

全国学力・学習状況調査の実施工程とCBT化による影響

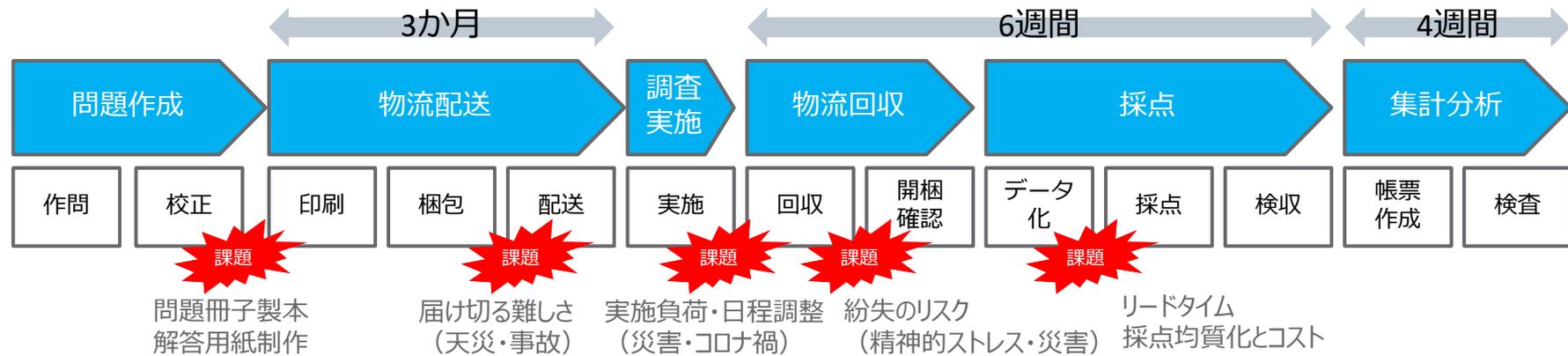
第4回 CBT化検討ワーキンググループ発表資料

2020.07.31

JIEM 株式会社教育測定研究所

- CBT化によるメリット
 - ・**大量の資材を扱うためのリードタイムの短縮（準備期間を3か月程度短縮可能）**
印刷・製本、梱包ピックアップ時間の削減（実施主体の時間的猶予）
天災・事故等による不測の事態へのスムーズな対応が可能となる（設置管理者、学校の安心）
紙媒体の電子化作業のリードタイム削減（児童生徒へのタイムリーなフィードバック）
 - ・**物としての移動に伴うコストの削減**
調査資材の未達、不足による実施ができないリスクの低減（学校、児童生徒の安心）
解答用紙紛失のリスク低減（設置管理者、学校の安心）
 - ・**確実な調査の実施**
CBTでは問題を一定期間非公開とする設定が容易に行え、地域ごとの事情に合った時期に実施日を決定することも検討可能（新型コロナウイルス対応）
- CBT化による発展
 - ・**PBTでは表現できない多様な形式による設題**
多様な形式に対応することにより測定可能な構成要素が増える
 - ・**アダプティブ化**
受験者の習熟度に合わせて問題の提示・フィードバックが可能
より短い時間で能力測定が可能（学校・児童生徒の負担減）
 - ・**パフォーマンス分析（解答ログ解析）**
解答プロセスやログを解析することにより、より細やかな能力測定、指導への活用が可能

■ 現状の調査プロセス



■ 当社が考えるCBTプロセス

➢ 学校PCにて実施



No	項目	教育的メリット
1	受験から結果までの受け取り時間の短縮	<ul style="list-style-type: none"> 結果をより早いタイミングで指導に活かせる
2	CBTで多様な形式に対応することにより測定可能な構成要素が増える	<ul style="list-style-type: none"> より「真正度（authenticity）」の高い問題形式や問題の提示方法、提示タイミングや解答のさせ方などのバリエーションが増え、テストの目的に応じて対応可能 今まで問うのが難しかった構成要素の測定が可能
3	アダプティブ化	<ul style="list-style-type: none"> 受験者の習熟度に合わせて問題の提示・フィードバックが可能 受験時間の短縮
4	アイテムバンク化	<ul style="list-style-type: none"> 個人単位・団体単位、過去の履歴に応じた結果分析・フィードバックが可能
5	解答ログの分析	<ul style="list-style-type: none"> より細やかな能力測定が可能（例：解答プロセス、解答時間からの躓きポイントの分析など）、分析結果を将来的な指導に活かせる
6	特別な配慮が必要な児童生徒への対応	<ul style="list-style-type: none"> 画面の文字サイズ変更、問題文読み上げなど対応のバリエーションが増える

No	課題	対応策
1	児童生徒が解答を手書きで行うことの重要性を尊重したい	<ul style="list-style-type: none"> ・入力デバイスが整うまでは解答を解答用紙に手書きとすることで、児童生徒の手書き解答を主体とする問題形式にもスムーズに対応可能
2	学校現場におけるPC環境の多様性への対応	<ul style="list-style-type: none"> ・事前に実施環境を調査することで、必要な環境で動作するCBTを隈なく用意する <p>実施例) 実施OSに依存し難いJVM上で動くCBTとして実装。JDKも合わせて用意することで、実施環境のバージョンの違いも回避</p>
3	トラブルへの対応	<ul style="list-style-type: none"> ・トラブルへの対応方法を画一化し、監督者（教員）への負荷を低減する <p>実施例)</p> <ul style="list-style-type: none"> ・サーバ、PCの2つに解答状況や解答自体を冗長的に保持する実装を行う ・どの環境で実施しても、開始時に最新のステータスを確認し適切な位置より開始する

- コンピューター技術を活かし、テストの目的に応じて多様な問題形式が可能
 - 受験者へのInput（課題の提示）
 - 音声、動画、VRなど複合的な要素
 - 検索など受験者が情報を選んで取得
 - チャット形式など解答に応じてインタラクティブに課題を分岐
 - Webクラスルームなどグループ単位の課題
 - 受験者の解答方法
 - マウス、キーボード、タッチスクリーンなどのデバイスを利用した解答方法のバリエーション
 - 発話による解答
 - 受験者のヘルプやカスタマイズ機能
 - 受験者による画面の文字サイズ変更
 - 問題文読み上げ
 - 採点方法
 - 人工知能やコーパスなどを利用した自動採点
 - 解答パターンや確率などを利用した採点の自動細分化



従来の測定領域だけでなく、紙のテストでは問うのが難しかった
教科の枠を超えたこれからの時代に必要とされるスキル
の測定も可能に

- 問題：英語の文章を読み、講義を聞いて、内容をまとめる

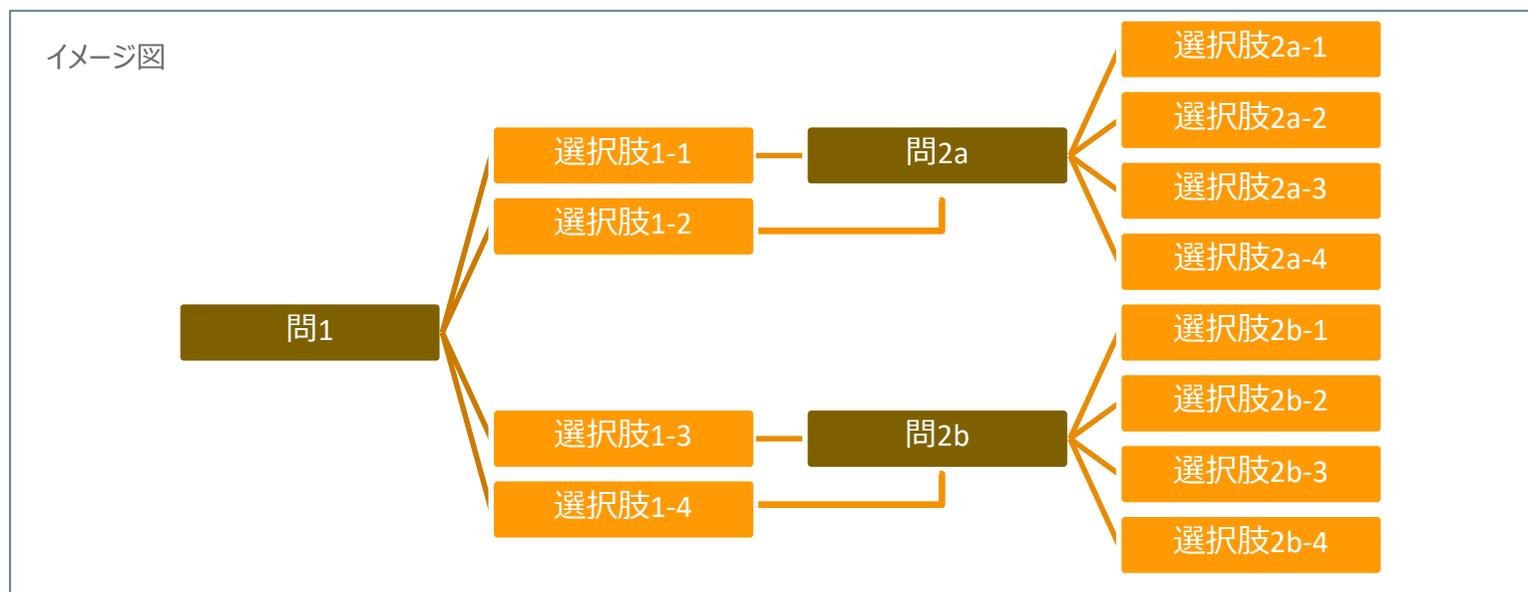


「読む・書く・聞く・話す」など技能ごとに分けるのではなく、英語を実際に活用する場（通常、複数の技能が統合されている）に近い形で **言語運用能力**を測定することが可能。

CBTでは、画像や映像と音声の提示タイミングを調整することにより、**リアルなシチュエーションの再現**が可能

CBT問題形式の例（将来的な可能性）

- 問題：仮想の他者と協同で問題を解決する問題状況の場面が設定され、他者と会話しながら問題を進める。解答した内容により、会話の内容が分岐し、インタラクティブな問題設計が可能
- 解答：問題状況の客観的分析、他者の意見や考えを理解した上で、自分の意見に近い選択肢を選ぶ



CBT化することにより**解答に応じた課題の分岐**が可能

課題解決など**答えが複数**考えられる要素において、**インタラクティブ**なプロセスを得ることで
思考力や課題解決能力の測定も問うことができる

- 問題：日本語（ニュース、討論、物語など）を音声または動画で聞く

- 解答

- 内容理解を問う
- 内容を正確に他の人に伝える設定で説明する
- 内容に関する自分の意見を言う



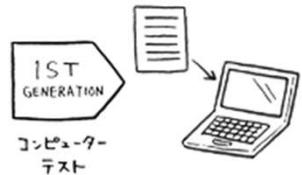
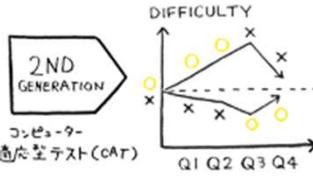
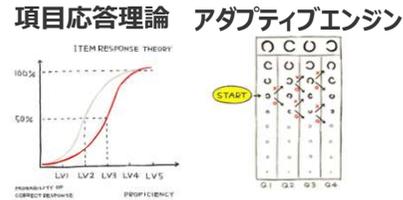
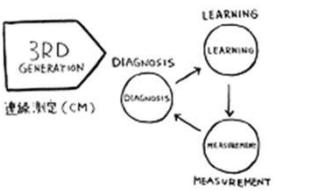
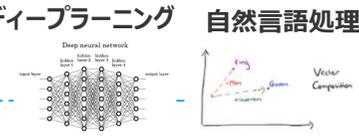
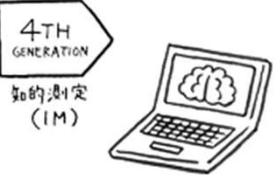
「聞く」能力は、相手の言っている内容を正しく聞き取り、言葉として理解するだけでなく、相手の「真意」や「ニュアンス」も感じ取ることが必要である。
またそれは、**コミュニケーション能力**にも密接に関係している。

CBT化することにより問題として使える題材も増え、**従来の国語のテストでは問い難かった「聞く力」**も測ることが可能となり、また**「話す力」**も問うことができる

- 問題：ある人が調べた社会問題（地球温暖化など）の調査結果を提示
- 解答
 - 調査結果のデータと内容の妥当性をインターネットを活用して調べ、問いに答える
 - 調べた結果をプレゼンテーション資料としてまとめる



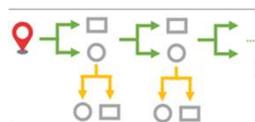
情報過多のIT社会において**情報の妥当性を自ら調査し**、客観的に判断する**情報選択力**、情報を論理的にまとめる**思考力**は現代社会で求められている能力であり、CBT化によって**インターネット検索も活用**した問題作成も可能である

EdTechフェーズ	期待されるメリット	サービス提供	技術モジュール
<p>1ST GENERATION</p> <p>コンピュータテスト</p> 	<ul style="list-style-type: none"> コンピュータ化されているが、未調整のスコアまたは偏差スコアが使用される紙のテスト ショートアンサーテストに利用可能な自動採点システム 	✓	
<p>2ND GENERATION</p> <p>コンピュータ適応型テスト (CAT)</p> 	<ul style="list-style-type: none"> 前の設問解答のパフォーマンスに基づいて異なる設問を提供するシステム 受験者の能力の正確な測定を維持しながら、テスト時間を短縮 	✓ CASEC 2001~	<p>項目応答理論 アダプティブエンジン</p> 
<p>3RD GENERATION</p> <p>連続測定 (CM)</p> 	<ul style="list-style-type: none"> 学習支援のためこのアプローチにより十分に学習されていない領域を特定 再テストと自習を通じて、測定を効率的な学習に直接リンク 	✓ ☆ ルールベース適応学習	<p>ディープラーニング 自然言語処理</p> 
<p>4TH GENERATION</p> <p>知的測定 (IM)</p> 	<ul style="list-style-type: none"> AI技術により、採点およびその他の手動タスクを自動化 効率的かつ正確な採点と学習カリキュラムのカスタマイズを実現 	✓ ☆ AI推奨エンジンを備えた非ルールベースの適応学習システム ☆ ブロックチェーンタイプ、個人情報、個人パフォーマンス管理システム	<ul style="list-style-type: none"> - 推奨エンジン - NLPエンジン - 文字認識エンジン - 音声認識エンジン - 適応学習エンジン - アダプティブテストエンジン
<p>5TH GENERATION</p> <p>5G</p> 	<ul style="list-style-type: none"> データレートがはるかに高いため、ユーザーはコンテンツをより迅速にダウンロード可能 はるかに少ない待時間により、ユーザーはネットワークからデータを要求する際の遅延問題解消 	開発中 ☆ AR /VRプラットフォーム 第4世代+グループ型学習機能搭載モデル	

リコメンドエンジン

推奨と分析のための協調的適応エンジン

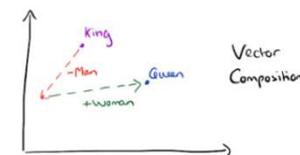
機械学習に基づく汎用推奨エンジン
さまざまな方法と推奨事項の特性を組み合わせた高性能エンジン



自然言語処理(NLP) :

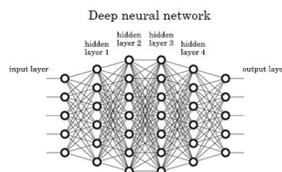
自然言語をコンピュータに処理させる技術

当技術応用により、文章の解析が可能になり自動採点、自動翻訳、対話システムを実現



文字認識エンジン

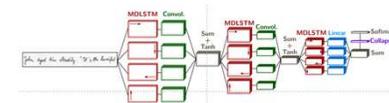
人工知能 (AI) の急速な発展を引き起こした機械学習技術
非常に正確な認識を可能にし、自動運転システム、医療技術、および幅広い分野に適用可能



音声認識エンジン

LSTM: 長・短期記憶

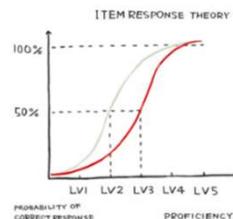
LSTMは、時系列データの非常に強力な予測モデルとして、また音声認識や文字認識などの分野での効果的な方法として注目されている



項目応答理論(IRT)

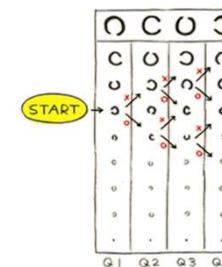
項目応答理論では、テスト難易度と受験者能力を別々に把握し、「ある難易度の問題に対し、ある能力をもった回答者の回答確率はX%である」

ということを根拠に、確率的に受験者の能力を求めることが可能



適応型測定・学習

コンピュータを用いて受験者の回答に応じた質問を適宜変えていく、コンピュータ適応型テスト (CAT) によって、短時間で受験者の能力が高い精度で測定可能

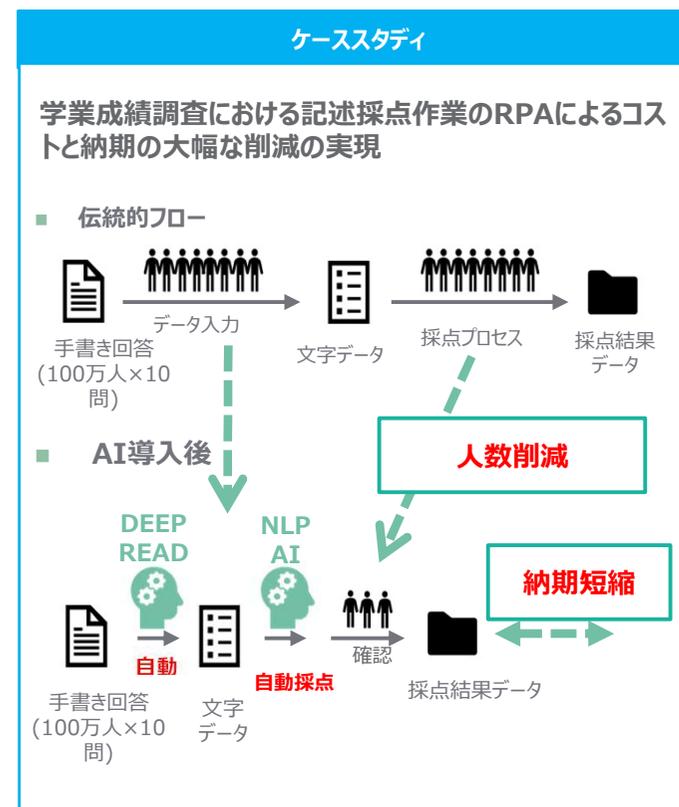


■ 導入目的

採点の高速化と品質の向上により、より効率的な教育サービスに貢献

■ 導入メリット

	採点エンジン	伝統的なシステム
児童生徒	<ul style="list-style-type: none"> - 迅速で正確な結果 - 学業成績向上を加速 	<ul style="list-style-type: none"> - 教師のパフォーマンスに依存
先生	<ul style="list-style-type: none"> - 効率 - 参照による採点品質の向上 - 採点作業負担の軽減 	<ul style="list-style-type: none"> - 心理的および身体的負担
学校/地区	<ul style="list-style-type: none"> - 採点品質の統一と改善 - 全体的な採点時間の短縮 - 採点者数の低減 	<ul style="list-style-type: none"> - 教師の採点品質に依存 - 採点時間の確保 - 採点者リソースの確保



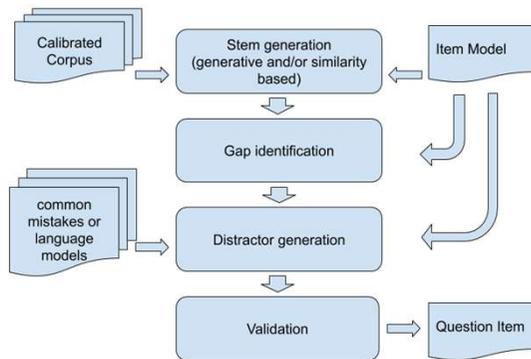
■ 導入目的

アイテム(問題)作成を自動化し、多数の高品質アイテムを低コストで生成、テスト機会を増やす

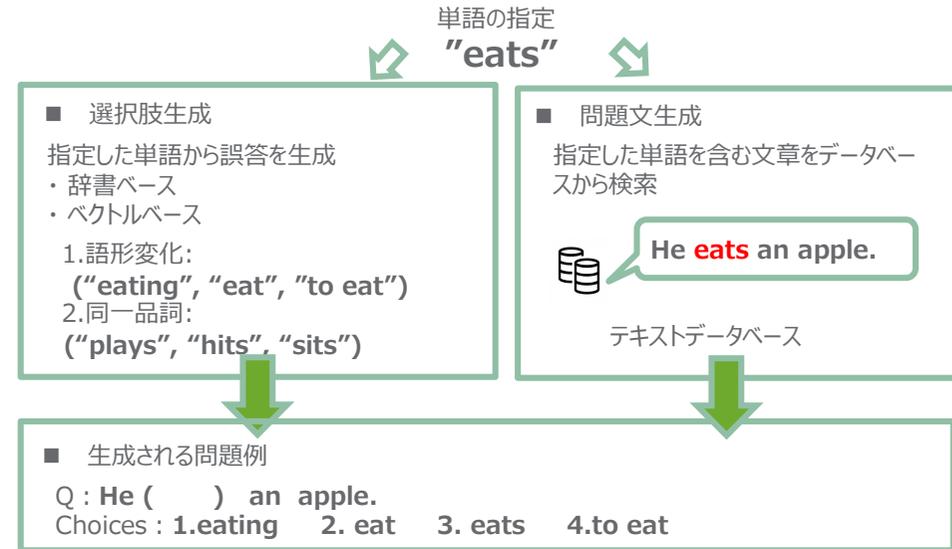
■ 導入メリット

	AIベース自動アイテム生成システム	伝統的なシステム
児童生徒	- 受験機会の増加	- 制限的な受験機会
先生	- アイテム作成の負荷を削減	- 短時間で質の高い大量のアイテムを作成するための心理的および身体的負担
学校/地域	- 問題作成品質の統一と改善	- 教師の問題作成の品質に依存

□ アイテム作成プロセス



□ 質問文と代替案の自動生成



□ システム出力例 1

Where did your parents _____ meet each other? Missing word: first
Glove:
'second' (0.92), 'third' (0.87), 'one' (0.84), 'fourth' (0.83), 'last' (0.83), 'time' (0.83), 'only' (0.82), 'the' (0.82), 'later' (0.81)
BERT:
'first' (0.82), 'ever' (0.09), 'even' (0.01), 'last' (0.01), 'really' (0.01), 'finally' (0.006), 'eventually' (0.004), 'actually' (0.003), 'originally' (0.003)

□ システム出力例 2

Query: A man is eating pasta.
Top 5 most similar sentences in corpus:
A man is eating a piece of bread. (Score: 0.8518)
A man is eating a food. (Score: 0.8020)
A monkey is playing drums. (Score: 0.4167)
A man is riding a horse. (Score: 0.2621)
A man is riding a white horse on an enclosed ground. (Score: 0.2379)

実施方式への対応

「英検 S-CBT」の実施方式の特徴である、Speaking（吹き込み式）と Writing（手書き式）について、それぞれ7割程度の方が概ね対応できていることが伺えた。

スピーキング試験（ヘッドセットで問題を聞きマイクに回答を吹き込む）はどう感じましたか。



ライティング試験（解答を手書きする方式）はどう感じましたか。



※アンケート実施日：7/11 英検S-CBT受験者187名に実施 有効回答170人
出典：https://www.eiken.or.jp/s-cbt/info/2020/pdf/20200714_info_scbt.pdf

オンラインプロクタリングサービス概要

- 基本的な流れ：オンラインの受験の様子を動画で保存→それを人が（受験後に）チェックして本人確認・不正がなかったかを確認する



CBT受験：
受験の様子を録画し、動画で保存

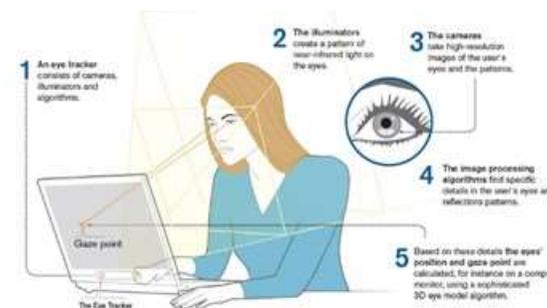


データ送信：
受験後、録画した動画をデータチェック拠点
に送信



データチェック：
録画映像をAIプリチェック、人手で確認
不正がないか確認

- 受験者動向チェック
 - 目線のチェック、試験画面と関係のない場所ばかりを見ていないか(アイトラッキングの活用)
 - 試験中に複数人が映っていないか
 - 他の人との会話をしていないか(音声チェック)
- 画面動作チェック
 - 試験画面を閉じて再開する等の行為がないか
 - 別ウィンドウを開き、web検索などを行っていないか
 - タイピング型の解答時に、明らかなペースト行為などが行われていないか



- 補足：試験の内容・目的に合わせて、チェック度合いを調整
(例：顔認証の定点チェックのみ、常時動画によるサンプルチェックや全件・全内容チェック等)