Sequences with stopping rules

Dr John Little, Sarah Gott
and Professor Peter Tymms
Centre for Evaluation and Monitoring
Durham University

Defining sequences with stopping rules

A ***sequence with stopping rules*** is an assessment whereby;

- items are arranged in order of increasing difficulty
- stopping rules are invoked depending on the performance of the pupil on items in a section of an assessment.

The ***stopping rule*** may be invoked depending on;

- the number of items wrong in a row OR
- the number wrong in total in the sequence.

CEM
Centre for Evaluation & Monitoring

Background

- Despite the widespread use of SWSR we are unaware of any in depth investigation of the technique.
- The PIPS assessment, administered to 3 million children over 20 years, uses a series of SWSRs.
- The BASE assessment, introduced this year, also uses SWSRs.

The purpose of an assessment and its interpretations

- An efficient and accurate measure of pupil performance – computer adaptive test?
- An assessment for a clearly defined developmental stage in which misconceptions are to be addressed for formative purposes – flat test?
- Or an efficient test which can be used for formative assessment and provides a measure of pupil performance – sequences with stopping rules (SWSRs)?

Aims of the research

For sequences with stopping rules, what is the impact of different stopping rules on;

- the assessment experience for the pupil (difficulty of hardest item seen compared with pupil ability) and
- the errors in measurement of pupil performance?

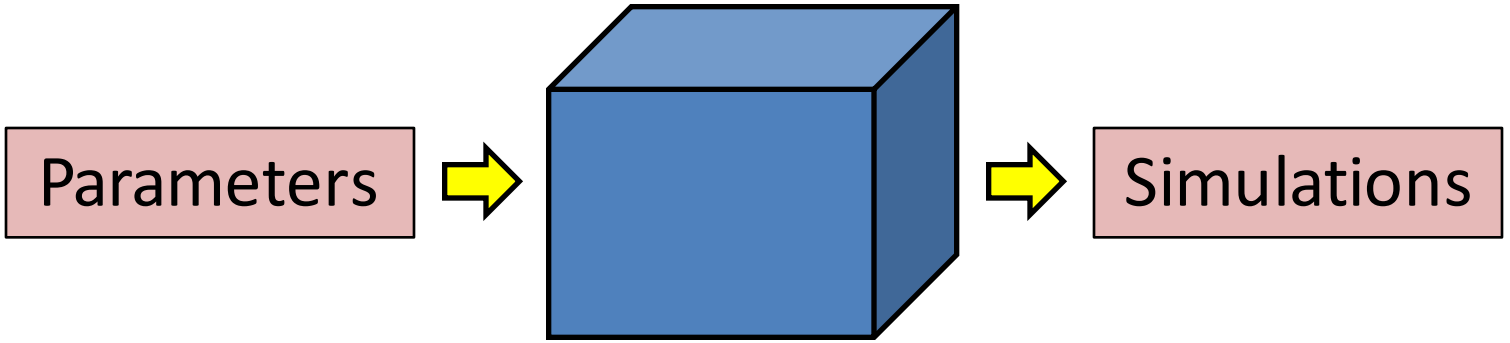
The BASE Assessment

- Data is from a baseline assessment of pupils' developmental stage at the start (and end) of reception in mathematics and literacy.
- The mathematics assessment is formed of sections, one of which is ***Number Recognition***.
- Items in each section are arranged in order of increasing item difficulty.
- The current stopping rule is 3 items wrong in a row, or 4 wrong in total in that section.

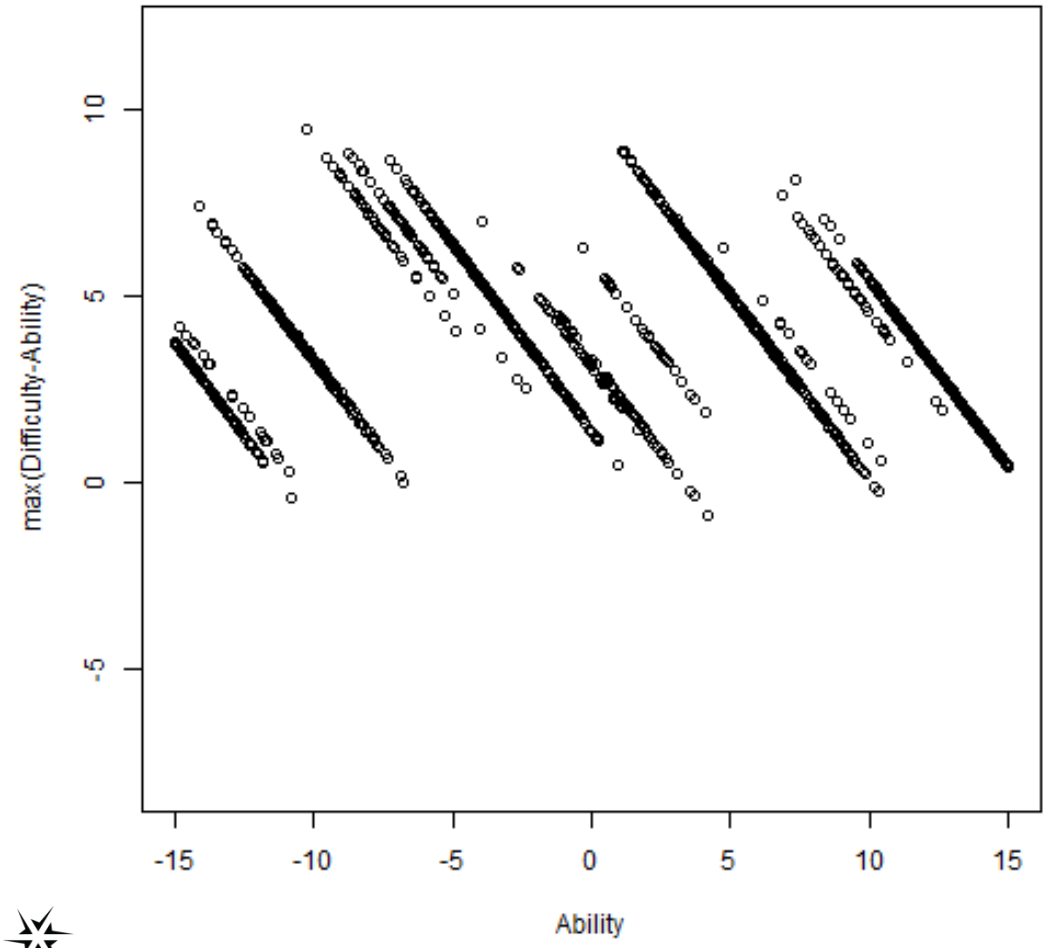
The BASE Assessment

CEM
Centre for Evaluation & Monitoring

Simulation methodology



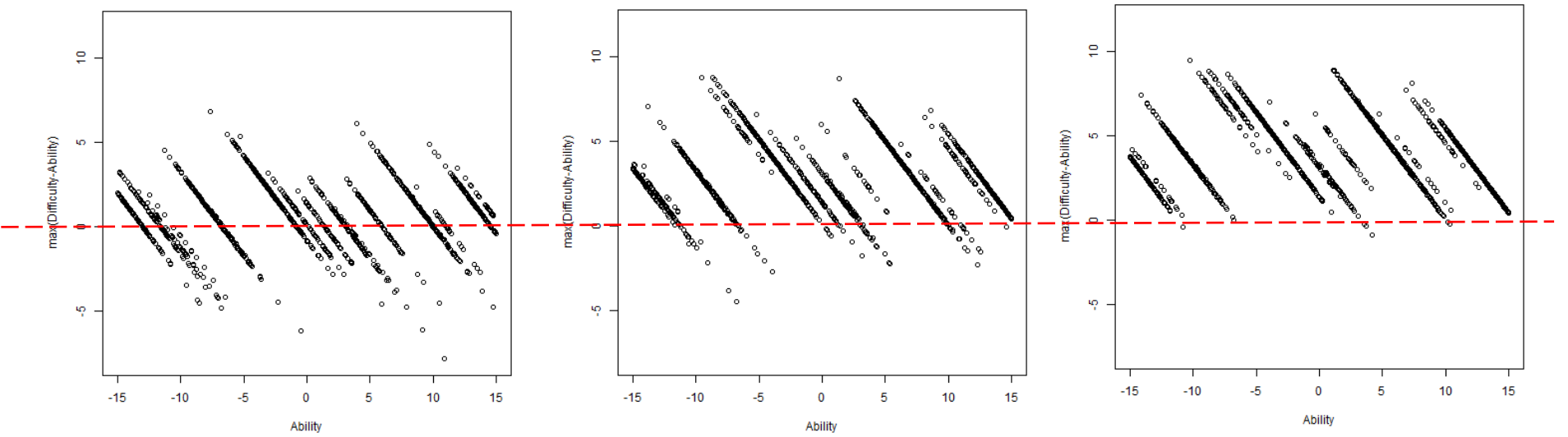
Difficulty – ability (max)



Number	Difficulty
1	-12.98
2	-11.58
4	-11.22
3	-11.58
5	-10.57
8	-6.73
7	-8.41
6	-6.7
9	-6.61
20	-0.79
12	0.14
15	1.41
11	-0.61
23	3.1
45	3.12
96	3.34
300	5.98
231	10.03
579	9.47
996	10.1
1000	11.06
4231	14.59
1049	15.45

CEM
Centre for Evaluation & Monitoring

Number of items wrong in a row and pupil experience



1 wrong

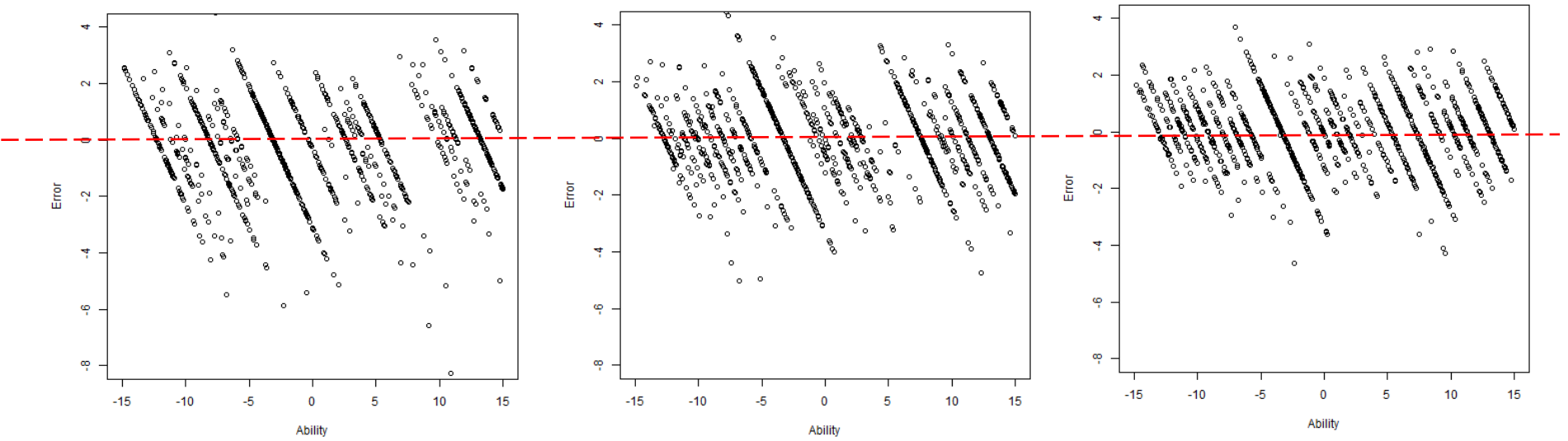
2 wrong

3 wrong

The fewer items wrong in a row, the better the pupil experience.

CEM
Centre for Evaluation & Monitoring

Number of items wrong in a row and errors in measurement



1 wrong

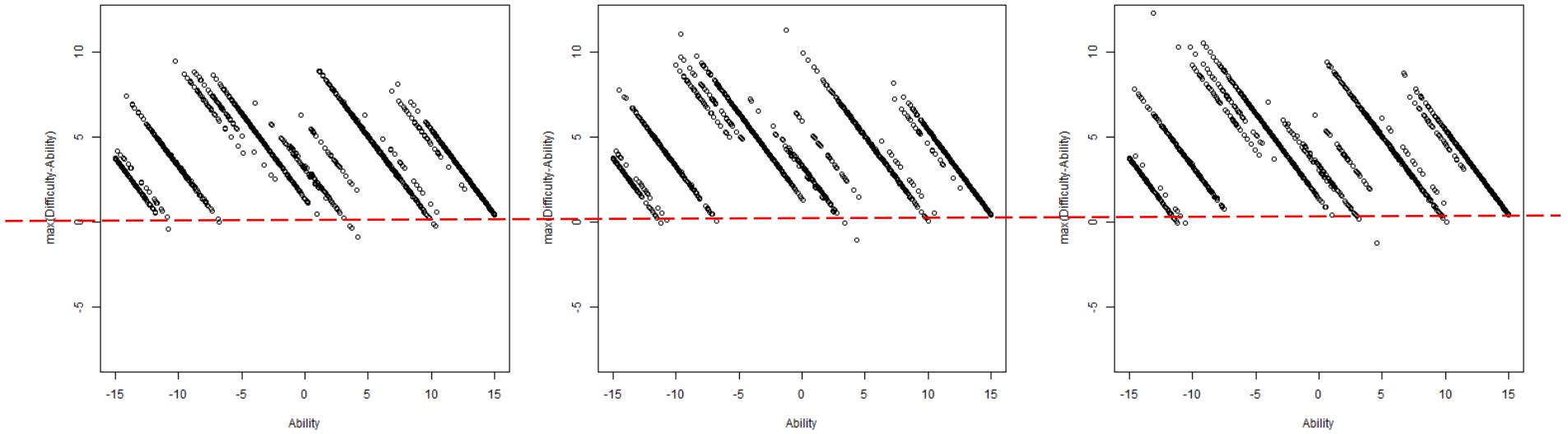
2 wrong

3 wrong

The number of items wrong in a row has minimal effect on errors.

CEM
Centre for Evaluation & Monitoring

Total number of items wrong and pupil experience



4 wrong

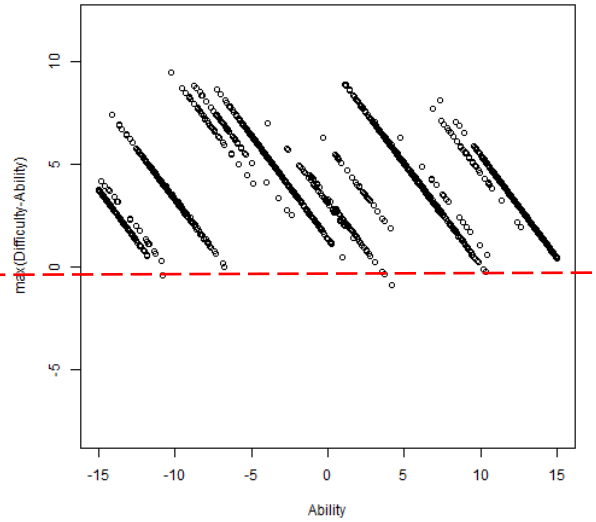
5 wrong

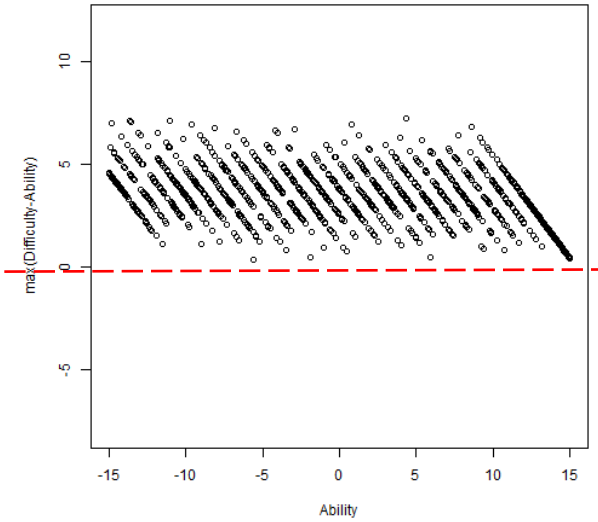
6 wrong

The total number of items wrong has minimal effect on pupil experience.

CEM
Centre for Evaluation & Monitoring

Distribution of item difficulties





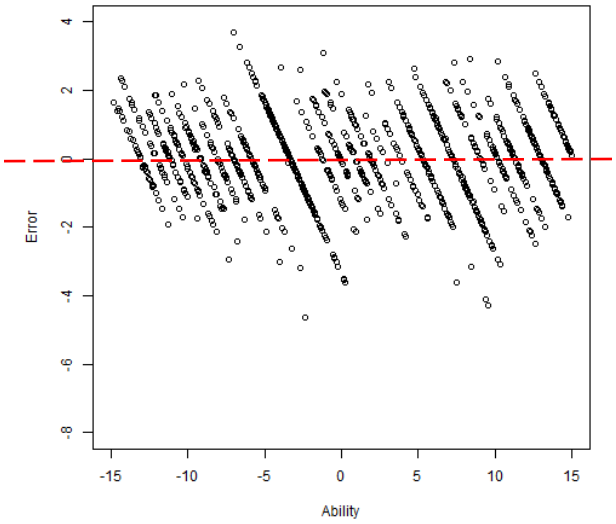
Unequal

Equal

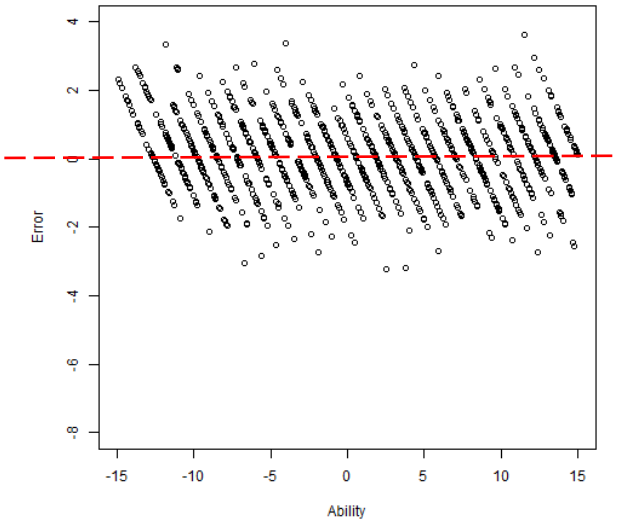
A more equal distribution in item difficulty improves the pupil experience.

CEM
Centre for Evaluation & Monitoring

Distribution of item difficulties and error in measurement



Unequal



Equal

A more equal distribution in item difficulty has minimal effect on error in measurement.

CEM
Centre for Evaluation & Monitoring

General comments for creating SWSR

- There is no simple SWSR assessment design. The interpretations to be drawn and the errors deemed permissible all influence design.
- More conservative stopping rules improve estimates less than might initially be expected for the data in our study.
- SWSR has the advantage that pupils all have the opportunity of answering the same questions, depending on their abilities, in an efficient manner. Hence fair comparisons may be made about their learning.

Future developments

More research is required to establish guidelines for designing assessments using SWSR.

Future research includes application of the simulation methodology to investigate optimal stopping rules for other sections, or assessments, with;

- different difficulty profiles or distributions,
- different target audiences,
- different interpretations such as screening tests or selection tests.