



## Introduction

- Predictability in exams is of interest to Ofqual
- Constant noise in the media that some exams are too predictable – social media may suggest the opposite!
- Overly predictable assessments can:
  - narrow the taught curriculum
  - encourage use of pre-prepared answers
  - threaten validity of assessment
- However, some predictability is necessary – candidates should:
  - know what is expected of them in an exam – e.g. paper layout/structure
  - understand what each question requires them to do to answer well – e.g. command words, marking criteria

## Previous research

- Several qualitative reviews of papers, evaluating several aspects of predictability:
  - Baird and team – Irish Leaving Certificate (Baird, Caro and Hopfenbeck, 2016)
  - Ofqual (2008) – A levels and GCSE
- Rich data, but:
  - labour intensive
  - reliant on standardisation of experts (their use of a rating scale)

## Current research

- Aim:

- to define the **factors** which can affect the (unwanted) predictability of papers

- Devised a framework of predictability factors – used a **process** of actually predicting future questions to see what underlies these predictions

- Predicted papers constructed and compared to actual live summer 2017 papers:

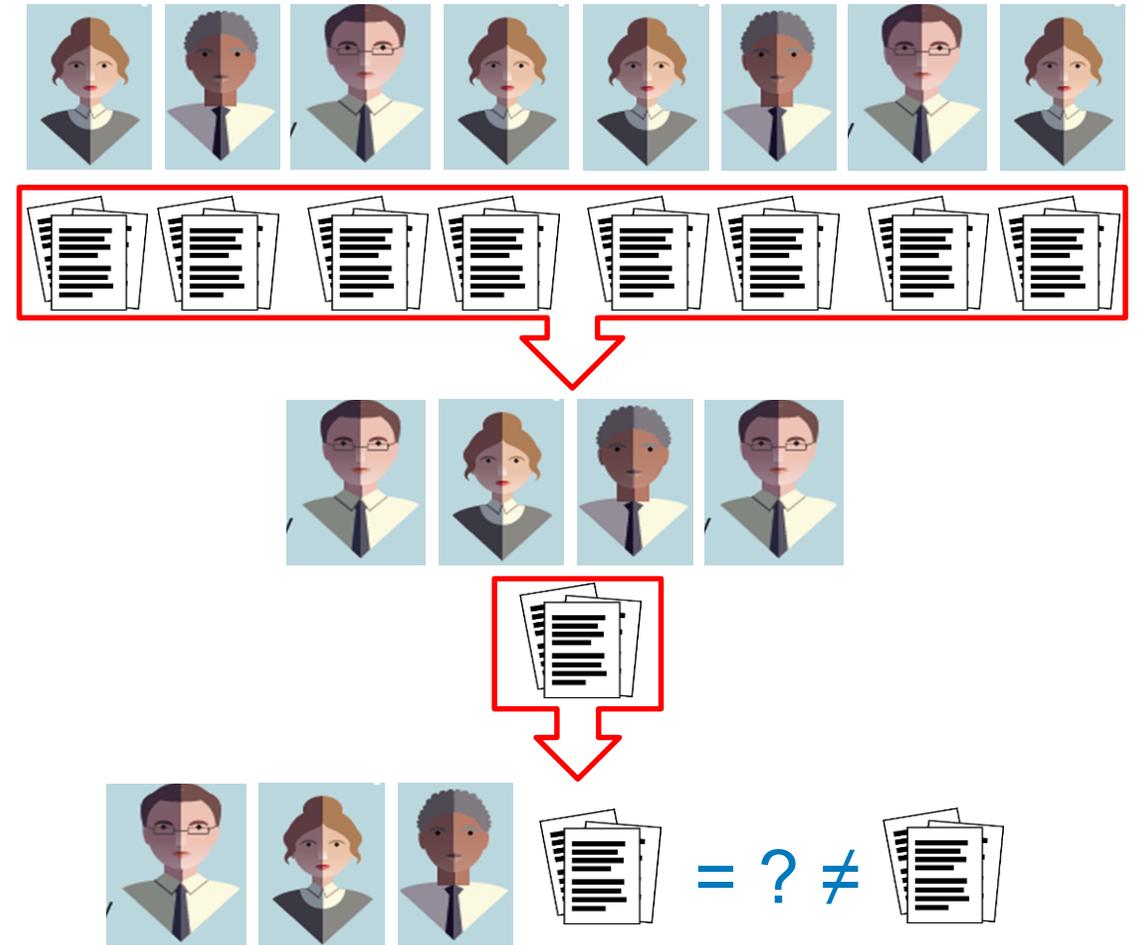
- allows some comparison of paper predictability within- and between-subjects

- accuracy of predictions can potentially validate the predictive power of the identified factors

# Design

## ■ Three stage study:

1. Independent predictions of future questions by teachers
2. Meeting to pick the “best” of the predictions and construct predicted papers
3. Review of accuracy of predictions against summer 2017 papers



## Papers selected

- Study carried out on papers from 6 syllabuses
- AS Government & Politics, GCSE History, A level Psychology – 2 exam boards per subject
- In total, 9 papers were used, either one or two per syllabus, adding up to 2-3 hours of examination time

# Design

## ■ Three stage study:

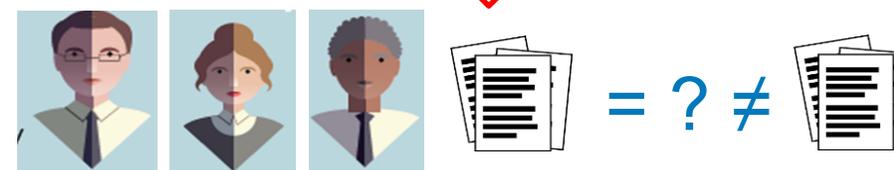
1. Independent predictions of future questions by teachers



2. Meeting to pick the “best” of the predictions and construct predicted papers



3. Review of accuracy of predictions against summer 2017 papers



## Stage 1 - Independent predictions

- 55 teachers recruited: 7-11 per syllabus
- Materials provided:
  - 4 years of summer exam papers (2013-2016) or, Psychology (new syllabus) – sample assessments and practice/specimen papers (one or two)
  - Specification document outlining content
- Pro-forma sent out
  - Contained a draft (not exhaustive!) list of possible factors which may inform predictions to get teachers into right mindset
  - Predictions for each question slot on the paper (where fixed structure)
  - Set of predictions for each section on the paper (where flexible paper structure)
  - Reasoning and confidence for the predictions – factors which influence prediction and allow narrowing down of the possibilities

## Results

- Stage 1 predictions coded against a framework of factors generated iteratively by repeat coding runs and discussion
- Provisional - only single coded at present, full double coding to come

## Framework factors

- Factors related to appearances on past papers – frequency, cycling, not appeared for a while, topic appearance in a different form
- Specification document related - importance of question/topic, content listing narrowing type of question and wording
- Appropriateness of topic for the type of question including size, difficulty, differentiation, wide (efficient) coverage of content
- Factors related to logic of paper – coverage of topics, logical flow of questions, distribution of types of questions, balance
- External factors such as textbook coverage, language in textbooks, practice questions, sample assessments, availability of sources, topicality
- Syllabus lifetime – less predictable questions at end of life, likelihood (or not) of new topics on new syllabus

# Top 5 coding categories (% of total coded)

## ■ Government and Politics AS:

- Non-appearance in recent years (21%)
- Topicality (16%)
- Patterns of question cycling (13%)
- Alignment of wording to specification (7%)
- Logical order of questions (7%)

## ■ History GCSE

- Non-appearance in recent years (19%)
- Patterns of question cycling (17%)
- Topic fits position on paper (10%)
- Topics in textbooks (7%)
- Content coverage (6%)

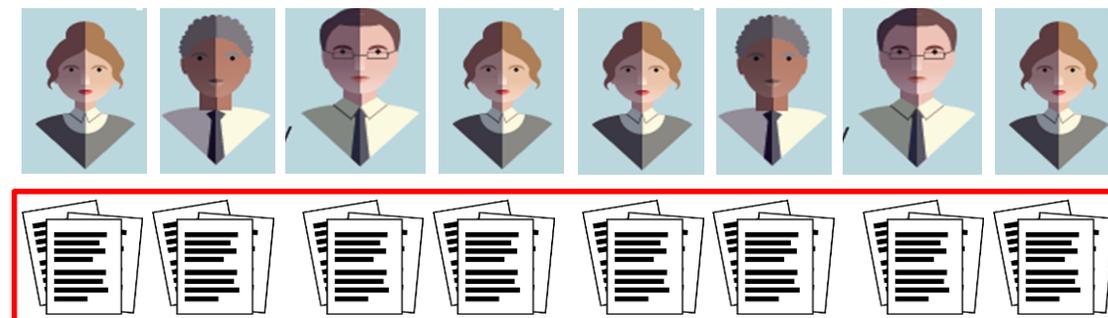
## ■ Psychology A level

- Non-appearance in recent years (12%)
- Alignment of wording to specification (12%)
- Question type often comes up (10%)
- Question type fits position on paper (9%)
- Example questions in SAMs/textbooks (9%)

# Design

## ■ Three stage study:

1. Independent predictions of future questions by teachers



2. Meeting to pick the “best” of the predictions and construct predicted papers



3. Review of accuracy of predictions against summer 2017 papers



## Stage 2 - Predicted paper meeting

- Four teachers per syllabus attended meeting to construct “best guess” paper
- Provided with same materials plus a summary of the predicted questions from each first stage participant
  - Not the rationales as this would be too much information to sift
- Predicted both their own syllabus (**‘home’**) and the other syllabus in the subject (**‘away’**)
- Encouraged to reach consensus on single question for each slot, or set of questions giving appropriate section mark total (Psychology)
- Discussion recorded to determine factors behind their choice

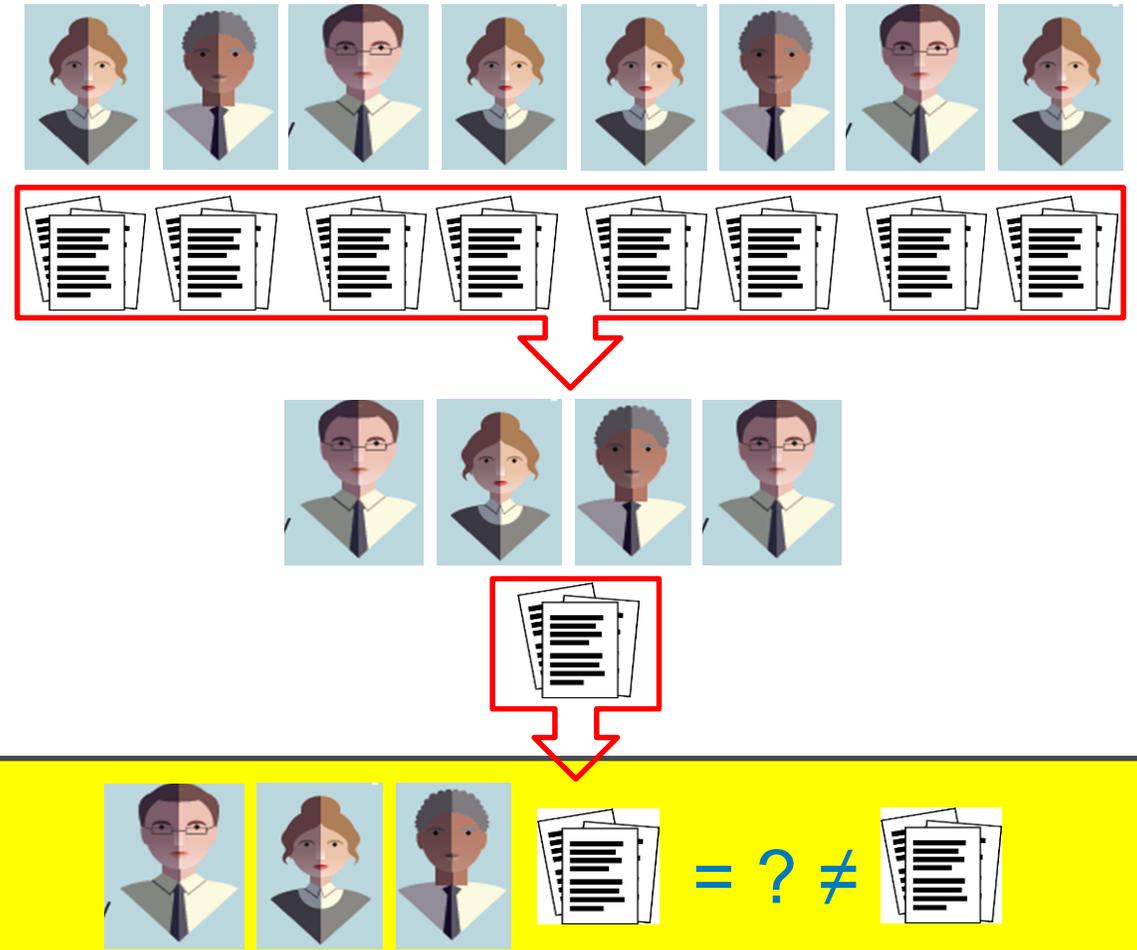
## Prediction process observations

- Some questions were felt to be more predictable than others due to the size/depth/nature of the question restricting possible content
- The usual process was to start with these questions – often the highest tariff questions - and work around them
- Working at whole paper level led to a puzzle kind of approach – careful fitting of appropriate topics into slots on the paper, addressing
  - Some core (vital) content that had to be included
  - General content coverage (depending on how fully each paper covered the content)
  - Difficulty and differentiation
  - Most importantly - use of past papers to narrow down this allocation of content

# Design

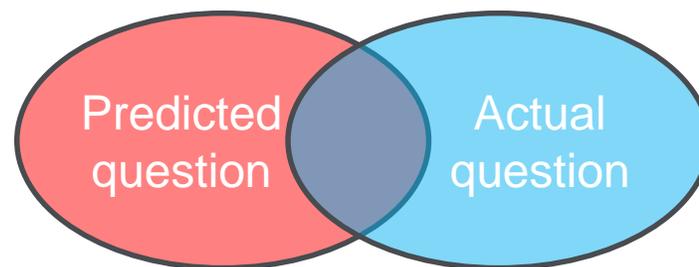
## ■ Three stage study:

1. Independent predictions of future questions by teachers
2. Meeting to pick the “best” of the predictions and construct predicted papers
3. Review of accuracy of predictions against summer 2017 papers



## Stage 3 - Evaluation of predictions against summer 2017 papers

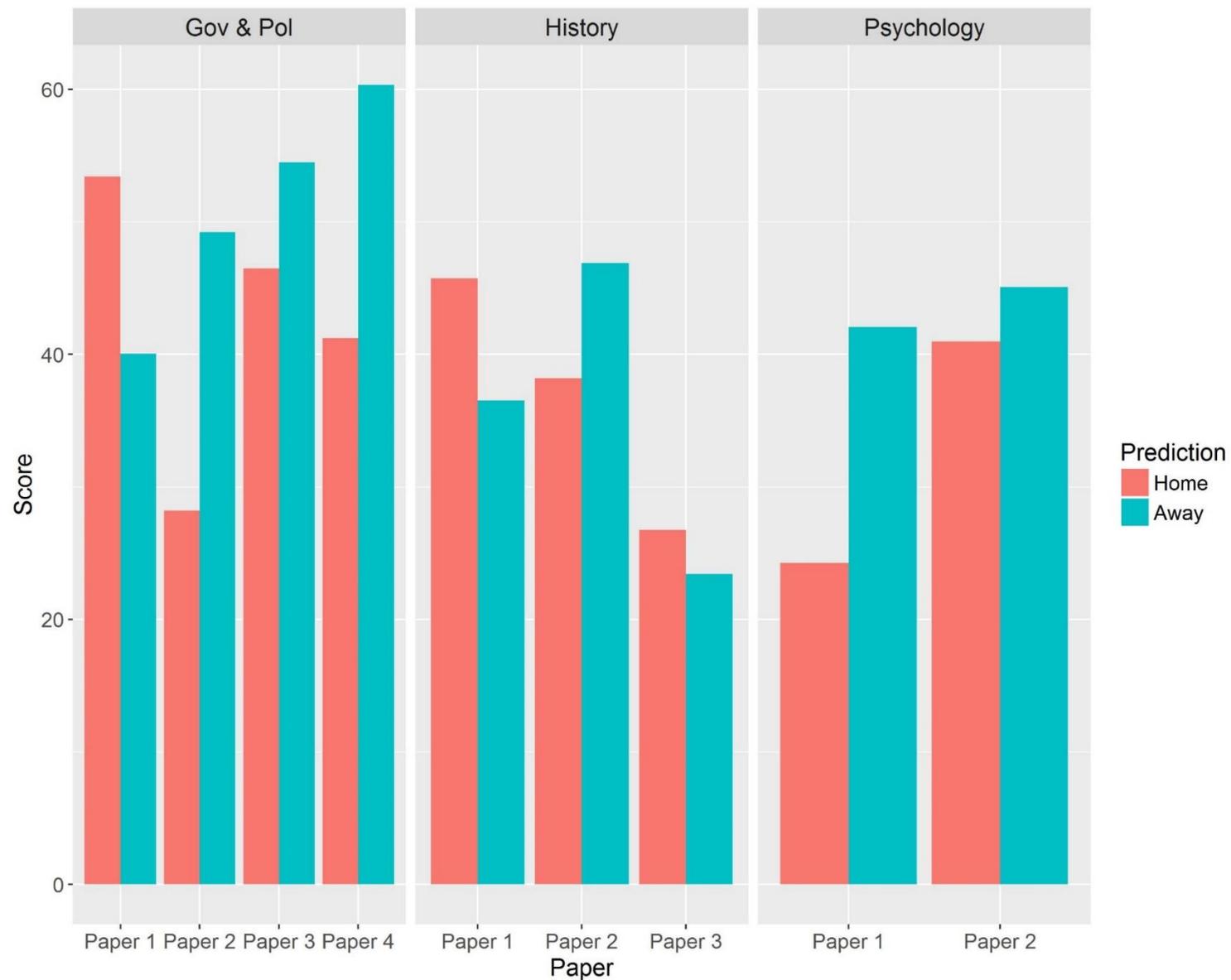
- 2-3 experts per subject evaluated overlap of predicted and actual papers
- Asked to assume that a candidate knew only what had been assessed on the predicted paper – how much of the actual paper could they answer?
- Given some guidance about outcome space – skills and knowledge required to give an answer



- Overlap of ANY question on predicted paper with each actual question in turn
- Score of 0-10 - averaged across experts
- Weighted by question tariff as proportion of paper total
- Summed to give an overlap score 0-100 for the actual paper?

## Accuracy of predicted papers

- Slightly higher prediction agreement for Government & Politics (47%) against History (36%) and Psychology (38%)
- Away board predictions more accurate (44% vs 38%)!



## Further work / questions

- Once properly double-coded, can relate factors to accuracy of prediction – which factors most associated with accurate or inaccurate prediction
- When does a certain degree of predictability become too much?
  - Possibly relate to our stage 3 metric – Psychology was thought to be suitably unpredictable

## Conclusions

- Use of this whole-paper prediction process replicates process an individual paper-setter may go through - should reveal most factors which can be used to narrow down predictions
- Most obvious difference between papers with fixed structure (same question tariffs each year) and unstructured papers – but confounded by newness of the unstructured paper syllabuses
- Teachers using wide variety of factors to narrow down their predictions
- Wide range of confidence (% certainty from single digits to over 90%) – much due to individual personality

## Questions?

- Email at: [stephen.holmes@ofqual.gov.uk](mailto:stephen.holmes@ofqual.gov.uk) and [nadir.zanini@ofqual.gov.uk](mailto:nadir.zanini@ofqual.gov.uk)