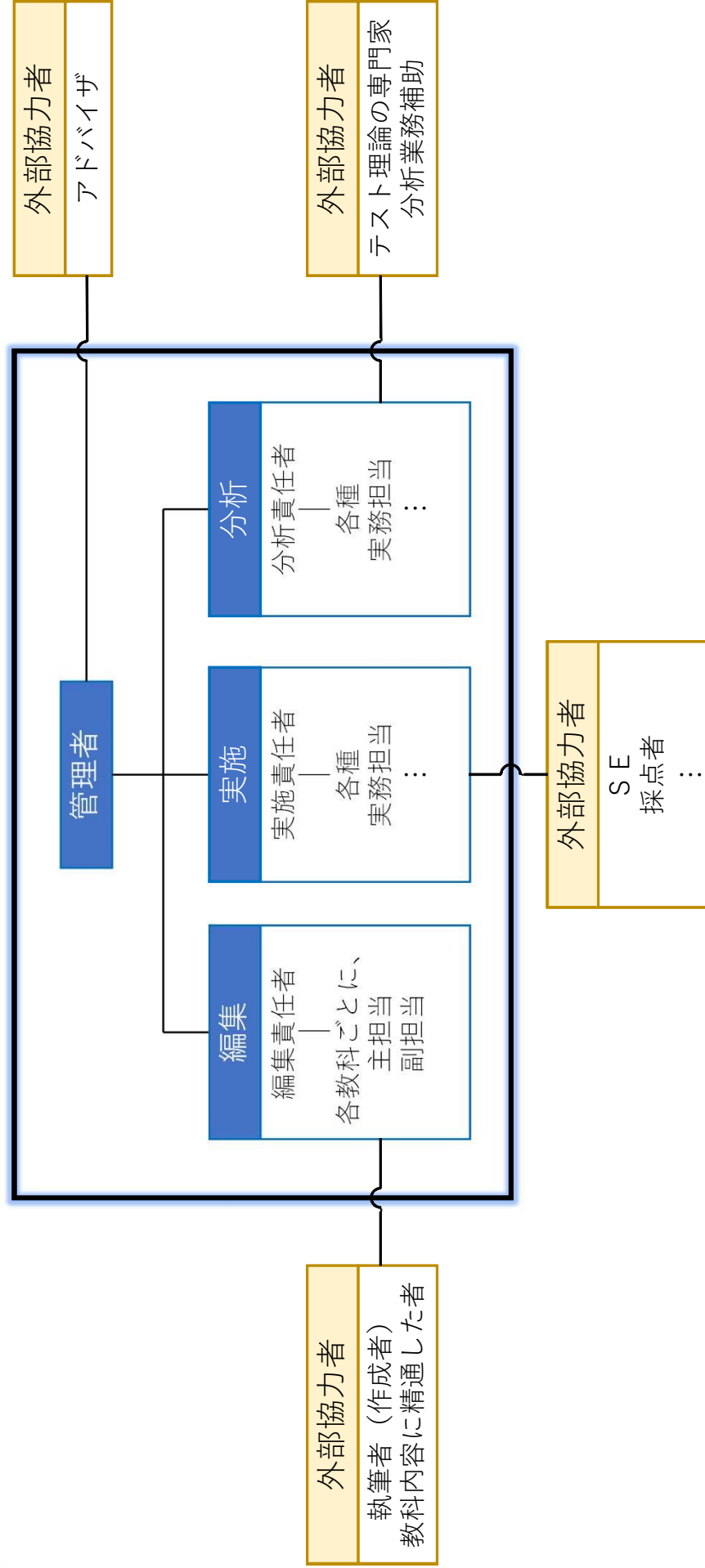
◎ 強く求められる ○ 求められる

△ 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



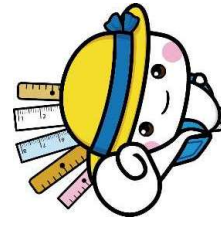


目々の学習での活用例



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																							

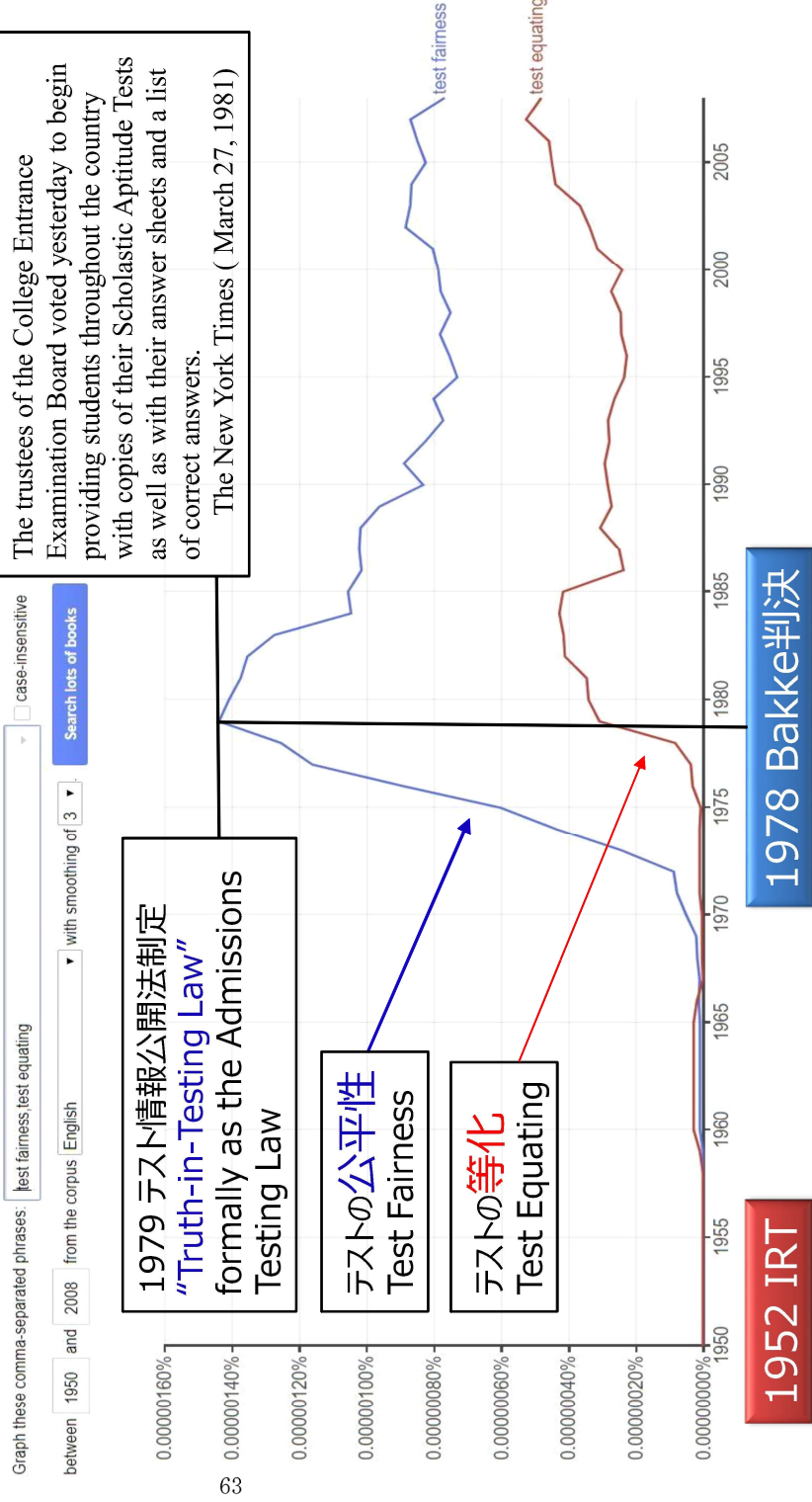


※問題の困難度に関する情報と児童生徒の回答結果から得られる学力スコアを照らして、より効果的な復習や宿題を項目プールから提示することが可能になる

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)

Google Books Ngram Viewer



◆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

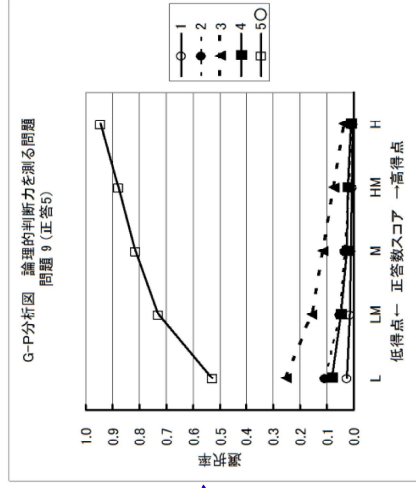
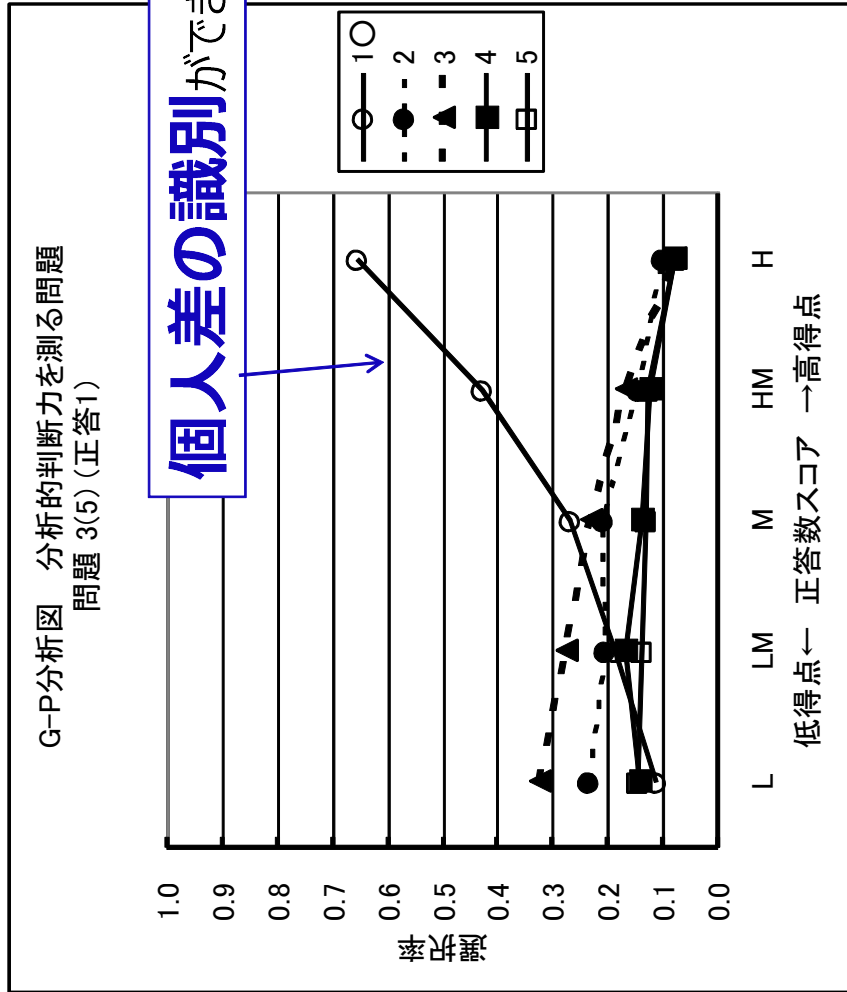
(◆**追跡**◆のための測定)

- 学力発達のサポート
- 埼玉県学力調査

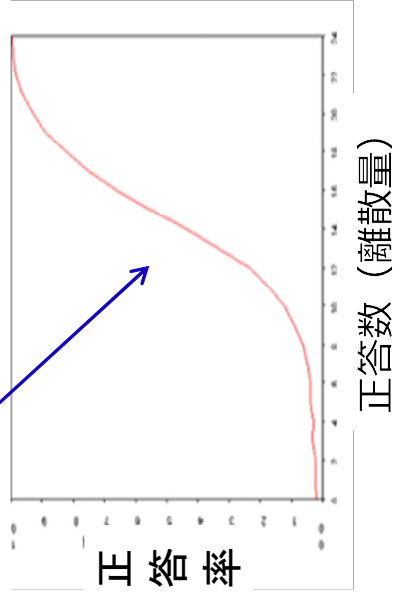
選択肢から見た良い項目

(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)



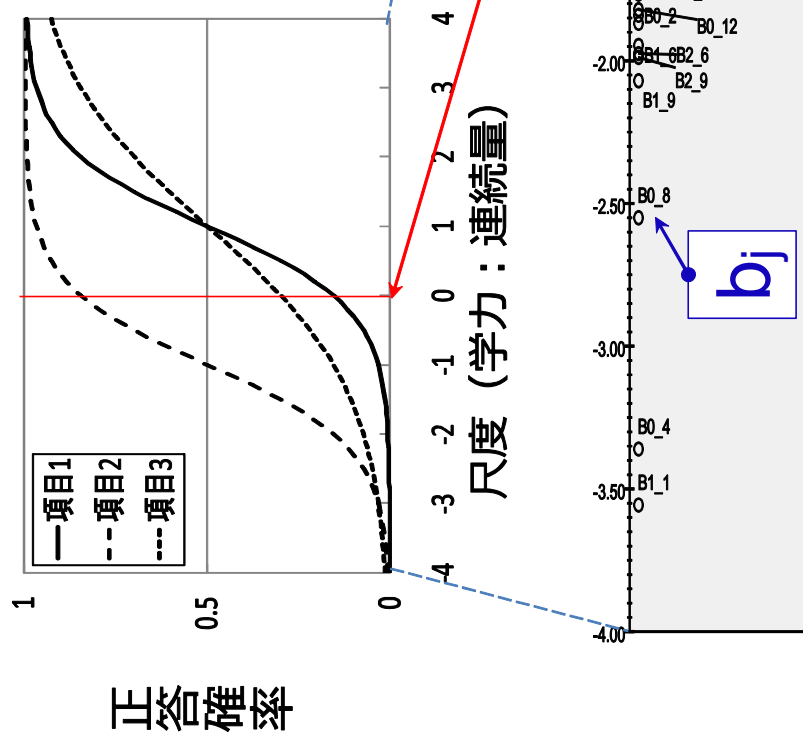
IRTモデルは統計モデル

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

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

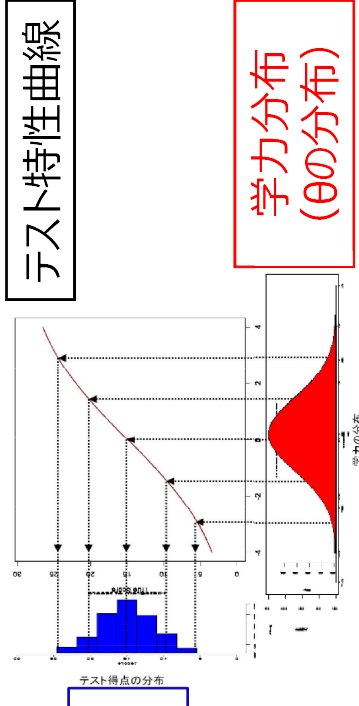
- 1) 学力 θ と項目困難度 b_j を分離
- 2) θ と b_j を同じ数直線 (尺度) 上で表現
- 3) b_j の組み合わせと正誤情報から θ を推定



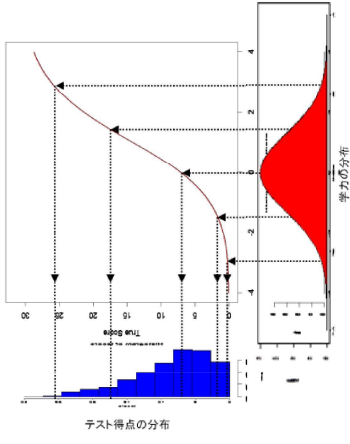
IRTモデルによるテストの得点分布の予測

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/shingai/chousa/shotou/085/shiryo/attach/1312362.htm

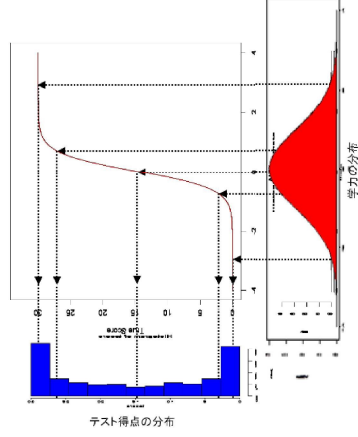
1) 一般的なテスト



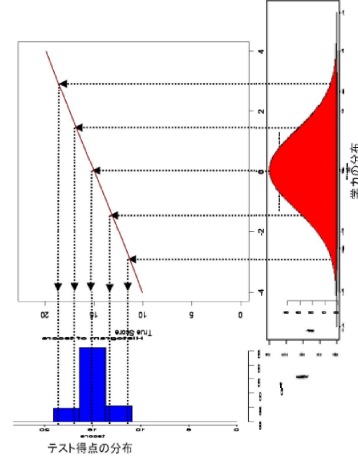
2) 高い学力層を選抜する場合



3) 資格の認定に使う場合



4) 個人差を小さく見せる場合

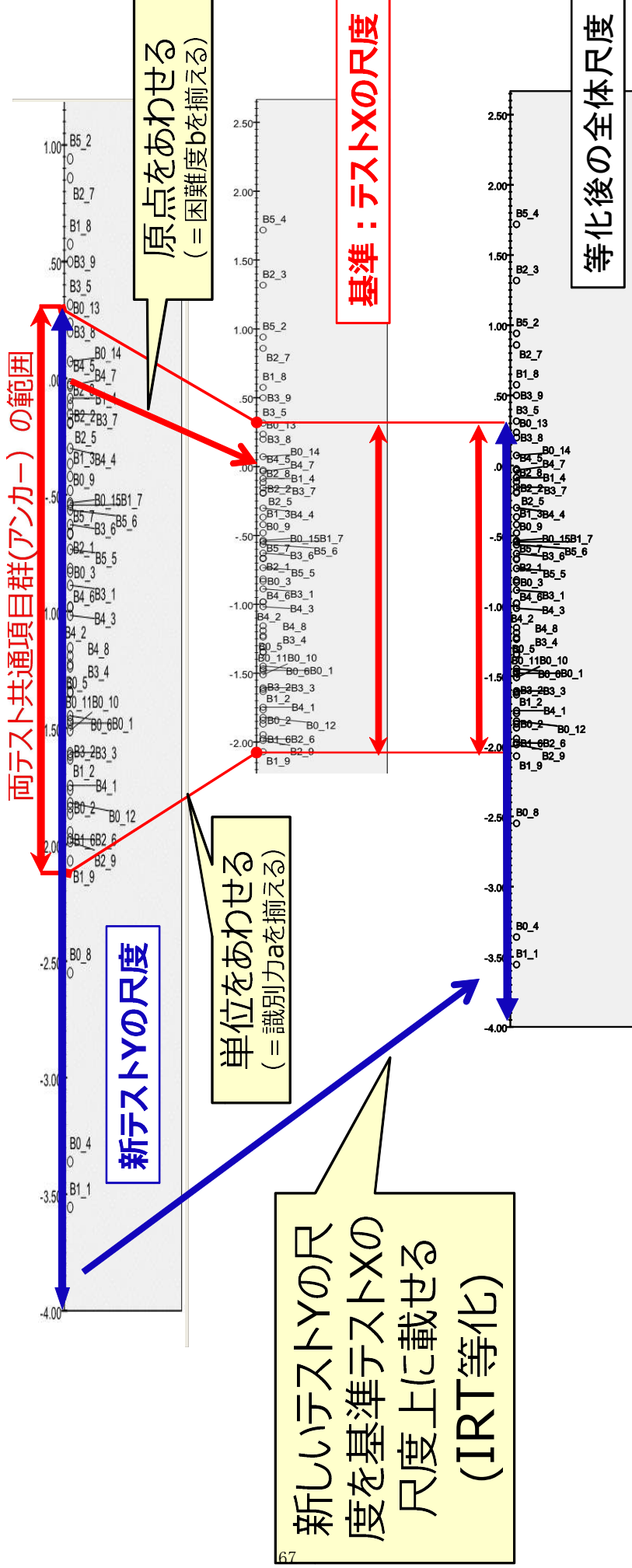


重要:

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

新しいテストYの尺度を基準テストXの尺度上に載せる (IRT等化)

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。