

パネルディスカッション

CBT入試への期待と課題

公益社団法人医療系大学間共用試験実施評価機構仁田善雄 理事学
術顧問

早稲田大学教育・総合科学学術院澤木泰代 教授

長崎大学教育開発推進機構アドミッションセンター 吉村宰 教授

電気通信大学大学院情報理工学研究科 植野真臣教授

仁田 善雄 先生

- 昭和52年4月 千葉大学助手（医学部・公衆衛生学）に採用
- 平成14年5月 東京医科歯科大学講師（医歯学教育システム研究センター）に採用
- 平成15年4月 東京医科歯科大学准教授（医歯学教育システム研究センター）に採用
- 平成20年4月 社団法人医療系大学間共用試験実施評価機構研究部長に採用
- 平成26年4月 社団法人→公益社団法人 に変更
- 平成30年 日本テスト学会賞を受賞
- 令和2年 大学入学共通テスト企画委員会CBT活用検討部会委員に委嘱
- 令和5年3月 公益社団法人医療系大学間共用試験実施評価機構研究部長を退職
- 令和5年4月 公益社団法人医療系大学間共用試験実施評価機構学術顧問に任命

澤木 泰代 先生

経歴

- 熊本県生まれ。熊本大学教育学部卒。熊本県公立中学校教諭を経て英語教授法修士号取得（イリノイ大学アーバナ・シャンペーン校）。昭和女子大学助手を勤めた後、2003年応用言語学博士号取得（UCLA）。同年よりEducational Testing Service（米国ニュージャージー州）研究員。2009年早稲田大学教育・総合科学学術院准教授。2014年4月より現職。専門は応用言語学。現在アジア言語テスト学会共同会長、日本言語テスト学会理事。

CBT関連

- ETS研究開発部 主にTOEFL iBTの妥当性研究を担当 2003年8月～2009年2月
- TOEFL Committee of Examiners (第三者専門家委員会) 2018年10月～2022年9月
- 大学入試センター CBT活用検討部会委員 2020年4月～2022年3月
- 大学入試センター 入学者選抜におけるCBT活用に関するワーキングチーム委員 2021年4月～2023年3月
- 科研費 要約ライティング形成的評価用自動フィードバックの開発と妥当性検証 2020年4月～2024年3月(20H01292)、2024年4月～2028年3月(24K00095)
- 早稲田大学 LANGX Speaking test (AI搭載のやり取り型英語スピーキングテスト) の開発と妥当性検証 2020年10月～現在まで (『早稲田大学Podcasts: 博士一步前』 Vol. 3 参照)

吉村 宰 先生

- 1999年 4月 岡山大学 教育学部
- 2002年10月 大学入試センター
大学入試センター在籍時に、IRTを用いた1990年度から2004年度の
大学入試センター試験のデータベース（問題公表済みのアイテムバンク）
を作成。
- 2005年 長崎大学 アドミッションセンター
（現：教育開発推進帰国アドミッションセンター）
主として入試に関するコンサルティング業務、その他調査・研究、
広報活動を行っている。

ディスカッション

1. パネリストから電通大CBT報告への質疑応答
2. パネリストへの質問
 - CBTを入試に展開していくことは可能か？
 - それを実現するためには どのようにすればよいか？

仁田先生からのご質問

質問1 露出数関連について

3パラメータモデルを用いること、実行、デバックを評価に含めるようなより詳細なモデルを維持するには、より多くの露出数が必要となるのでは？また、項目バンクを作成するにはより多くの問題数が必要となる。これらを両方クリアするための方法としてはどのようなことを考えているか？

回答

毎年新入生学力調査で700人以上の学生が受検しており、十分なデータを収集できると考えています。

LLM生成AIによる自動作問支援

問題

ある郵便局には、複数の荷物があります。各荷物には特定の重さがあり、トラックは特定の最大積載量を持っています。以下の短冊から適切に並べ替えて、荷物の重さをすべて載せるために必要なトラックの最小数を計算するプログラムを完成させてください。

短冊

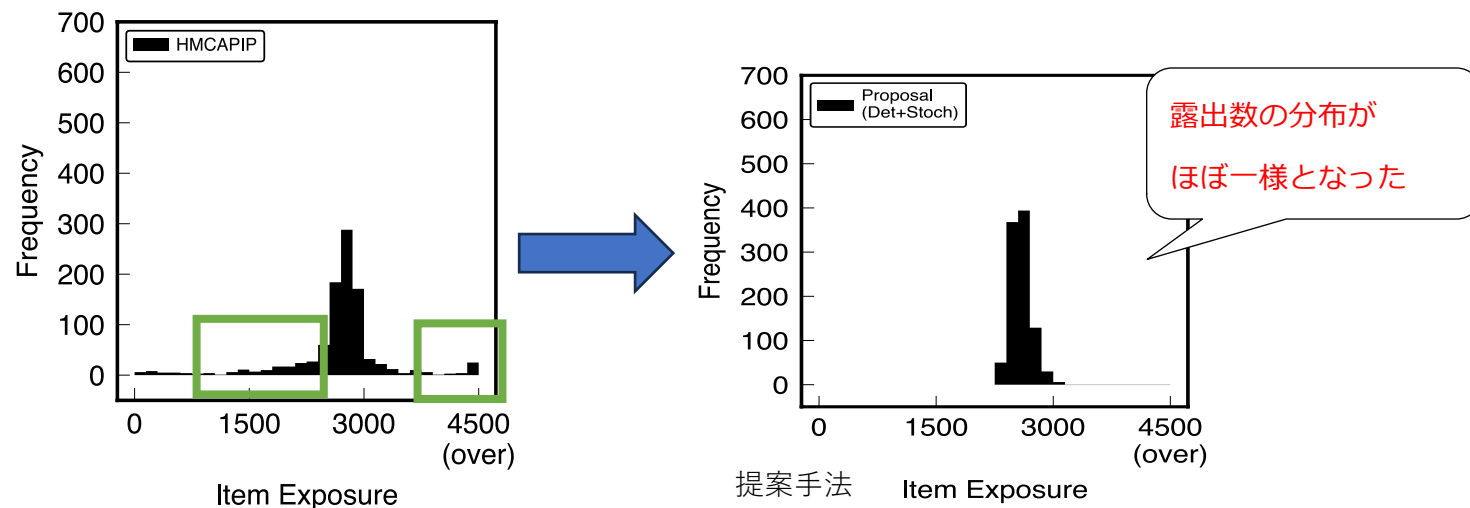
```
1. `def min_trucks(n, packages, max_load):`  
2. `truck_count += 1`  
3. `packages.remove(pkg)`  
4. `current_load = 0`  
5. `return truck_count`  
6. `current_load += pkg`  
7. `if current_load + pkg <= max_load:`  
8. `while len(packages) > 0:`  
9. `for pkg in packages[:]:`  
10. `packages.sort(reverse=True)`  
11. `truck_count = 0`  
12. `packages.append(pkg)`  
13. `if current_load < max_load:`  
14. `for weight in packages:`  
15. `truck_count -= 1`
```

今の時点では人間のチェックが必要であるが現在でも相当のクオリティの作問を実現できる。将来、その場で受検者の能力に適応した問題項目（情報量を最大にする問題）を自動生成することも可能である

露出数の偏り軽減のための整数計画法

$$\text{Maximize } \sum_{i=1}^n \left(\lambda_i x_i - \underbrace{\frac{1}{1+e^{-z_i}}}_{\textcircled{1}} - \underbrace{M_{\{i,p\}}}_{\textcircled{2}} \right) x_i$$

- 露出数の大きさに応じたロジスティック関数による
①決定論的ペナルティおよび②確率的ペナルティを与えた
整数計画法を用いて逐次的にテスト生成



澁本 壱真, 植野 真臣. 項目露出ペナルティを用いた整数計画法による自動並行テスト構成, 統計数理 (2024)

質問2 最低基準の設定

BOOKMARK法などでは、多数の専門家が問題を確認して、合格レベルを決定する方法がとられるが、今回の電通大ではかなり自動的に決める様な方法がとられているように思われる。入試においては、入学定員が決められていることから、これを含むモデルも必要かなと思われる。自動化システムの先生の手応えと入学定員も含むモデルなのかどうか、もし含んでいない場合は今後含めるお考えはあるかどうか、教えてください。

回答

- BOOKMARK法は医師のスキルのような人間の判断を必要とする合格基準設定に有効だと思います。一方、IRTを用いることの長所は1. 精度が高い将来の成績予測と2. 過去入学生との同一尺度上でのスコア比較が可能な点です。今回は、大学で留年しないで就学できることを最低基準としていてIRTを用いた予測でGPA成績によりかなり客観的に決定できます。
- 総合型選抜では留年確率を最低限下げたスコアを足切り点にして第1次審査の参考にしてもらっていますが、思ったより優秀な志願者が多く、第一書面審査ではこの足切り点よりだいぶ上のスコアが最低限の条件であり、このような指標は合否意思決定に有用であると考えています。
- また学校推薦では足切り方式ではないですが、数学CBTでは足切り基準を大きく上回っています。
- 今後は、実際の入学後のGPAデータを検討しながらスコアを総有効活用するかの検討を行っていきます。

澤木先生からのご質問 1

高校レベルの学習がどの程度身についているかを測るうえで、教科書レベルの問題に照準を合わせることは大変望ましく、真正性の高い問題作成ができるのではと考えます。問題作成においては、分野や各出題者間での内容のすり合わせはどの程度必要とお考えでしょうか。つまり「等質なテスト」作成のうえで、アイテムバンクに含める問題間である程度の互換性が必要となると思われるが、スライドに示された細目表をさらに具体化したレベルで、問われる力や問い方、形式などの項目特性（問題形式や作題方法）をどの程度統一すべきか、また項目の品質管理をどのように行うかについて、お考えがあればお聞かせください。（注：入試ですので言及できない部分もあるかもしれませんが、貴学での状況でも、より一般的な話としてでも、可能な範囲でお答えいただけましたら幸いです。）

回答

- プログラミングやデータサイエンス問題はPCIで問題形式が統一されていますが、基礎問題のフォーマットも評定平均の確認のためであり、教科書の演習問題や期末テストでの出題に似せたフォーマットに統一しています。
- 今回、作問委員会で各作問者が作問してきた問題項目を持ち寄って修正点を洗い出し、それを数回繰り返して完成させています。作問者は問題バンク中の問題内容およびIRTを含む統計データを閲覧することができ、どの程度の問題項目が本学生にとって有効かも共有していきます。作問委員会で承認された問題項目は別の問題確認者もしくはアドバイザーによってチェックされます。

澤木先生からのご質問 2

プログラミングやデータ解析等のタスク（目標が示された活動）についての採点方法についてお伺いしたく存じます。比較的時間がかからない基礎的な問題について、例えば二値（「正答」「不正答」）で採点することは理に適っていると考えます。一方で、このような大きなタスクは時間がかかるのと同時に、解答の正確性にも段階があり、CBTを行うことにより解答プロセスに関する情報を取得することも可能とと思われます。その段階ごとに得点を与えることで診断的情報（どこができていてい、どこでつまづいているのかに関する情報）を取り出すことはどの程度お考えでしょうか。入試後のテスト結果活用も視野に入れておられるため、それができれば事後指導のうえで参考となるのではと考えた次第です。

回答

プログラミング問題の診断については、回答プロセスデータと他の下位問題（この問題を解くために正答しなければならない問題）のデータを用いてどこで躓いているかがフィードバックできるようになっています。回答プロセスのデータは数理的に受験者の能力パラメータに吸収されて診断に利用されます。ただし、その用い方は入学後のUEC検定であり、入試では診断は特に行いません。むしろ、テストの測定誤差をなるべく小さくするようにテスト構成を行います。

吉村先生のご質問

<課題>

○共通テスト，個別試験では試験時間の確保が最重要視される（1秒の不足も許されない）。これに対し10分以内ならIRT補正值を用いるという手法は社会的に認められるのか。入試ミスに気をつけろとのお達しが毎年来るが文科省の意見も聞きたい。

○非公開のアイテムバンクについて

・秘匿の限界：受験対策は徹底的に行われる。受験報告→高校でのとりまとめ→某企業がとりまとめ→某オンラインソフトウェアで利用可能（無料）

・作題ミス：作題ミスへの対応はどうなるのか。他の入試区分との整合性についてどう考えてるのか聞かせていただきたい。

○コンソーシアム参加にはアイテムバンクへの問題提供がセットになる。現状で作題に四苦八苦している大学へのインセンティブとして何が考えられるか

吉村先生のご質問への全体的回答

ご指摘のように試験の公平性のルールなどをCBTが順守するのは難しいと考えています。ただ、ペーパーテスト入試のために決められたルールをそのままCBT入試に当てはめること自体難しいのも事実であり、その制約は実質 CBT入試の導入を不可能にするものと考えています。CBT入試のための新しいルールを模索していくことが必要と考えています。例えば、公平性についてペーパーテストではテストの問題内容や制限時間の公平性が重視されますが、CBTでは測定精度の公平性が最も重視され、それらの公平性が満たされていれば問題内容や所要時間は問題とされません。

次ページに各質問への回答を列挙いたします。

回答

<課題>

○共通テスト、個別試験では試験時間の確保が最重要視される（1秒の不足も許されない）。これに対し10分以内ならIRT補正值を用いるという手法は社会的に認められるのか。入試ミスに気をつけるとのお達しが毎年来るが文科の意見も聞きたい。

回答：おっしゃるような公平性を満たすようによりよい方法を模索すべきと思います。例えば、個別に試験中に試験時間を延ばせる機能があれば一番良いとも思っています。ただし、ペーパーテストとCBTは全く違うものなので同じルールを課すことは難しいことをご理解いただかないといけません。説明でも述べたようにCBTの長所は受験生が同じ問題項目のテストを受検していなくてもフェアに評価できることにあり、フリーズした学生が不利にならないように精度高く補正することが可能です。現在のCBTシステムの制約の下ではCBT実施者、受験生の負担を上げない解決法の一つと考えています。

○非公開のアイテムバンクについて

・秘匿の限界：受験対策は徹底的に行われる。受験報告→高校でのとりまとめ→某企業がとりまとめ→某オンラインソフトウェアで利用可能（無料）

回答：受検時にCBTでは問題項目の守秘義務の契約（暴露が発覚すると法的措置もあり得ます）を行っています。最終的に将来の学生が不利益になるので高校側、業者にも問題バンクの趣旨を理解し、問題を暴露する活動に加担することは辞めていただかないといけないと思います。また、社会的にそのような文化が出来上がるとよいと考えています。

・作題ミス：作題ミスへの対応はどうなるのか。他の入試区分との整合性についてどう考えてるのか聞かせていただきたい。

回答：受験者から問題ミスの指摘があった場合は、その場では何もしません。指摘後、作問委員会で問題に誤りがあることが確認されるとその問題への回答を推定から除くルールになっています。今のところUEC検定を含めてそのようなことはありませんでした。

○コンソーシアム参加にはアイテムバンクへの問題提供がセットになる。現状で作題に四苦八苦している大学へのインセンティブとして何が考えられるか

回答：アイテムバンク、システム、データ解析などの共用が目的となるはずですが、コンソーシアムがどのような形で進められるかはこれから議論になっていく予定ですが、入試の精度向上と教員負担減になればよいと考えています。