

## テストのパラダイムシフト

### テスト作成者の求めに テスト研究は応えているか

2026年9月15日 (火)

石井 秀宗  
名古屋大学大学院教育発達科学研究科

## 問題意識

- テスト (定期考査) の作成者は学校教員
- **テストの作成や分析に関して学ぶ機会は殆どない**
  - 過去問, 先輩からのアドバイス等をもとにテストを作成
- 業務の多忙化
  - テストの作成や検討, 分析に使える時間は少ない
- **時代の変化**
  - 人材不足, 以前では教員にならなかったような人も
  - 社会の風潮, 世代の違い
  - AI, ICTの急速な発展

2

## 問題意識

- テストやテスト作成に関する知識やスキル, 能力, 理念などの伝達, 継承が難しくなっている
- AIなど新しいスキル の活用に対応していく必要がある
- **テストに携わる教員にどのようなことを伝えられるとよいのか?**
- **今後のテスト研究で必要とされることは何か?**
- 教員対象の研修会やインタビューなどから考察

3

教員を対象とした, テストの  
作成や分析に関する講義に  
寄せられた意見・感想

## テストの作成や分析に関する講義

- 愛知県内の私立高校に勤務する中堅教員の研修会
  - 2026年6月開催
- プログラムの1つとして開講された, テストの作成や分析に関する講義
  - 参加者は50名程度
- 講義内容 (120分)
  - テストデータから得られる基本情報
  - テスト問題の分析
  - テスト問題作成ガイドライン
  - 良い問題・悪い問題
  - 項目反応理論

5

## 講義内容

- テストデータから得られる基本情報
  - 得点分布, 平均点, 正答率, 選択率, 無答率, 識別指標
- テスト問題の分析
  - 必要性, トレースライン, 項目分析の具体例
- テスト問題作成ガイドライン
  - ガイドラインの説明, 修正例
- 良い問題・悪い問題
  - 要件, 具体例
- 項目反応理論
  - 必要性, 概要, 適用例, 要件

6

• テストに対する認識

- テストの評価、品質管理がなされていないことを怖く思った
- テストは単に点数をつけるものではなく、生徒の学びを理解し、成長を支えるとともに、授業改善につながるものだ分かった
- テストも、生徒の学びや人間形成に資するものであるべき

• 良い問題の理解

- 測りたいものを測れるのが良い問題だというのはその通りだと思った
- 思考力を問ういわゆる凝った問題が良い問題だと思っていた
- 知識の定着を測りたければ、知識問題は良い問題だという話はありがたかった
- 心がけていたことが良い問題の要件だったので、安心した

7

• 問題作成ガイドライン

- ガイドラインを知らなかったが、役に立つと思った
- ガイドラインで言われていることを意識したことがなかった
- 正答率が低いのは、生徒の勉強が足りないからだだけ思っていて、問題に原因があると考えたことは無かった
- フォーマットが得点に影響するとは思っていなかった
- 選択枝を無理に多く作ることがあったので、少なくともいいと聞いて安心した
- よく勉強している生徒が点を取るように「裏をかく問題」「ひっかけ問題」を作っていたが、測りたい能力が違っていた
- これまで、あまりよく考えないで「すべて選べ」問題を出題していた
- とくに理科や数学の場合、設問間の依存関係はどうしてもできてしまう。そういう問題の割合を減らせばよいということか？

8

• 項目分析

- 問題の性質を統計的に分析する方法があるとは知らなかった
- これまでは、テストの平均点、最低点、最高点、各問題の正答率しか見ていなかった
- これまでは、正答率のみに注目し、易しい問題と難しい問題の量を調整して、平均点を一定水準にすることしか考えていなかった
- 識別力を見ることの重要性、有用性を理解した
- トレースラインで項目の特徴、とくに識別力が分かるのは良いと思った
- 識別力の高い問題の作り方を知りたい
- テスト問題の検討、生徒の学びの理解、授業改善などに項目分析は役に立つと思った
- 誤答枝から生徒の学びを理解できるのは良いと思った
- 項目分析は有用だと思うが、時間がなく理想論

9

• IRT

- 項目反応理論 (IRT) というものを初めて聞いた
- IRTという言葉は聞いたことがあったが、中身は知らなかった
- 問題が異なっても比較できるということに魅力を感じた
- 定期考査や個々の入試にあまり向いていないのは残念に思った
- 過去問文化が発達している本邦で、大学入学共通テストをIRT化できるかは疑問に思った

• AI・ICT

- AI採点を導入しているので、分析はしやすいと思った
- そもそも回答をデータ化することに時間がかかる
- デジタル採点の導入により、個々の生徒へのフィードバックが薄れている
- 知識はAIで補えるので、AIでは代用できない能力を育成すべき

10

• 作成環境・その他

- テストの作成や分析の話を聞く機会はいまだなかった
- 測りたいことは何かを明確にすることの必要性を理解した
- これまではどういう問題にするかということだけに意識が向いていた
- 検討会で意見を言ったりして、テストを作成するのは有効だった
- 各問題がどの観点に関するものか、観点の割合をどの程度にするのが適切かなど、観別評価との対応づけが難しい
- 教科指導が十分できない中で平均点を上げることを求められると、とても易しい問題を出すことになり、識別力の低い悪い問題が多くなる
- 時間がなく、採点しやすい問題を作ってしまう
- 良い問題・悪い問題の要件がいろいろあって、こういうのが良い／悪い問題ということがパッと分かる例を示してほしい

11

教員を対象とした  
インタビュー

## 高校教員へのインタビュー



- 対象
  - 東海地方の高等学校に勤務する教員 6 名
  - 勤続年数：7～35年
  - 2026年7～8月に実施
- インタビュー内容
  - テスト作成について教わった経験、内容
  - テスト作成の体制
  - 作成時期
  - 1 回の定期考査で作成するテスト数
  - 心がけていること
  - 困っていること
  - テスト作成におけるAI, ICTの利用 など

13

## テスト作成について教わった経験、内容



- 6 名中 5 名が、教わった経験なし
  - 経験のある 1 名は、教員になる前に働いていた企業（教育産業）で、テスト作成について教わった
- 最初は、過去問、先輩からのアドバイス、授業での生徒の反応などを頼りに作成
  - 内容、構成、形式、配点など
  - 慣れてくるのは 3 年目あたりから
- 検討会で意見をもらう
  - 出題範囲、難易度
  - 問題文の一意性、適切性
  - 正答、誤答の確かさ
  - 出題の意義や意図の議論はわずか

14

## テスト作成の体制



- 学校や教科科目によって異なる
  - 1 名の教員が 100 点分（全部）作成
  - 作成担当でない科目は確認を担当
  - 同じ科目を教える複数の教員で大問を分担して作成
  - 100 点分作成するほうが作りやすい、調整しやすい
- 検討会で検討
  - 同じ科目を担当する教員が集まりテスト問題を検討
  - 他の教員が作成したところは意見しにくいことも
  - 時間がなければ回覧
  - 検討会をしなくなった、時間がない、予定が合わない
  - 単独で担当している科目は、同僚に確認してもらうだけのことも

15

## 作成時期、作成するテスト数



- 考査の約 2 週間前から作成
  - 単独で担当している科目は 1 週間前くらい
  - まだ授業終わっていない、部活もある
- 1 回の考査で 2 ～ 4 個作成
  - 常勤、非常勤込みで、年間スケジュールを立てる
  - 学年ごと、コースごとにテストを作成
  - 異なるコースに共通問題を入れ、比較することも
  - 追試問題を作成することも
- テスト数の増加
  - 学習指導要領の改訂
  - すべての教科科目で毎回テストを実施
  - 検討会を設定できない、作成者の重複

16

## 心がけていること



- 平均が 50～60 点になるようにする
  - 赤点にならないように
  - 勉強しなくても高得点が取れることがないように
  - 勉強へのモチベーションがなくならないように
  - 平均点が高かったり、100 点が多いのを歓迎しない雰囲気
- 過去問を避ける
  - 不公平にならないように
- 練習した問題、例題の数値を変えた問題を出題
  - 勉強すれば点が取れると思えるように
  - 努力が報われるように
  - 問い方を変えると答えられなくなってしまう

17

## 心がけていること



- 考えさせる
  - 記憶だけで正答に至らないように
- 書かせる
  - 文で記述できるまでの理解を促す
- 別解がある
  - 解法は 1 つでないことを伝えたい
- 違いを分かっているかを問う
  - 一見細かくても重要であるところを押さえているか

18

## 困っていること



- 時間の捻出が難しい
  - 問題の中身を検討する時間が十分とれない
  - 実施後に振り返る時間がない、得点分布・平均点・正答率の確認程度
- 担当教員の力量不足
  - 問いに不備あり、一意でない、正確性に欠ける
  - 担当クラスの平均点が教員の評価になるという意識があることも
- 若い教員に意見を言いにくくなっている
  - ハラスメント、「ではあなたが作って下さい」と言われてしまう
  - コスパ・タイパ重視、意義や意図よりも作成・採点しやすい問題
  - 自分の作った問題を見せることで学んでもらう
- 受験科目でない生徒のモチベーションが低い
  - クラス内で二極化すること

19

## 困っていること



- 観点別評価
  - 観点と問題の対応をどう決めるか
  - 割合をどう設定するか
- 受験対策にもなるテストをどう作るか
  - ある程度入試を意識してテストや授業を構成せざるを得ない
  - 難しくなり平均点が下がる
- どの程度、どの入試に寄せるか
  - 国立、私立、共通テストでだいぶ異なる
  - 会話型は正答率低下
- 答案読み取り機器の弊害
  - 採点はラクになるが、作問やフォーマットに制約がかかる
  - 解答形式、答案枚数

20

## AI, ICTの利用



- 科目や教員によって異なる
  - サンプル問題の作成には使える
  - YouTubeなどに上がっている解説動画を参考にする
  - AIが作った問題を定期考査に使うには課題あり
- 英語
  - 利用できる場面多い、だいぶラクになる
  - 単語数や難易度を調整して文章を作成
  - 例文作成
  - 記述問題、内容一致問題の確認
- 数学
  - 解答が公開されていない入試問題の解答確認
  - 作成した問題の解答確認

21

## AI, ICTの利用



- 理科
  - それらしい問題は作れる
  - 図を描かせてみたが、正確でなく使えなかった
  - 教科書や参考書の図を利用
- 国語・社会
  - 慎重、注意深さが求められる
  - 試験範囲・教材の範囲で作問するのは難しい
- 懸念事項
  - 意図を良く反映した問題にはなりにくい
  - 問題作成スキルが低下する
  - 問題を作る楽しさが低減する

22

## 見えてきたこと

## テスト作成に関する課題の現状



- テストの作成、分析等に関して **学ぶ機会**は殆どない
  - 依然として変わらず
  - 研修等でも減多に扱われない
- テストの作成、検討、分析に使える **時間**がない
  - 教員の多忙化
  - テスト数の増加
- 人材不足、**伝達・継承**が難しくなっている
  - 力量不足、そういう人に限って人の意見を聞かない
  - 検討会の消滅
  - 指摘しにくい雰囲気
- AI, ICTの急速な発展
  - 対応は科目や教員によって異なっている

24

## テストに携わる教員に伝えたいこと



- テストについて考えることが必要という意識
  - 品質の維持管理の必要性, 重要性
- **テストを行う目的の明確化**
  - テストの作成, 運用における規準となる
  - 教員自身のモチベーションの維持
- テストの作成や分析に必要な情報, 知識, スキル
  - 有用性の理解, 具体例による説明
  - **適正な評価, 生徒の学びの理解, 授業改善につながる**
- より多くの研修会, チュートリアル等が実施されることを期待
  - テストに関するものは減多に実施されない

25

## テスト研究で必要とされること



- テストにおけるAIやICTの利用
  - 既にいろいろなされている
  - 正確な図の作成
  - 試験範囲, 教材の範囲に限定した問題の作成
  - 大規模テストと定期考査の違い
  - 科目や教員個人の技量の違い
  - **技術がテストを制約してしまう問題**
  - **作成者の作題能力の変化・変容**
- テスト作成者の育成
  - どのように育成したらよいか
  - **何を, どのように, 伝えるか**
  - 教えにくい風潮, コスパ・タイパ重視
  - 教員の質の低下

26