ATP 2017
Innovations in Testing
Scottsdale, AZ

LEADING the CONVERSATION

Strategic Investments in Quality

A Guide for Developing Performance-Based Assessments

#atpconf

About the Presenters ...

ABOUT the PRESENTER

- **Michelle Laux, CFA**
  CFA Institute
  Director, Exam Development

- **E.W. Looney**
  BrightLink, LLC
  CEO

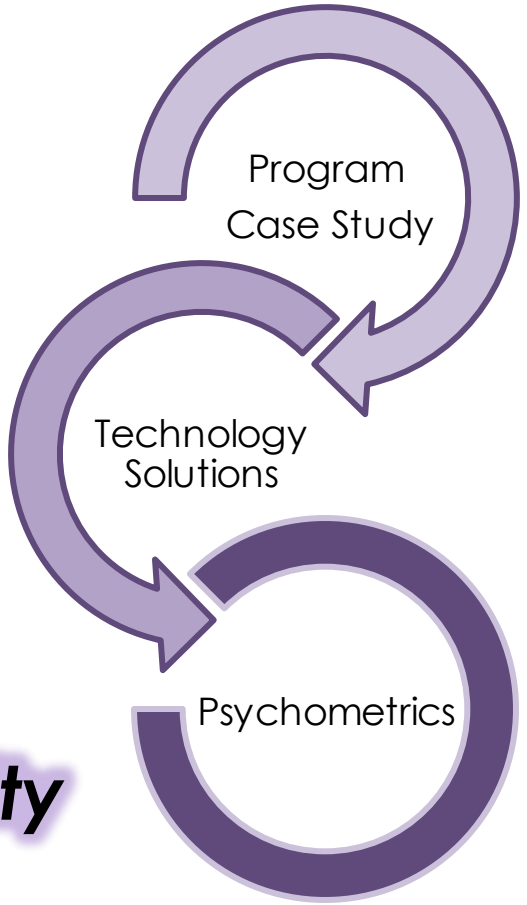
- **Chad Buckendahl**
  ACS Ventures, LLC
  Partner

The Conversation Today: Performance-Based Assessments

LEADING the CONVERSATION

Assessing Candidate Responses

- Design
- Implementation
- Data Analysis



Strategic Investments in Quality

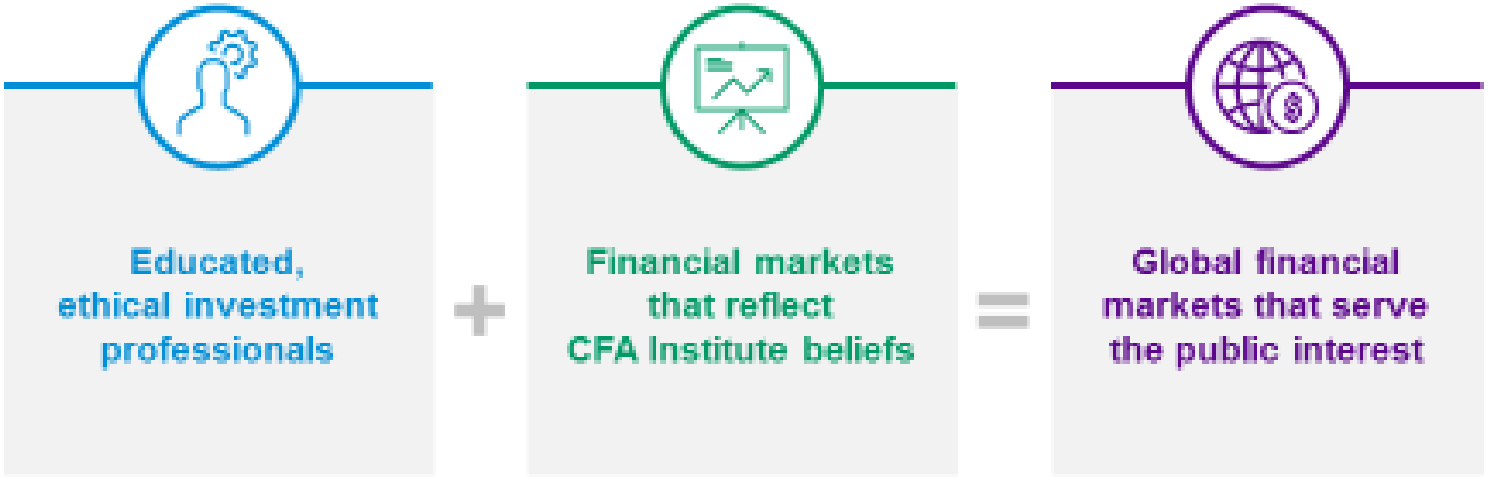
Program Case Study

Program Description

LEADING the CONVERSATION

CFA Institute

Leads the investment profession globally by promoting the highest standards of ethics, education, and professional excellence for the ultimate benefit of society



CFA Level III – Constructed Response Exam

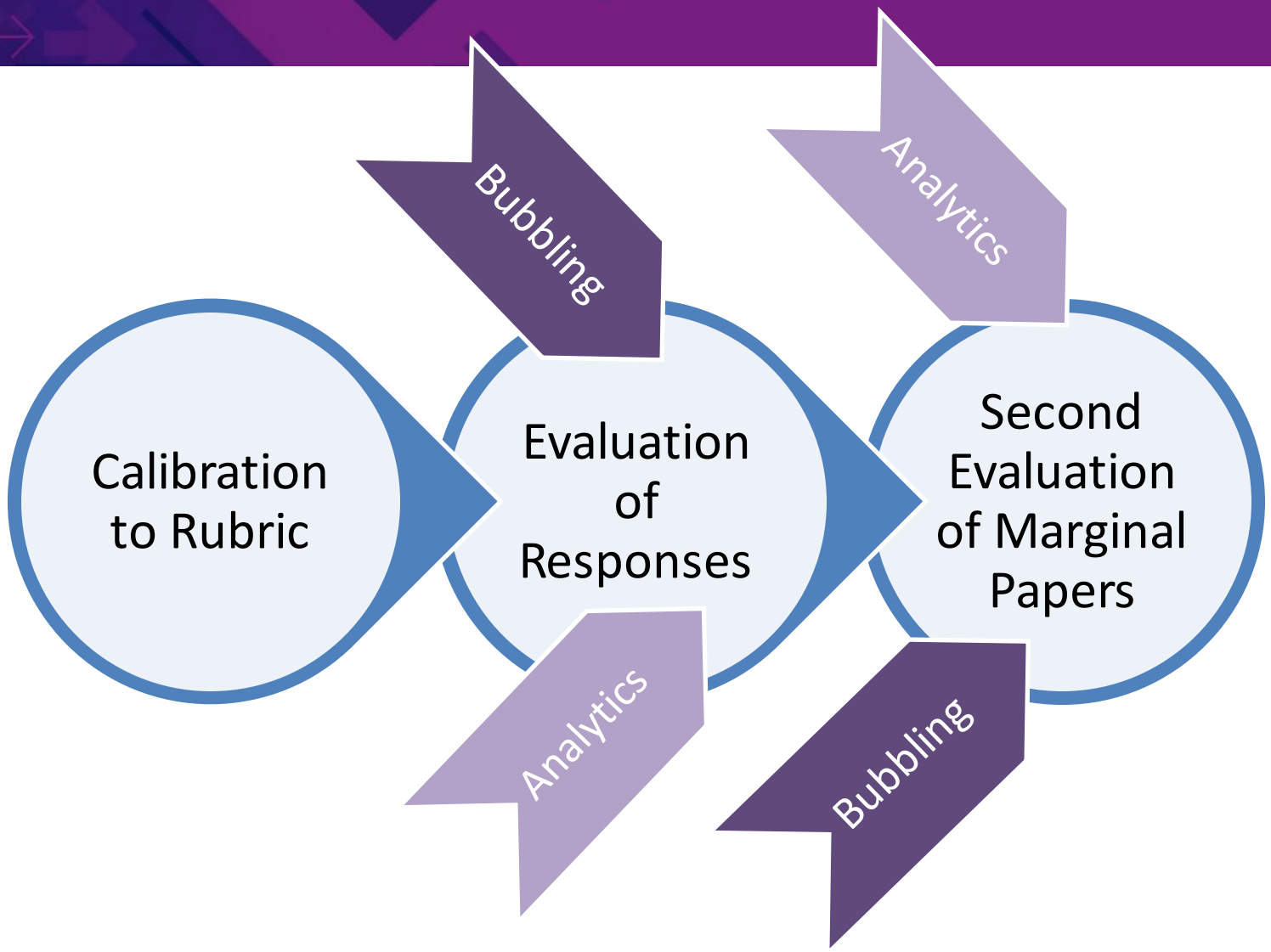
- 28,884 people took the Level III exam in 2016
- Pen & paper assessment

Description of Grading, circa 2014

LEADING the CONVERSATION

Description of Grading, circa 2014

LEADING the CONVERSATION



Sample scored candidate answer

Template for Question 1-E
Note: Assume no transaction costs or liquidity needs.

Asset class	**i. Determine the "sell" amount of bonds and the "sell" amount of equities to achieve the *most* tax-efficient allocation in *each* account (tax-deferred and taxable).**	
	Tax-deferred Account	**Taxable Account**
Bonds	*0*	*USD 500,000*
Equities	*USD 500,000*	*0*

Asset class	**ii. Determine the "buy" amount of bonds and the "buy" amount of equities to achieve the *most* tax-efficient allocation in *each* account (tax-deferred and taxable).**	
	Tax-deferred Account	**Taxable Account**
Bonds	*USD 500,000*	*0*
Equities	*0*	*USD 500,000*

iii. Justify, with *two* reasons, why this is the *most* tax-efficient allocation.

1. *Morbi faucibus orci fermentum ipsum euismod pellentesque. Sed lacus odio, laoreet a interdum porta, aliquam vitae arcu. Sed eu justo lobortis sapien varius mollis. Aenean pellentesque posuere fringilla.*

2. *Proin ut lacus ac ipsum auctor ultrices. Donec hendrerit ligula lacus. Integer vitae velit a erat imperdiet adipiscing. Cras ut erat gravida lorem feugiat tincidunt.*

1
1
1
1
2
0
E=6

Process Improvements

LEADING the CONVERSATION

We were ready for the next process improvement

Technology

Examiner Calibration (on paper)

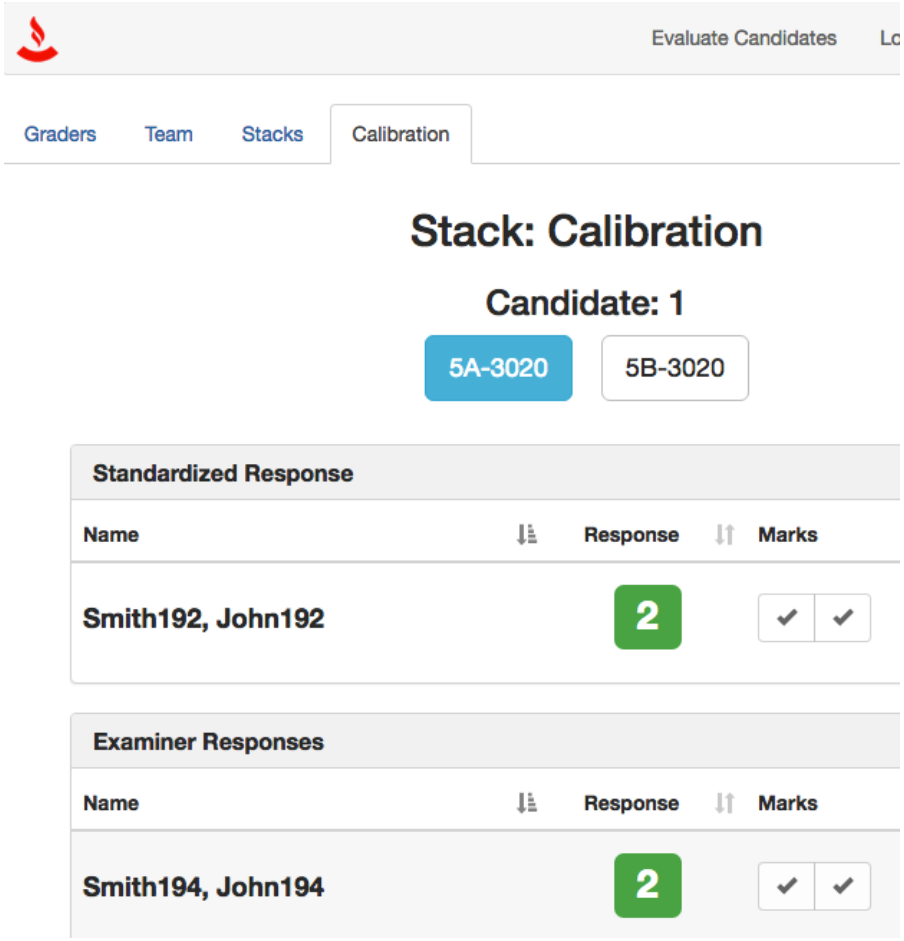
- Captains learn about results through volunteer information (or never)
- Bandwagon effect
- Scaling problems around oversight and quality

Examiner Calibration (on tablets)

LEADING the CONVERSATION

Scalable, Quality Interactions

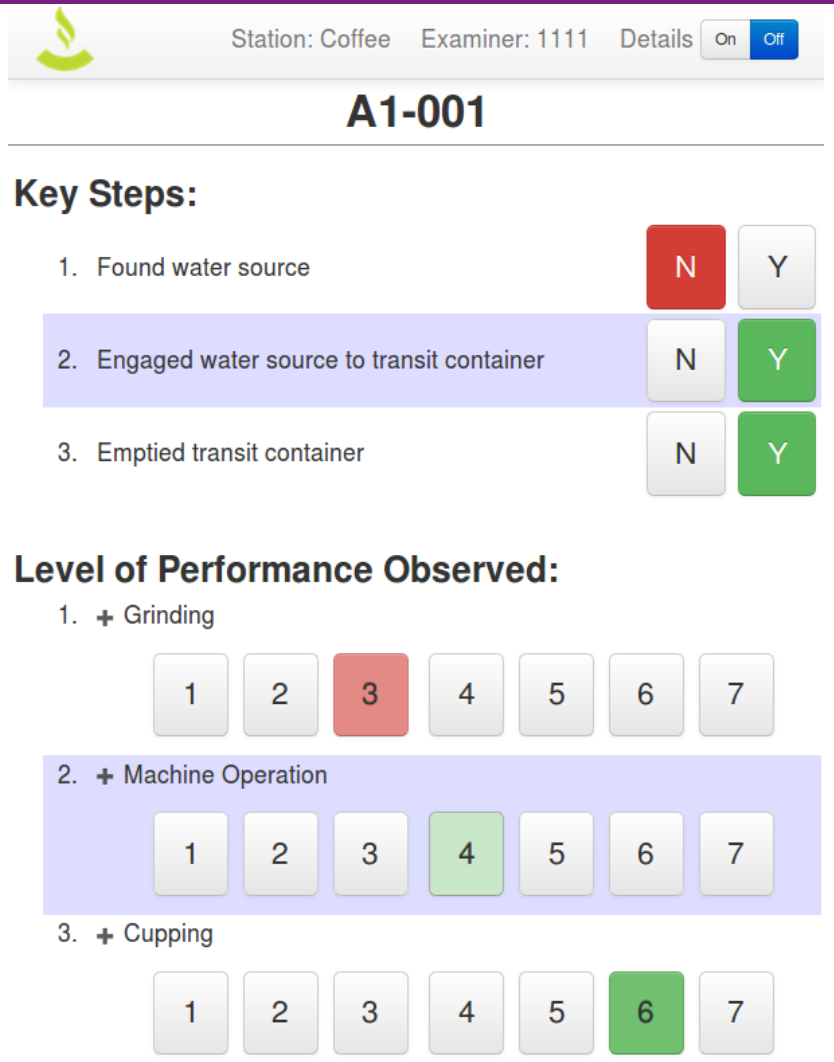
- Captains see *real* data immediately
- Independent grading



Reduce Cognitive Load

LEADING the CONVERSATION

- No more red pens
- No more double checking
- No more 2+1=4



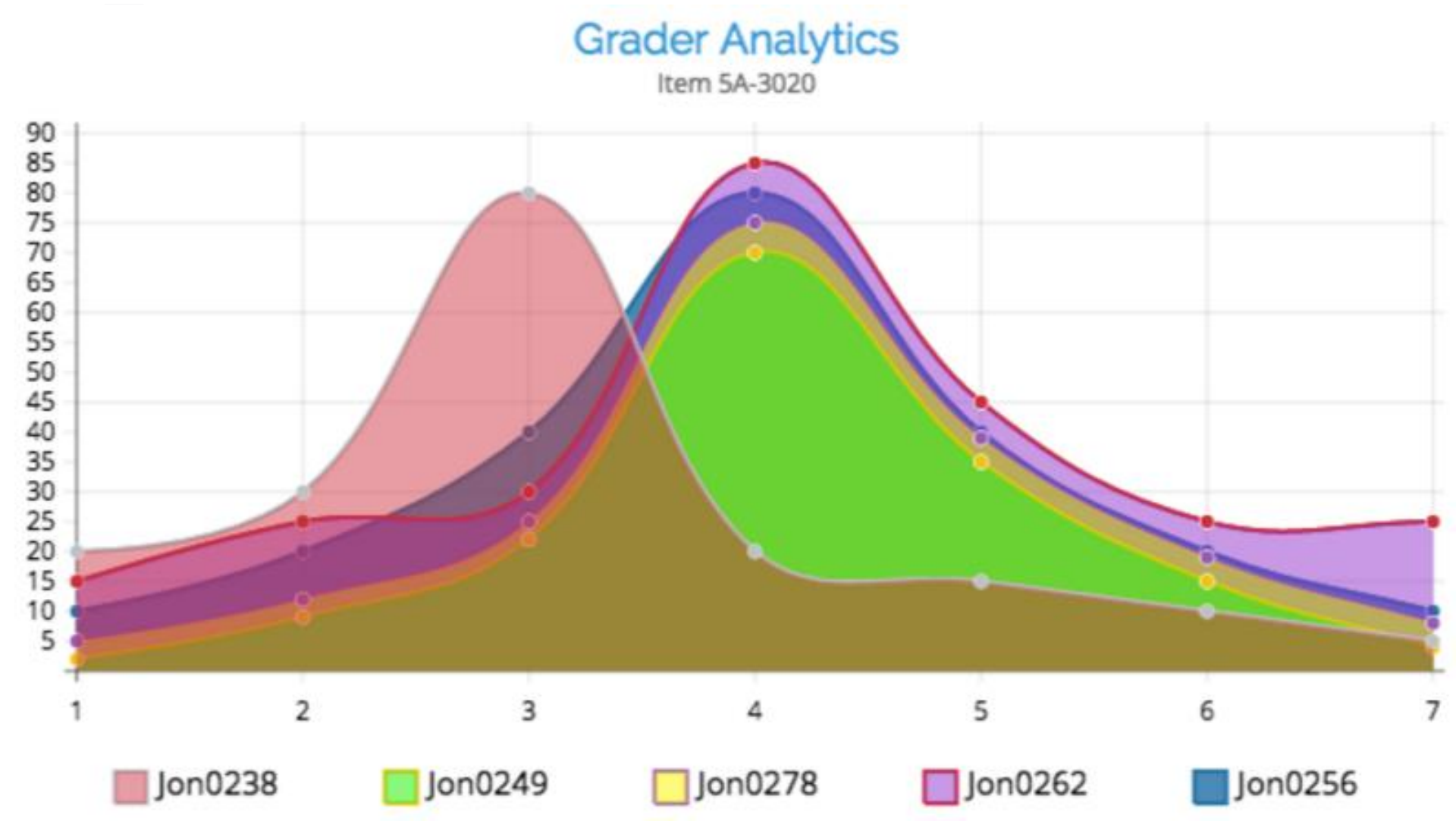
Real-time data

Grader Analytics

Name	Books	Mean	Std Dev	0s	Full	Calibration Rating
Smith, Jon0113	20	2.51	2.23	7	7	86%
Smith, Jon0179	10	2.60	2.37	4	4	95%
Smith, Jon0196	20	1.70	1.95	10	3	23%
Smith, Jon0225	10	2.65	2.20	0	9	92%

ATP Innovations in Testing 2017 Scottsdale, AZ

Real-time data



Usability: Paper vs. Tablets

Which is more usable, paper or tablets?

Ask Bob [redacted].

Usability: Paper vs. Tablets

LEADING the CONVERSATION

Based on Analytic data, Bob is one of CFA's most accurate graders

Usability: Paper vs. Tablets

LEADING the CONVERSATION

Bob is 89 years young

Psychometrics

Data collection

LEADING the CONVERSATION

Design

- Articulating scoring criteria
- Identifying important sources of error

Implementation

- Training and calibration
- Operational scoring
- Audit papers
- Captain's review

Using results of data analysis

- Formative uses
  - Accuracy of scoring
  - Consistency of scores
  - Consistency of decisions
  - Scoring criteria
  - Calibration
- Summative uses
  - Evaluating evidence for validity of candidates' scores
  - Quality control decision rules
    - Rescoring policies
    - Grader evaluation

Examples of feedback

Formative

Feedback from audit papers

- % of agreement with known scores/decisions
- Exemplars of higher/lower scores with mentoring

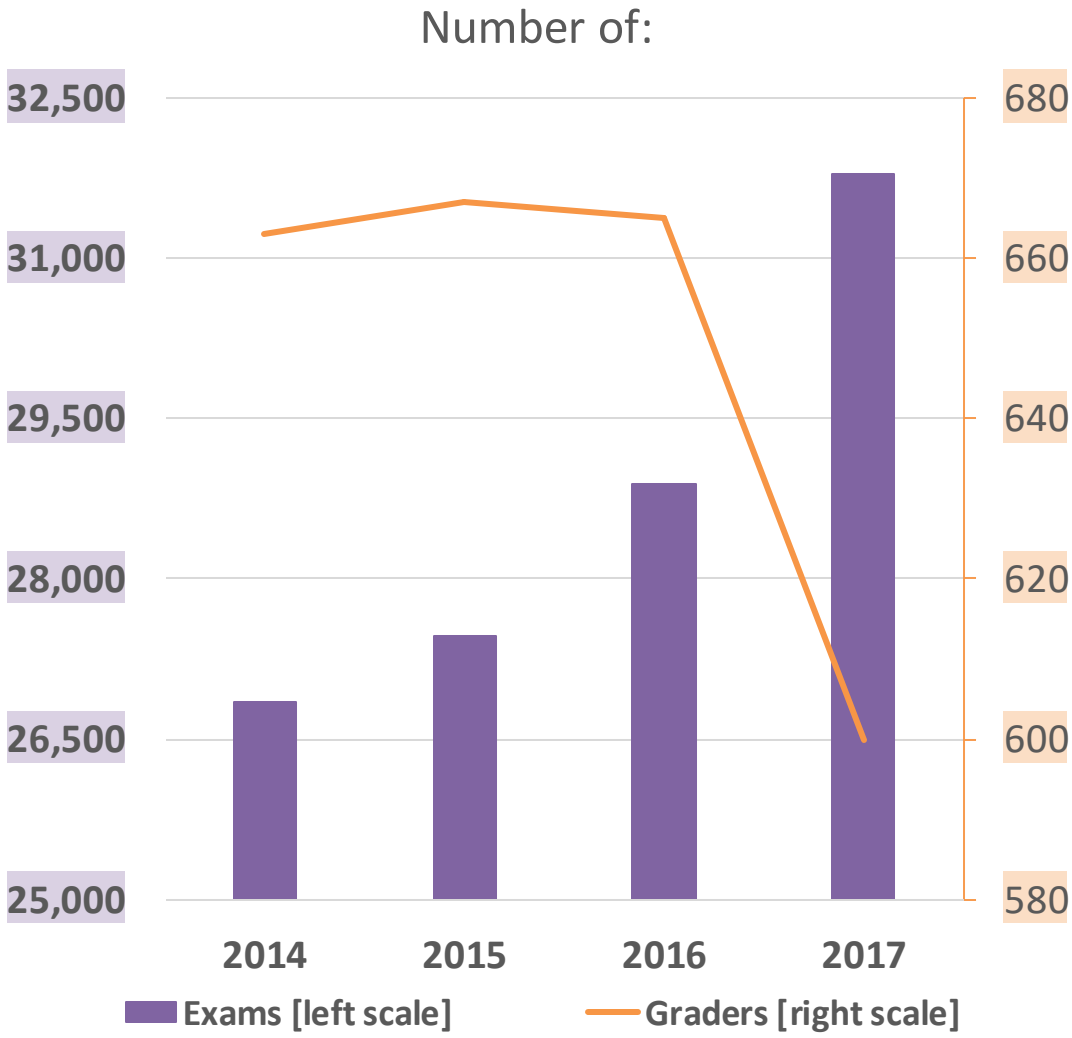
Summative

Grader evaluation

- 95% classification consistency on audit papers (dichotomous decisions)
- 80% exact agreement on known scores (e.g. 4-point scale)
- Rescore, retrain, rehire

Results

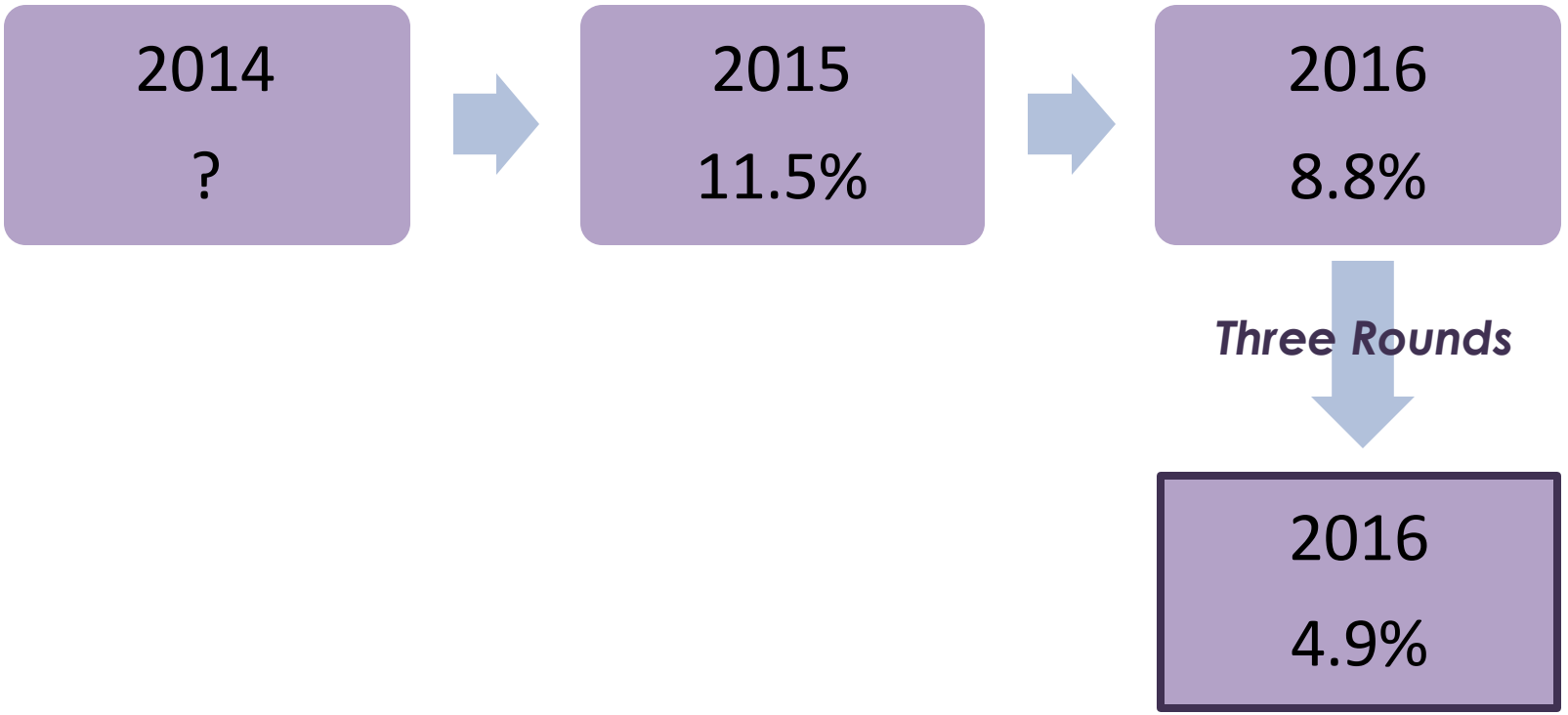
Results: Optimized Resources



Results: Improved Accuracy

LEADING the CONVERSATION

- Disagreement in Scoring, through two rounds:



What the Future Holds

LEADING the CONVERSATION

CFA Institute

2017	Future
• Rehiring decisions • Additional calibration – before live grading	• Disagreement • Additional calibration – during live grading • Digital capture of responses • Automated scoring engines

LEADING the CONVERSATION