
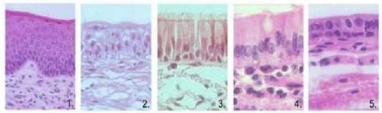
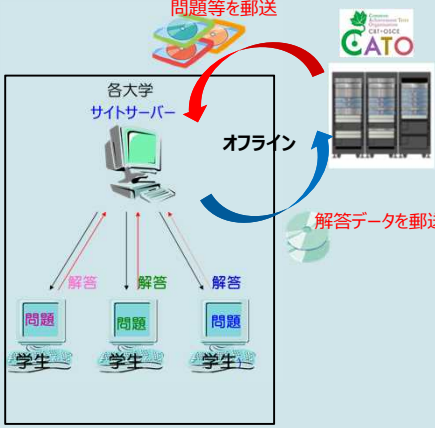


医療系大学間共用試験（CATO） -CBTを中心に-

| | |
|-----------------|---|
| 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)  <p>医学系82大学、歯学系29大学が参加</p> |
| 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; display: inline-block;"> <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;"> <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。

全国的な学力調査のCBT化検討WG 20200629 -CATO-

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

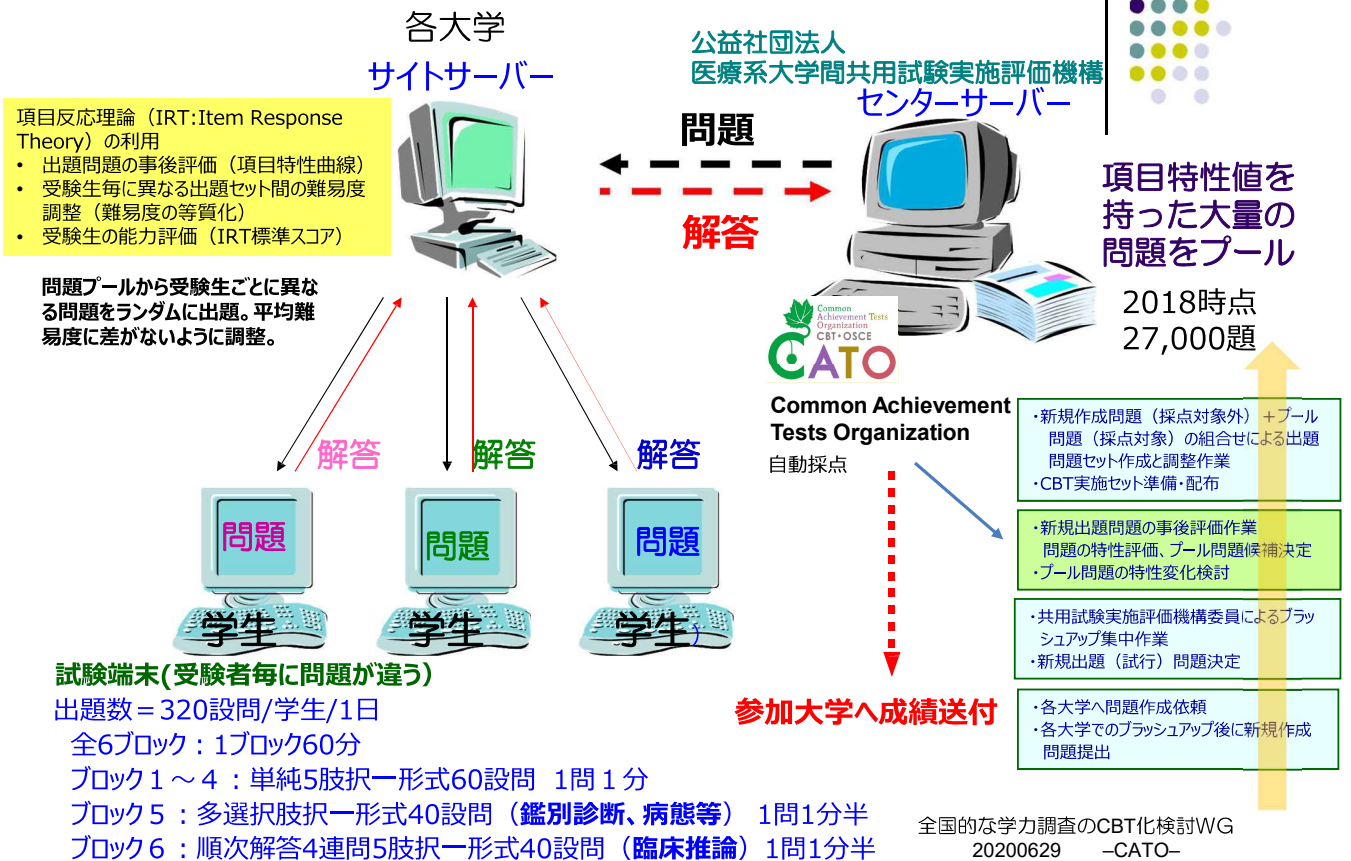


試験の信頼性・公平性・客観性・透明性等を担保するために我が国の代表的な試験評価の専門家が参画する試験信頼性向上専門部会(※)を設置し、恒常的に活動 (※2020年4月から試験信頼性向上委員会)

共用試験実施の概要 (2005年12月から正式実施)



CBT (Computer Based Testing) の実際



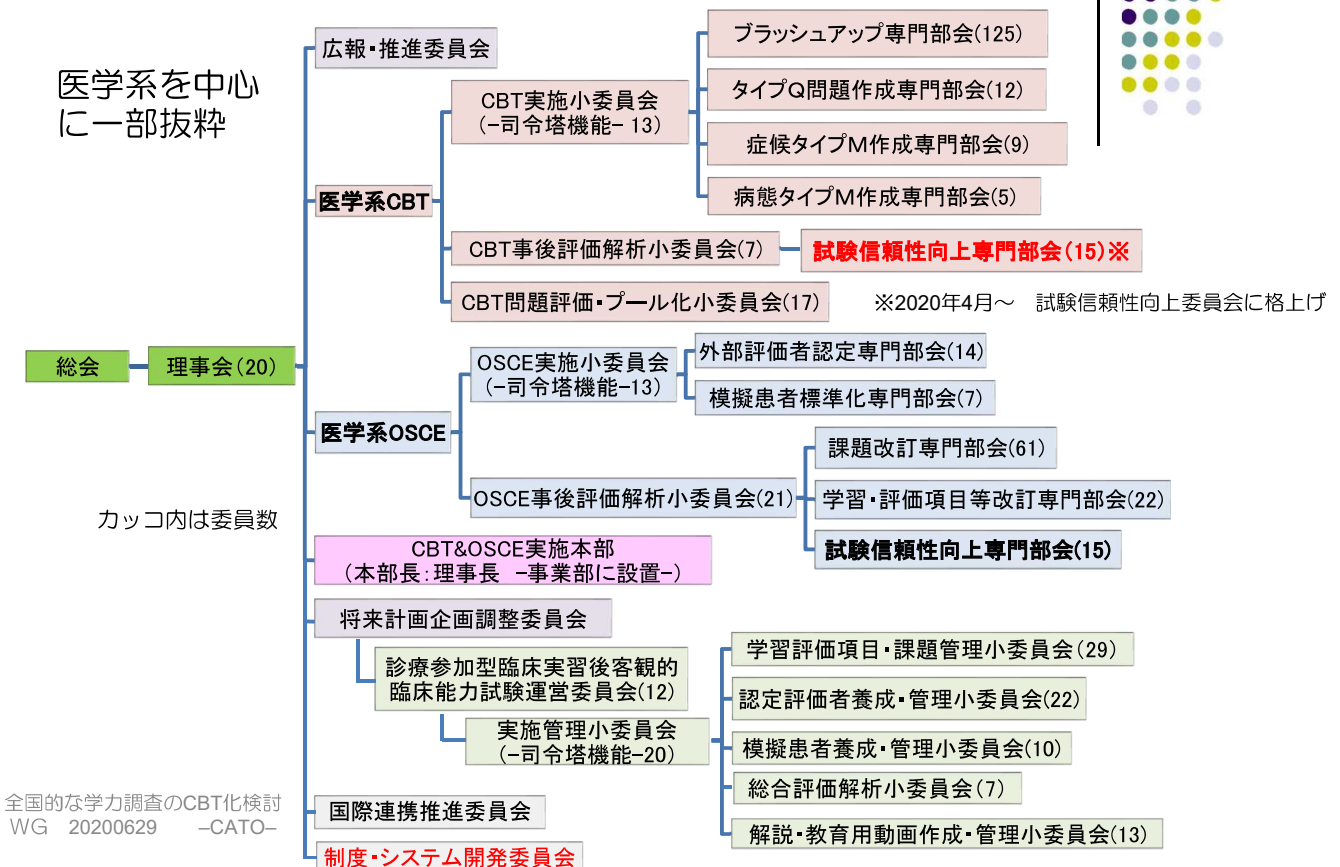


ブロックの構成

- 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

全国的な学力調査のCBT化検討WG 20200629 -CATO-

共用試験の運営を支える組織



全国的な学力調査のCBT化検討WG 20200629 -CATO-

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

ブループリントシミュレーション

2018実績 75回開催

- 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で）
- CBT実施キット準備・配布

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

全国的な学力調査のCBT化検討WG 20200629 -CATO-

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)



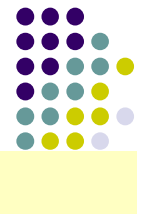
▶ 共用試験CBTの質保証

- ✓ 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

全国的な学力調査のCBT化検討WG 20200629 -CATO-

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（適切な露出回数の確保、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

全国的な学力調査のCBT化検討WG 20200629 -CATO-

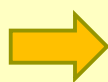
CBT&IRTをセットで -2-



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、トライアル実施計画を策定
並行して、

- ① トライアルサポート施設（※）の設置要請
 - ② トライアル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。

全国的な学力調査のCBT化検討WG 20200629 -CATO-

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

全国的な学力調査のCBT化検討WG 20200629 -CATO-

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

全国的な学力調査のCBT化検討WG 20200629 -CATO-

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エディションの開発（修正履歴、管理情報、ブラッシュアップ機能強化）

全国的な学力調査のCBT化検討WG 20200629 -CATO-

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（スタンドアローンモード、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

全国的な学力調査のCBT化検討WG 20200629 -CATO-

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。

全国的な学力調査のCBT化検討WG 20200629 -CATO-

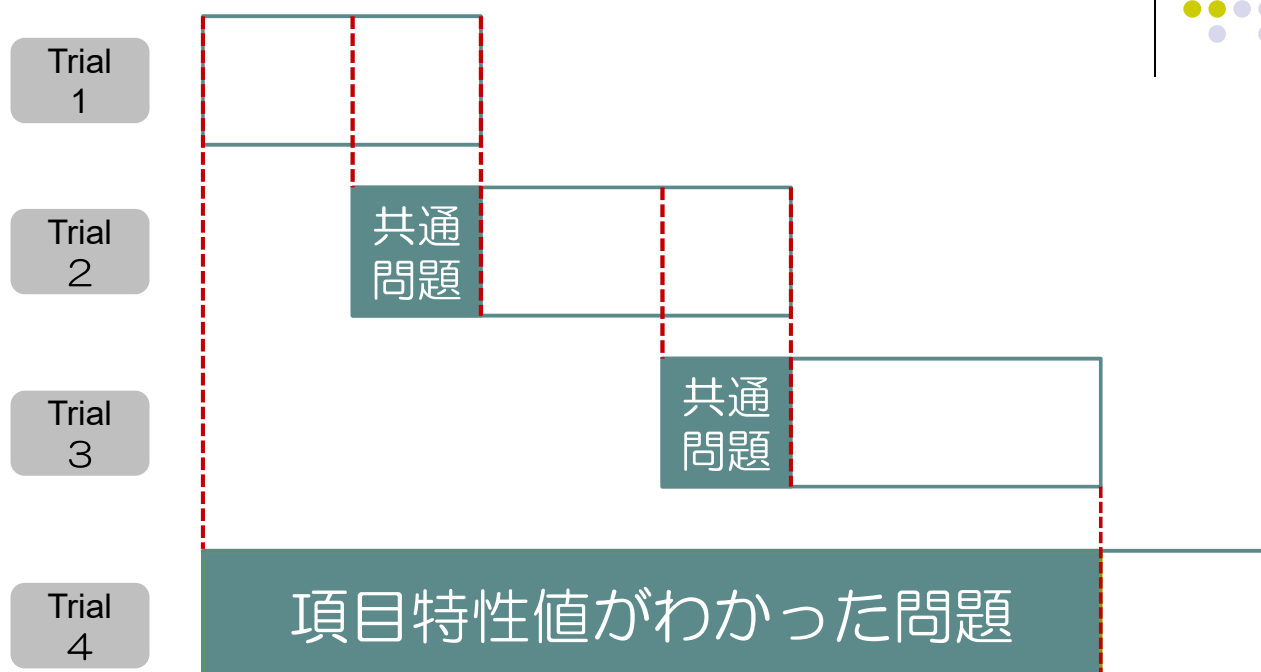
トライアル期間を通じたプール問題の蓄積



- 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陳代謝が可能
- 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題して新作問題の項目特性値の評価を行う

全国的な学力調査のCBT化検討WG 20200629 -CATO-

トライアルのテストデザイン

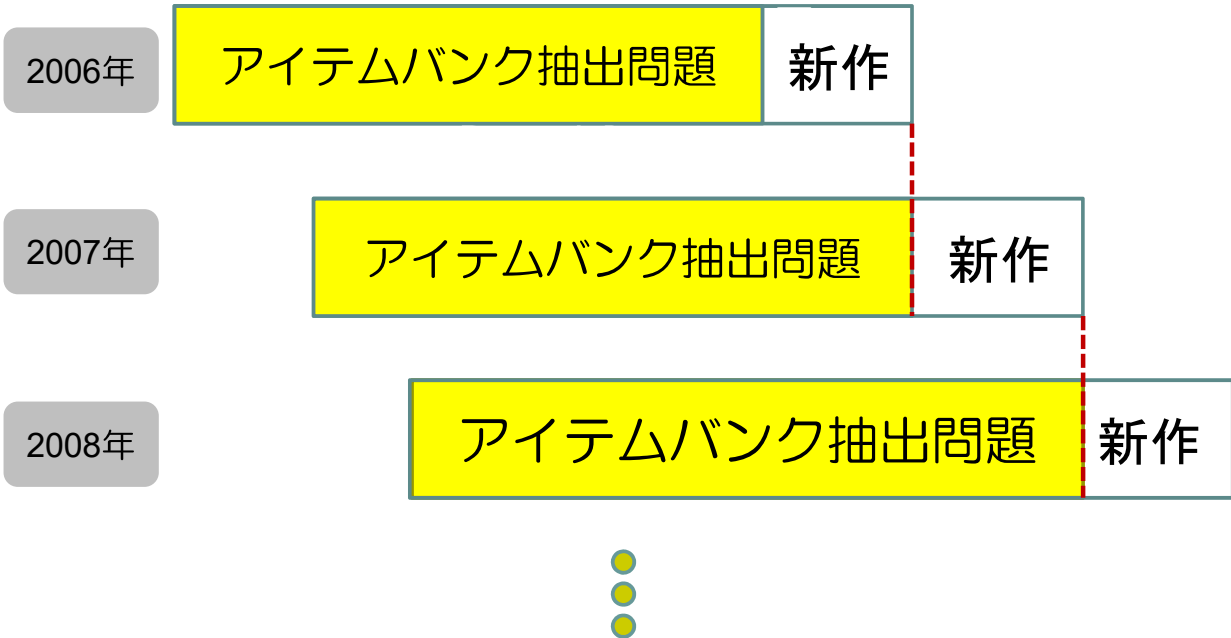


全国的な学力調査のCBT化検討WG 20200629 -CATO-

新作問題とアイテムバンク



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、異なる場所で、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った。
- 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なるように設定する。
- 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 of 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レーダーチャートによる学内平均との比較

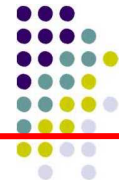
3) 学内評価

- 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

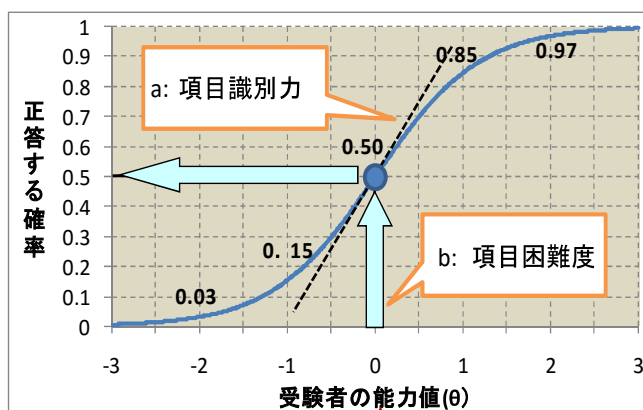
全国的な学力調査のCBT化検討WG 20200629 -CATO-



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で修正等
(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問題とする

全国的な学力調査のCBT化検討WG 20200629 -CATO-

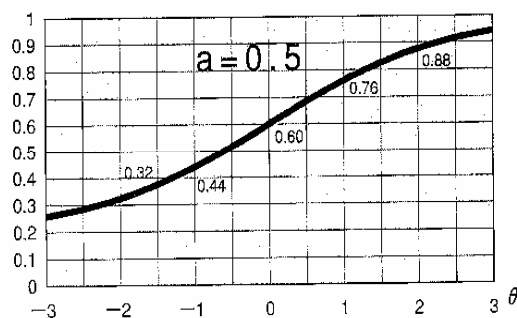
項目特性値 (項目識別力と項目難易度)



a 項目識別力

能力差のつく問題かどうかの指標

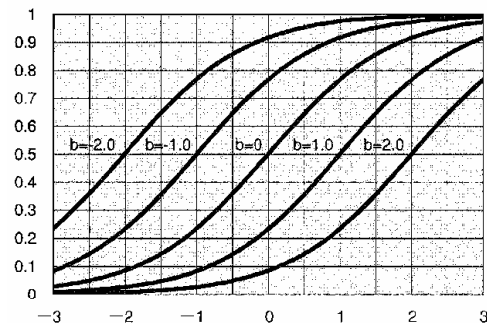
1. 特性曲線の最大の傾きを表す。
2. a が大きいほど識別力が大きい
3. 通常は0.3~2.0程度



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



全国的な学力調査のCBT化検討WG 20200629 -CATO-

「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

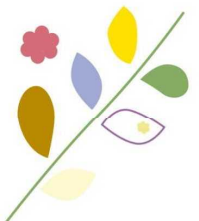
出題する問題が異なっても、児童・生徒の得点を比較できる。

- 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。



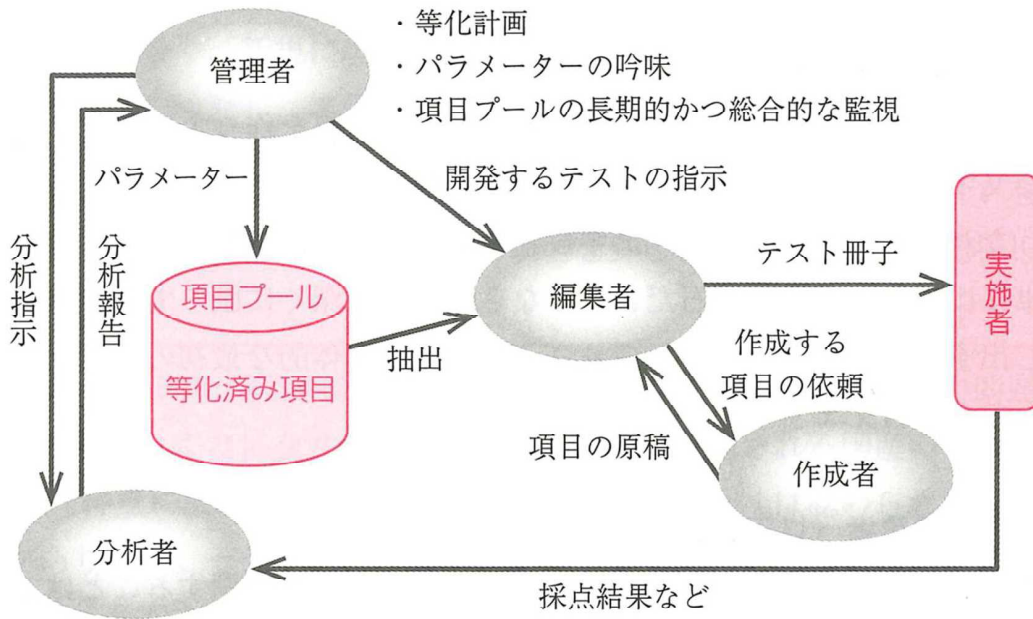
IRT調査の業務の流れ

- 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

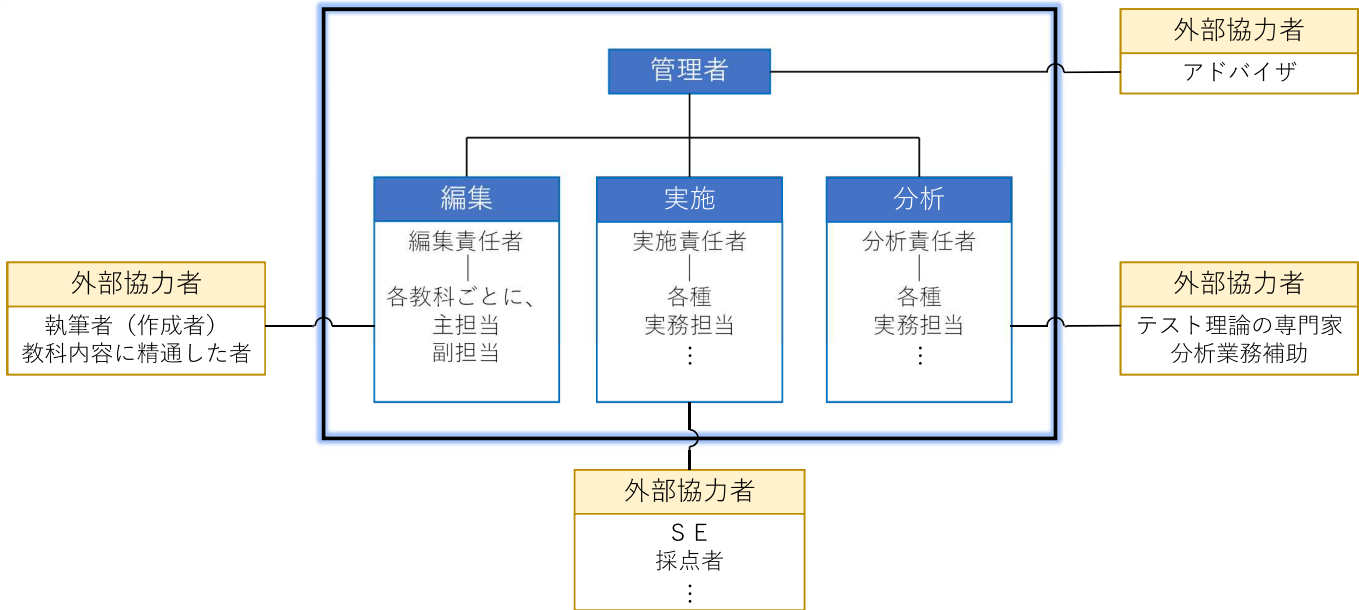
| 役割 | 技術 | | | 知見 | | | | 資質 | |
|-----|----|----|----|------|------|------|--------------|----------------------|--------------------|
| | 分析 | 編集 | 作問 | 測定理論 | 教科知識 | 教育情勢 | 他 | 共通 | 個別 |
| 管理者 | ○ | △ | △ | ○ | ○ | ◎ | 教育に関する総合的な知見 | コミュニケーション力 適度な柔軟性 | ・判断力 ・計画性 |
| 編集者 | ○ | ◎ | △ | △ | ○ | ○ | | | ・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。

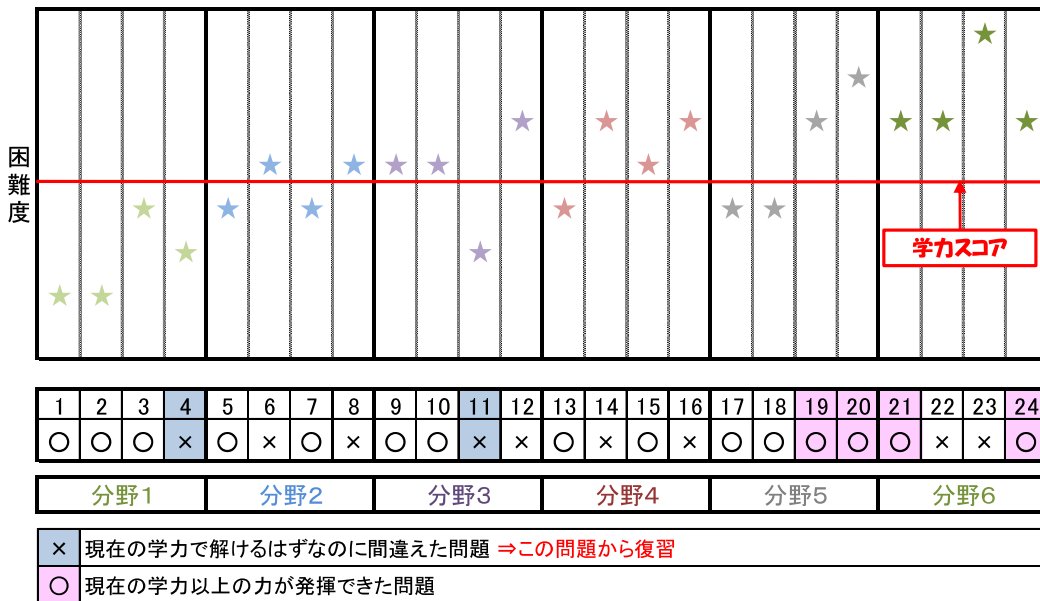
- ◎ 強く求められる
- 求められる
- △ 最低限求められる
- なくても良い



調査の実務に必要な体制



日々の学習での活用例

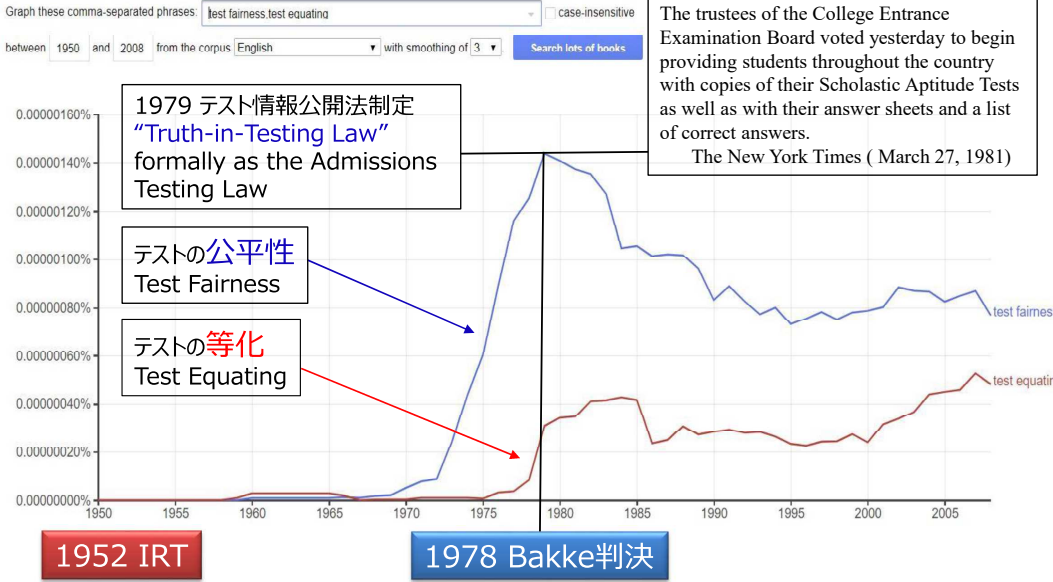


※問題の困難度に関する情報と児童生徒の回答結果から得られる学力スコアを照らして、より効果的な復習や宿題を項目プールから提示することが可能になる

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

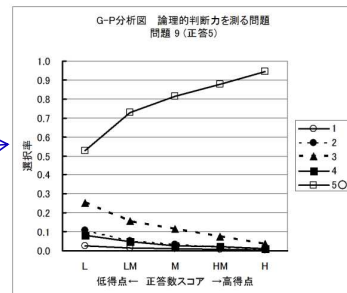
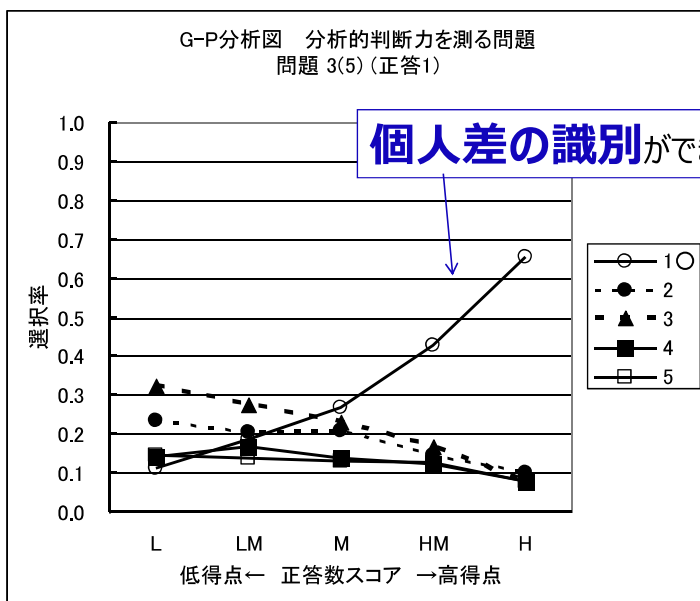
- 1) 個人の処遇：個人スコア
(◆総括的評価◆のための測定)
・選抜(対訴訟：極めて厳格な測定)
・医療系大学間共用試験CBT
・全国学調_本体調査
- 2) 集団の実態：集団スコア
(◆EBPM◆のための測定)
・全国学調_経年変化分析調査
・PISA・TIMSS・NAEP等
- 3) 個人の進捗：個人スコア
(◆「学び」◆のための測定)
・GIGA個別最適な学習
・形成的アセスメント
- 4) 個人の成長：個人スコア
(◆追跡◆のための測定)
・学力発達のサポート
・埼玉県学力調査

IRTモデルの
定式化の例 1/2

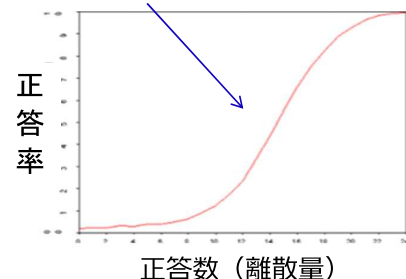
選択肢から見た良い項目

(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/

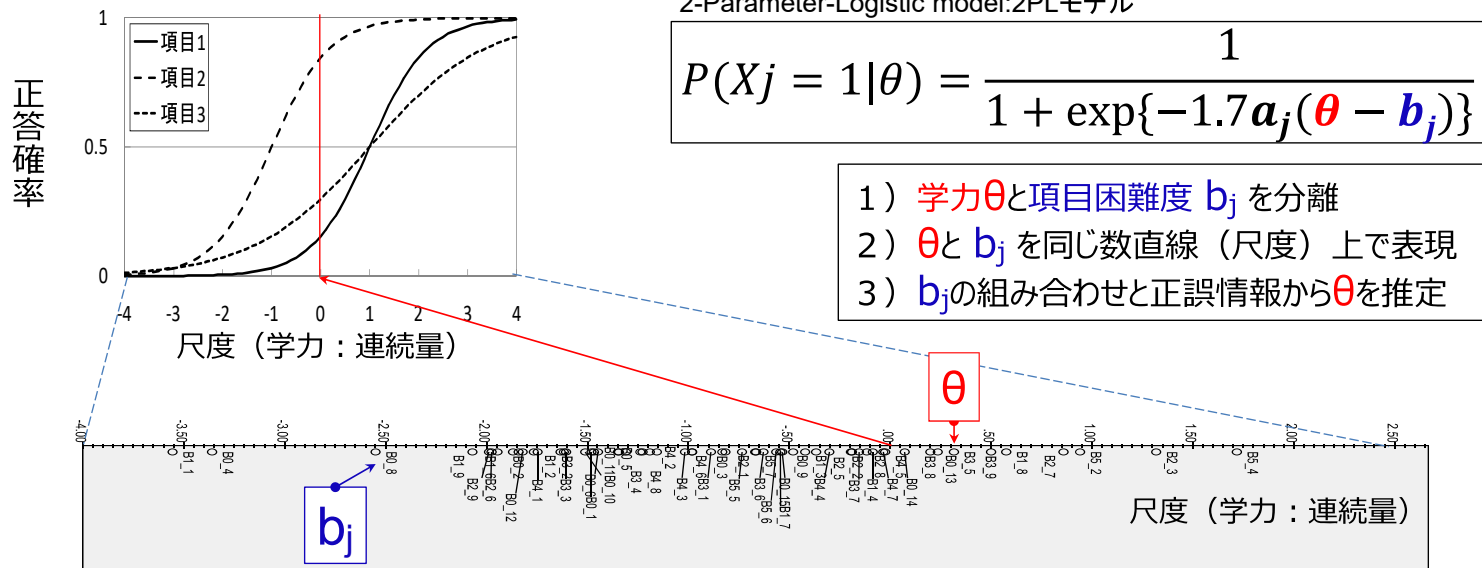


良い問題の項目特性曲線：DNC数学 I (柴山が復元)



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



平成22年度文科省委託調査研究報告書 p.77

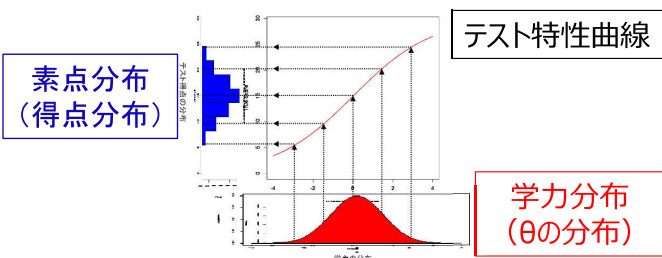
https://www.mext.go.jp/b_menu/shingi/chousa/shotou/085/shiryo/attach/1312362.htm

IRTモデルによるテストの得点分布の予測

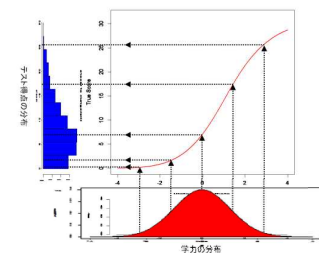
平成22年度文部科学省委託調査研究報告書p.5

https://www.mext.go.jp/b_menu/shingi/chousa/shotou/085/shiryo/attach/1312362.htm

1) 一般的なテスト

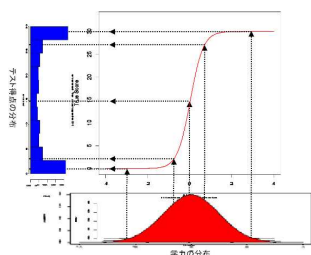


2) 高い学力層を選抜する場合

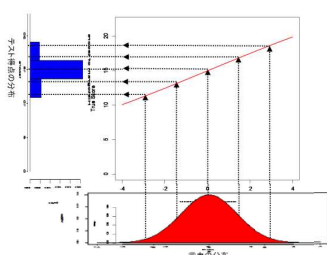


重要:
学力分布は同じでも、
テストの特性が変われば、
素点分布／得点分布の形
は変化する

3) 資格の認定に使う場合

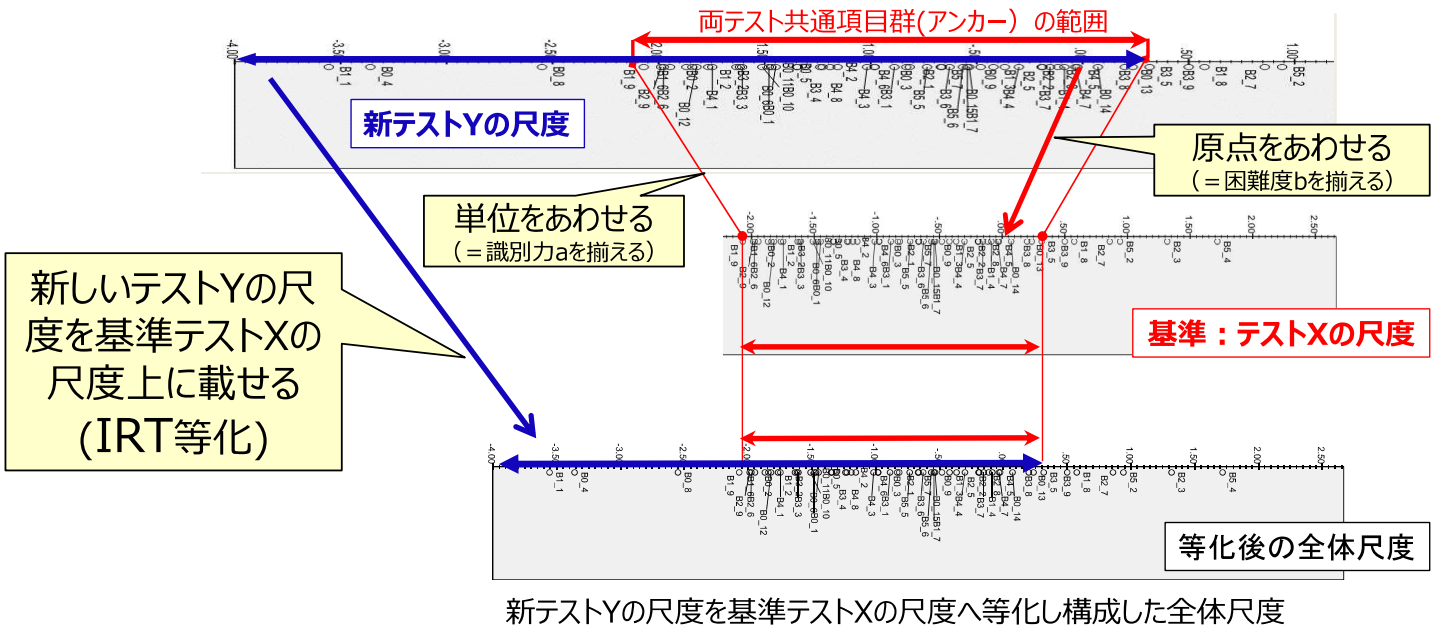


4) 個人差を小さく見せる場合



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



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、P、Q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。