

**AEA-Europe Annual Conference, 9 November 2018, Arnhem/Nijmegen**

# **The big crunch: combining sources of evidence in standard setting**

**Dennis Opposs (Ofqual) and Kristine Gorgen (Oxford University)  
with Tina Isaacs (UCL IOE) and Lena Gray (AQA)**

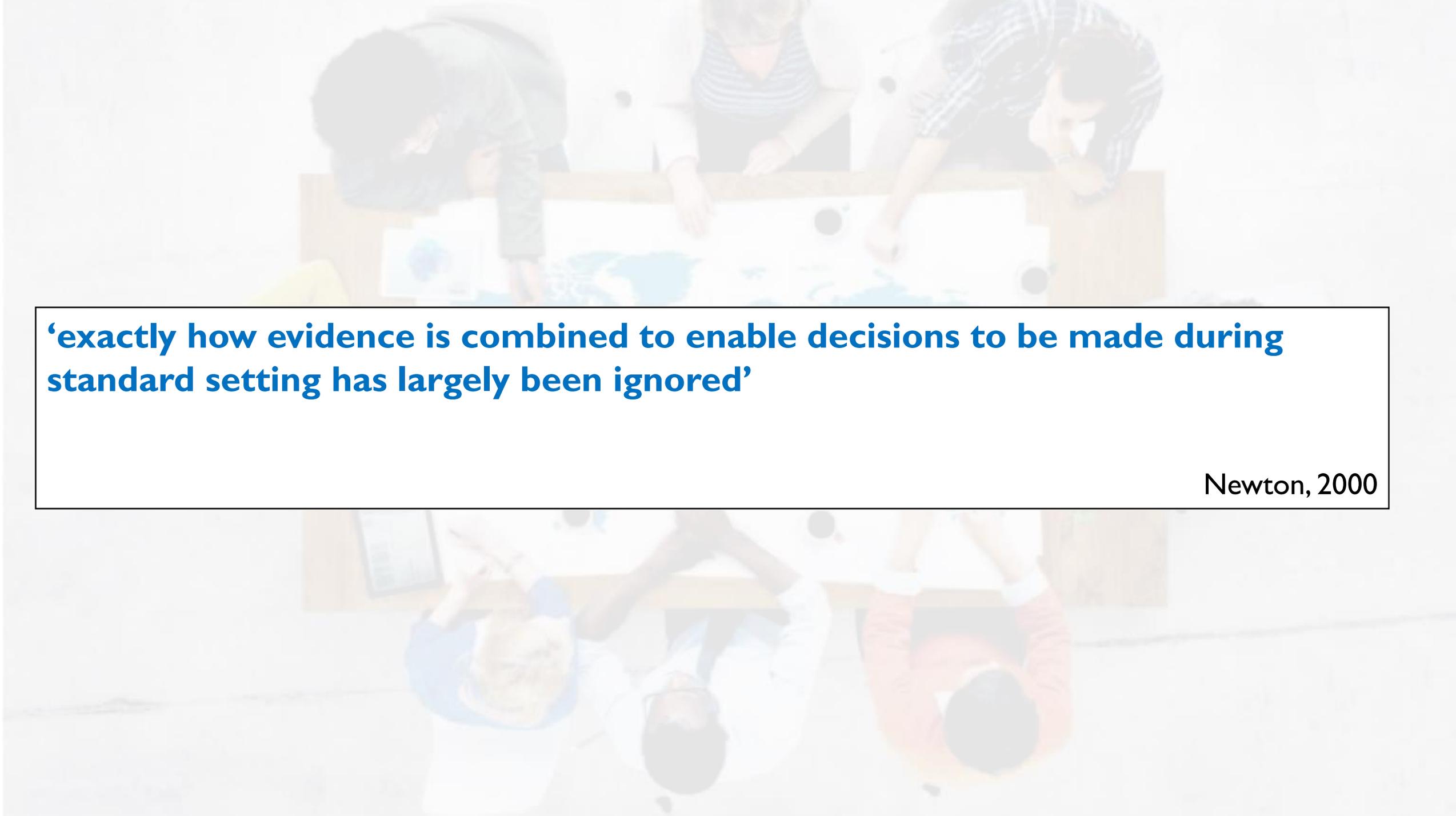


# EXAMINATION STANDARDS



How measures and  
meanings differ  
around the world

Edited by Jo-Anne Baird,  
Tina Isaacs, Dennis Opposs and Lena Gray



**‘exactly how evidence is combined to enable decisions to be made during standard setting has largely been ignored’**

Newton, 2000

## Atomistic methods

**Judgement about the examinations primarily involves making decisions about how well examinees might perform on individual items.**

Bookmark  
Angoff and variations  
Direct Consensus  
Nedelsky  
Ebel

For example, the US Advanced Placement® and the Swedish National Tests.

## Aggregate methods

**Judgement makes major use of the quality of examinees' responses or their marks allocated when sitting a whole examination paper.**

Contrasting Groups  
Borderline Groups  
Body of Work  
Awarding

For example, England's A levels and the South African National Senior Certificate.

## Statistical methods

**Methods which largely or wholly apply statistical techniques to the marks from students' responses.**

Item Response Theory  
Norming  
Scaling

For example, Chilean PSU and Georgian Unified National Examinations.

Gregory J. Cizek  
Michael B. Bunch

# *Standard Setting*

A Guide to  
Establishing  
and  
Evaluating  
Performance  
Standards  
on Tests



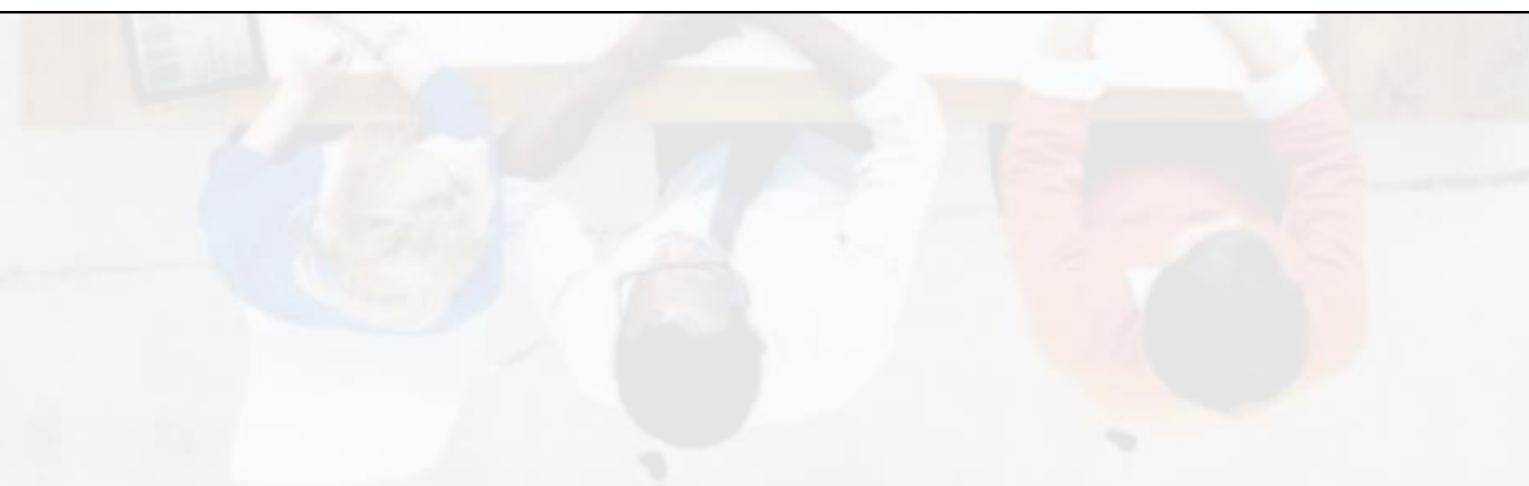
<p><b>Examinee centred</b> v <b>Test centred</b></p>	<p><b>Examinee centred:</b> judgements about whether real examinees show the necessary standard; could also be called 'holistic'.</p>	<p><b>Test centred:</b> each item or collection of items is considered and a judgement made of how a hypothetical examinee would perform.</p>
<p><b>Holistic models</b> v <b>Analytic models</b></p>	<p><b>Holistic models:</b> achievement or skill is assumed to be highly integrated.</p>	<p><b>Analytic models:</b> achievement can be assessed using relatively small parts of performance.</p>
<p><b>Norm referenced</b> v <b>Criterion referenced</b></p>	<p><b>Norm referenced:</b> performance standards are established with respect to the relative standing of examinees from a relevant population.</p>	<p><b>Criterion referenced:</b> performance standards are based on the level of knowledge, skill or ability necessary for a specified purpose and cut scores are established with respect to that level.</p>

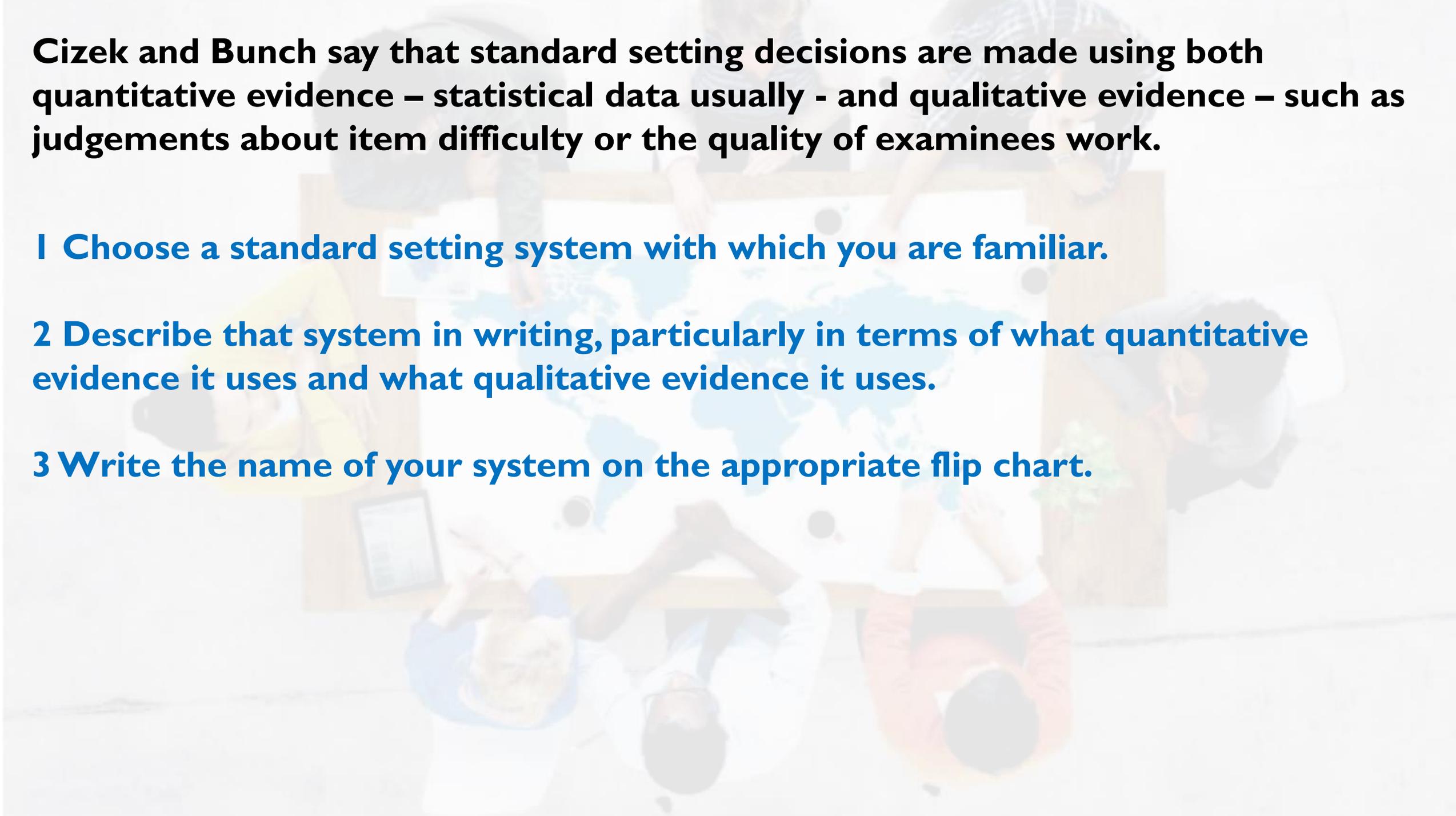
(Cizek and Bunch 2007)



**‘the demands and nature of standard setting in practice compel us to conclude that no simple distinctions between methods can be made and that well-conceived and implemented standard setting must recognise that any procedure requires participants to rely on both dimensions to effectively carry out their task’**

Cizek and Bunch 2007



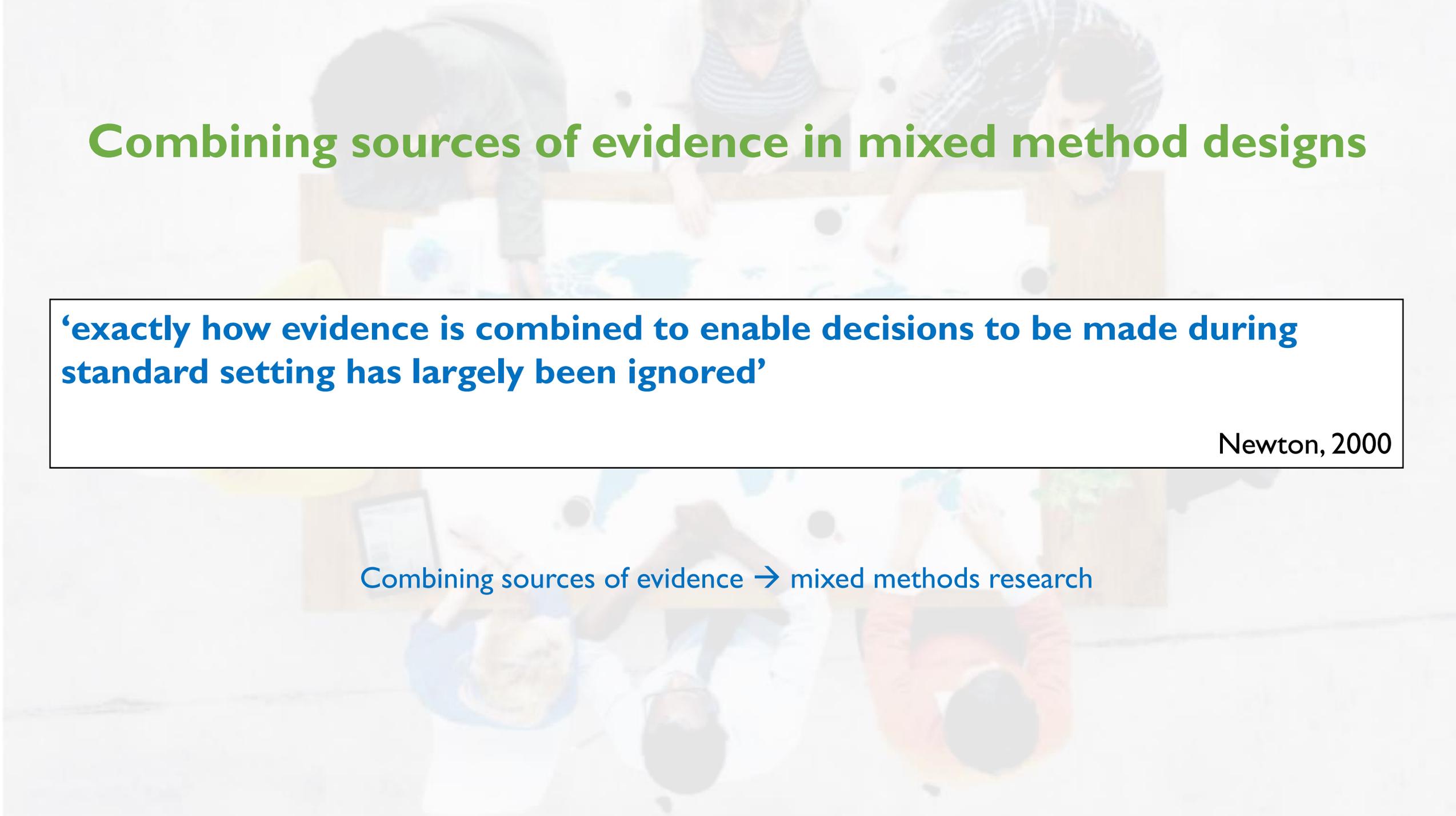
A group of people are sitting around a large wooden table in a meeting room. On the wall behind them is a large world map. The scene is brightly lit, and the people appear to be engaged in a discussion or collaborative work.

**Cizek and Bunch say that standard setting decisions are made using both quantitative evidence – statistical data usually - and qualitative evidence – such as judgements about item difficulty or the quality of examinees work.**

**1 Choose a standard setting system with which you are familiar.**

**2 Describe that system in writing, particularly in terms of what quantitative evidence it uses and what qualitative evidence it uses.**

**3 Write the name of your system on the appropriate flip chart.**

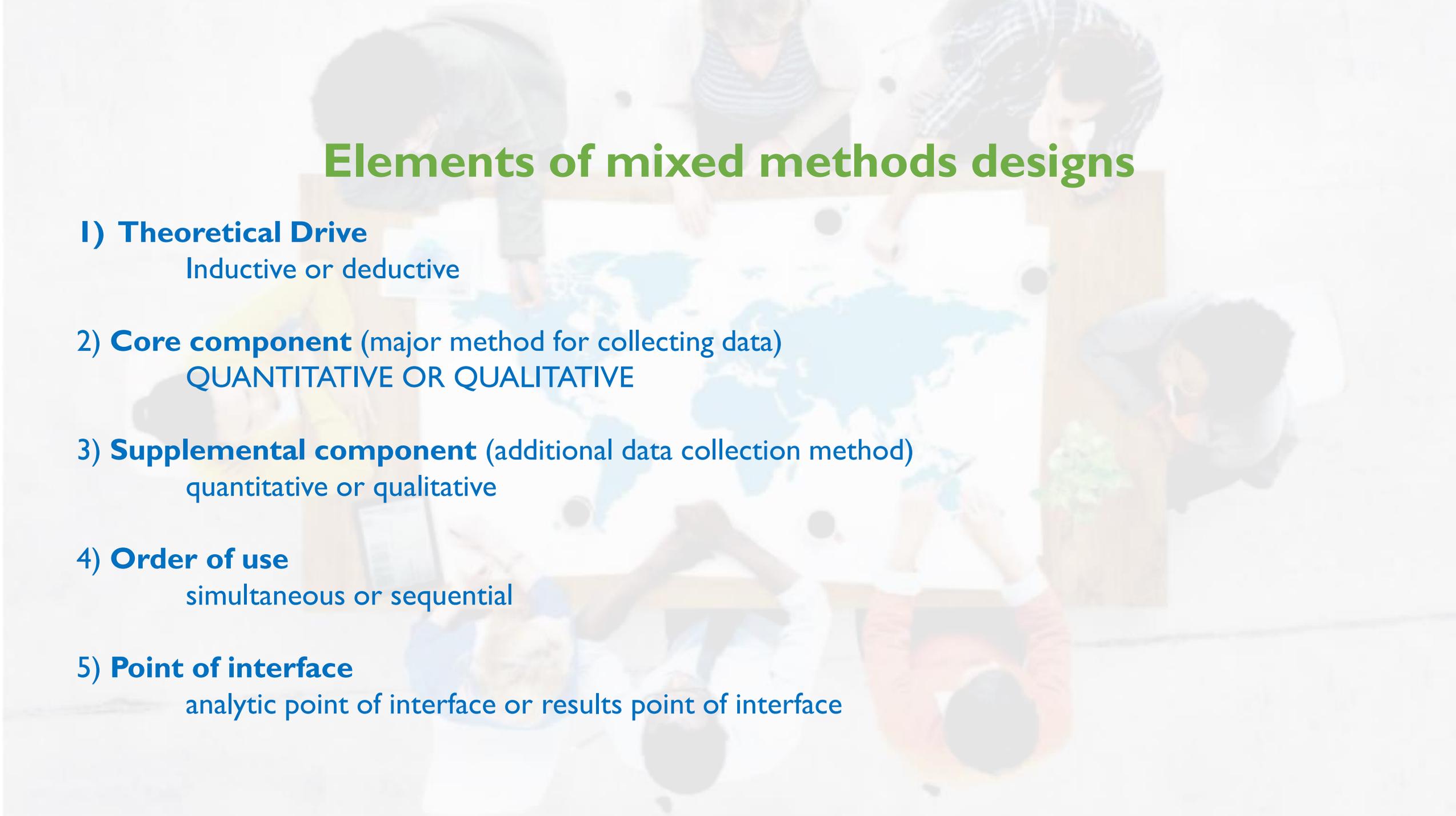
A group of people are sitting around a table, looking at a map or document. The image is a top-down view, showing the backs of the people as they lean over the table. The table is light-colored, and there are several dark-colored cups on it. The background is a plain, light-colored wall.

# Combining sources of evidence in mixed method designs

**‘exactly how evidence is combined to enable decisions to be made during standard setting has largely been ignored’**

Newton, 2000

Combining sources of evidence → mixed methods research

A group of people are sitting around a large wooden table, looking at a world map spread out on the table. They appear to be in a meeting or collaborative work session. The image is slightly faded and serves as a background for the text.

# Elements of mixed methods designs

## 1) **Theoretical Drive**

Inductive or deductive

## 2) **Core component** (major method for collecting data)

QUANTITATIVE OR QUALITATIVE

## 3) **Supplemental component** (additional data collection method)

quantitative or qualitative

## 4) **Order of use**

simultaneous or sequential

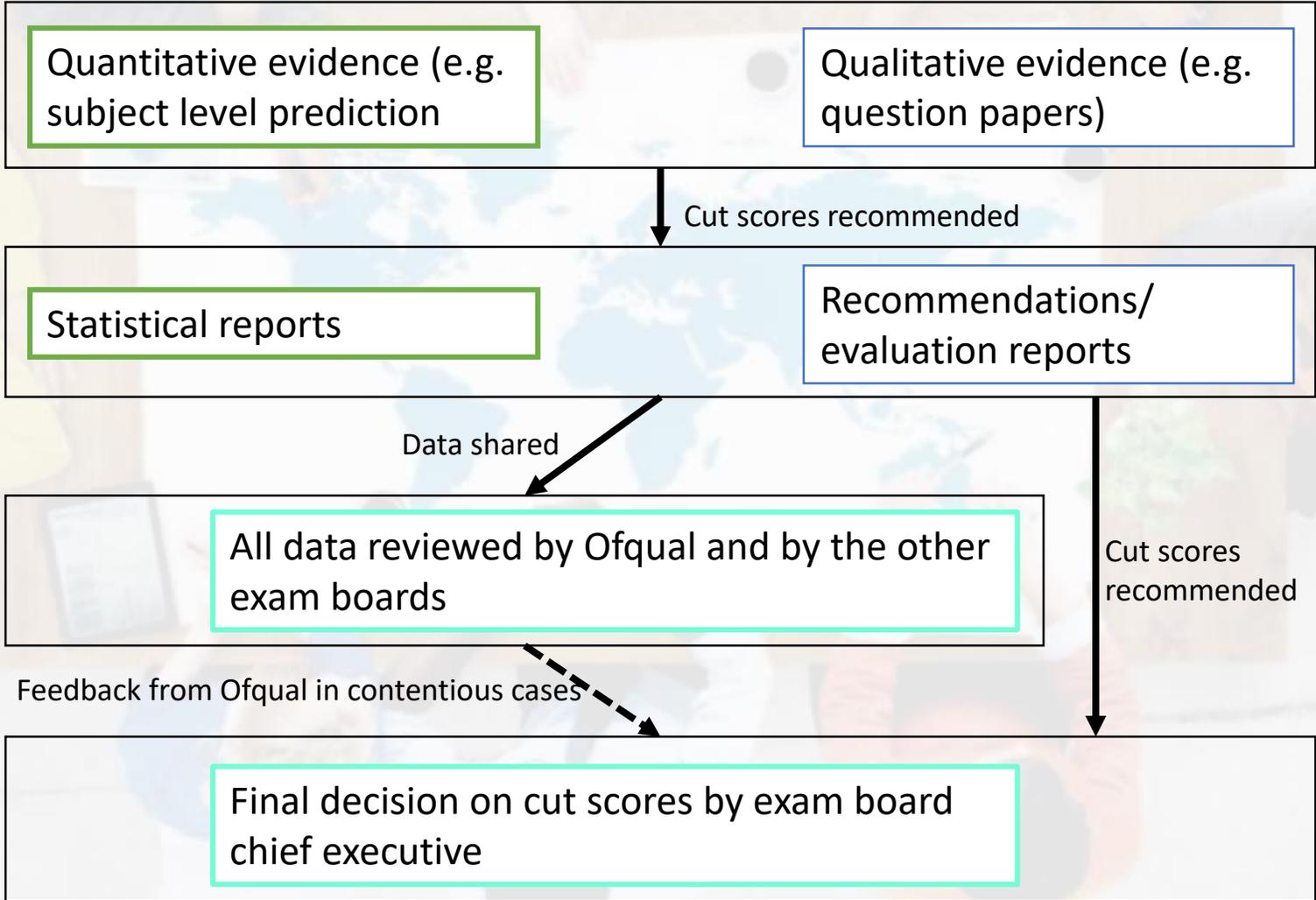
## 5) **Point of interface**

analytic point of interface or results point of interface

# England's A-level Standard setting system

CORE Component  
(QUANTITATIVE EVIDENCE)

Supplemental component  
(qualitative evidence)



Quantitative evidence (e.g. subject level prediction)

Qualitative evidence (e.g. question papers)

Cut scores recommended

Statistical reports

Recommendations/evaluation reports

Data shared

All data reviewed by Ofqual and by the other exam boards

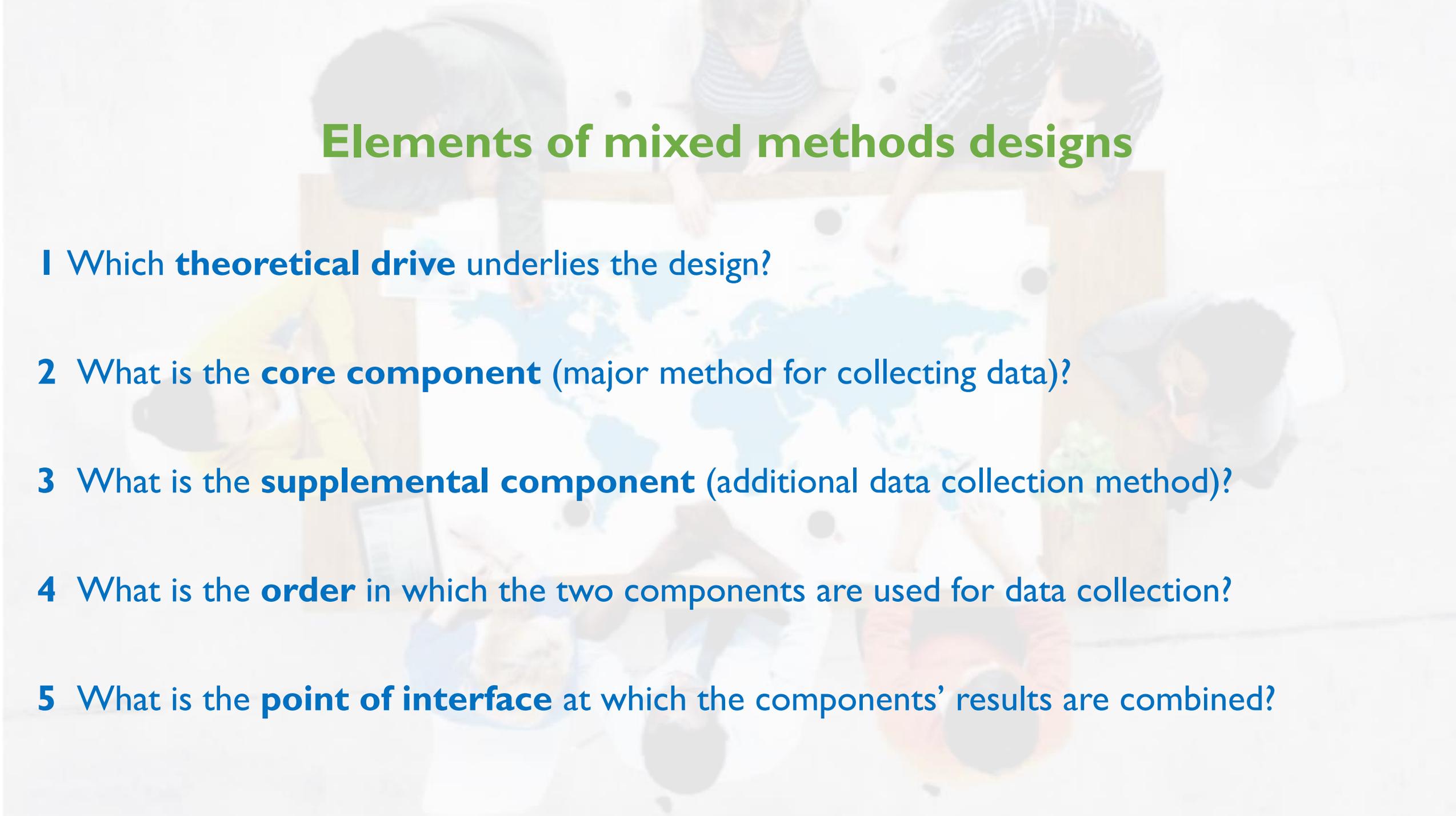
Cut scores recommended

Feedback from Ofqual in contentious cases

Final decision on cut scores by exam board chief executive

# Standard setting designs in 12 jurisdictions

<b>Quantitative</b> core component with a qualitative supplemental component (QUANqual mixed methods design)	England Hong Kong Ireland South Africa
Only <b>quantitative</b> evidence	Chile Georgia South Korea Victoria
<b>Qualitative</b> core component with a quantitative supplemental component (QUALquan mixed methods design)	Queensland Sweden US AP
Only <b>qualitative</b> evidence	France



## Elements of mixed methods designs

- 1 Which **theoretical drive** underlies the design?
- 2 What is the **core component** (major method for collecting data)?
- 3 What is the **supplemental component** (additional data collection method)?
- 4 What is the **order** in which the two components are used for data collection?
- 5 What is the **point of interface** at which the components' results are combined?

## References

Cizek, G.J. and Bunch, M.B. (2007) *Standard Setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: SAGE Publications.

Hesse-Biber, S., & Johnson, B. (2016). *The Oxford handbook of multimethod and mixed methods research inquiry* (Oxford library of psychology). New York, NY.

Johnson, R. B., and A. J. Onwuegbuzie. 2004. Mixed methods research: A research paradigm whose time has come. *Educational Researcher* 33:13–36.

Leech, N. (2013). *Mixed Methods Research*. Education. Oxford University Press.

Newton, P. E. (2000) *Maintaining Standards over Time in National Curriculum English and Science Tests at Key Stage Two: A report for the Qualifications and Curriculum Authority*. Slough: NFER.

Opposs, D. and Gorgen, K. (2018). What is standard setting? In Baird, J., Isaacs, T., Opposs, D. and Gray, L. (eds) *Examination standards: how measures & meanings differ around the world*. London: UCL IOE Press, 54-76.

Tashakkori, A., & Teddlie, C. (2016). *SAGE handbook of mixed methods in social & behavioral research* (Second ed., SAGE research methods). Thousand Oaks, California.

Watkins, D., & Gioia, D. (2015). *Mixed methods research* (Oxford scholarship online). New York.