

How do Latent Print Examiners Perceive Proficiency Testing? An Analysis of Examiner Perceptions, Performance, and Print Quality



AAFS Annual
Meeting
February 22, 2019



Sharon Kelley, Daniel C. Murrie, Brett O. Gardner, Karen D.H. Pan,
Karen Kafadar, & Kellyn Blaisdell

This work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement #70NANB15H176 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, and University of Virginia.



Purposes of Proficiency Testing

Identify strengths and weaknesses in procedures

Identify strengths and weaknesses in personnel

Provide performance indicators

Performance Indicators for:

Internal monitoring

Accreditation
(e.g., ISO/IEC 17025)

Courts
(i.e., evidence of
expertise)
Other legal
stakeholders

But...

- Commercial proficiency test materials may differ in important ways from routine, “real world” casework (e.g., Koehler, 2008; National Research Council [US] Whither Biometrics Committee, 2010)
- Proficiency tests are sometimes criticized as much simpler than real casework (e.g., Bayles, 2002; Koehler, 2017; Mnookin, 2008)
- Historically, no metrics existed for calibrating difficulty of a test

Objectives & Methods

1) Explore latent print examiners' opinions of proficiency testing

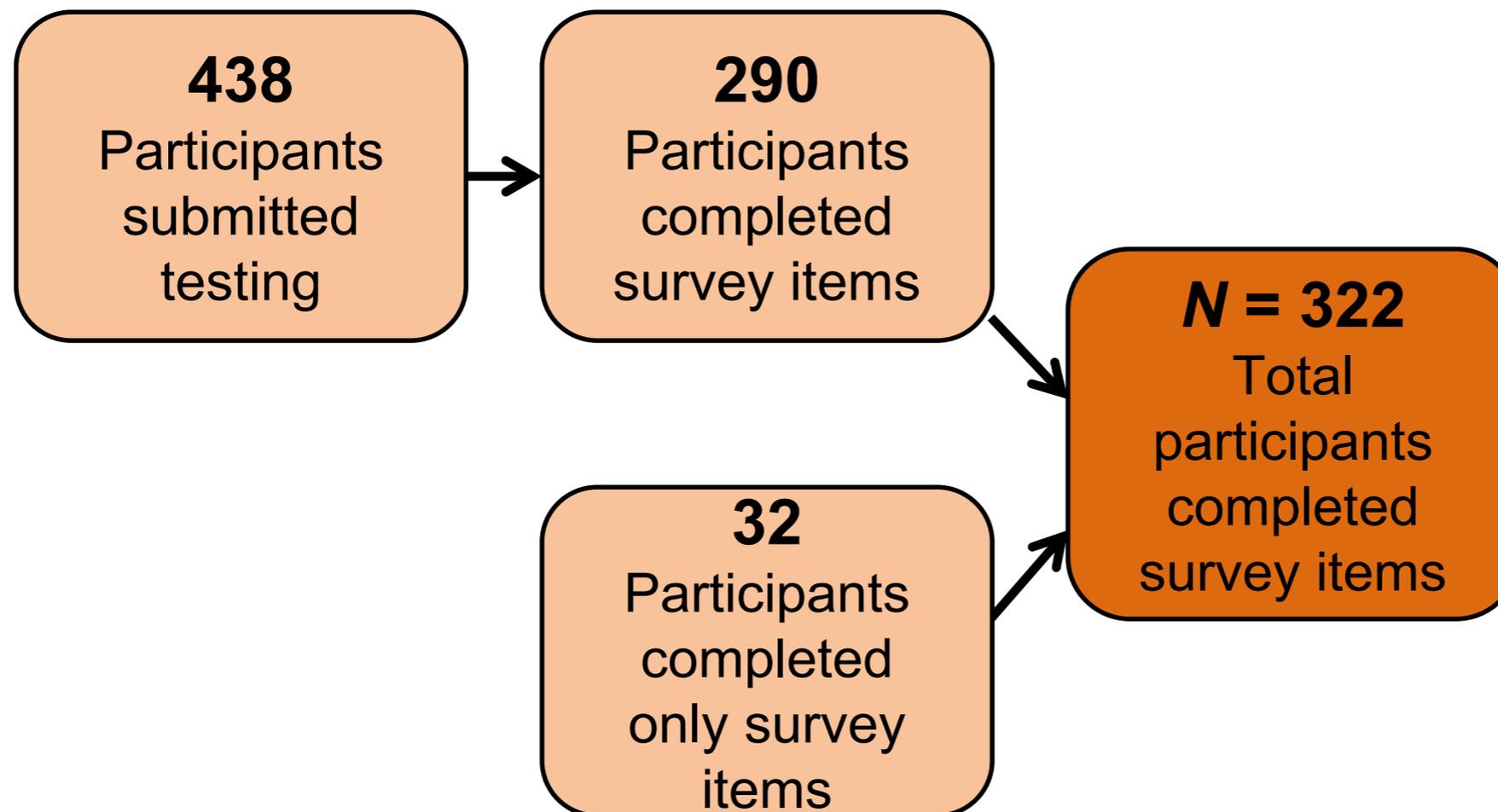
2) Explore how examiner opinions relate to performance on proficiency tests

3) Examine both subjective and objective indicators of print quality in current proficiency tests.

- Partnered with Collaborative Testing Services (CTS), leading provider of proficiency tests
- Distributed a survey with the 2nd latent print proficiency testing of 2017

Participants

- Participants were 321 fingerprint analysts who completed supplementary survey questions on the CTS latent fingerprint proficiency test due in October 2017.



Materials

Questions for each print (Q1-Q11)

On a scale of 0-10, how would you rate the challenge level of this questioned latent print?

On a scale of 0-10, how closely does this questioned latent print compare to latent prints received in your casework?

Questions for the overall test

Which latent print did you find to be the least challenging?

Which latent print did you find to be the most challenging?

What characteristic(s) of this latent print led to it being the most challenging?

- Image quality
- Distortion
- Limited points to compare
- Over/underdeveloped ridge detail
- Other (please explain)

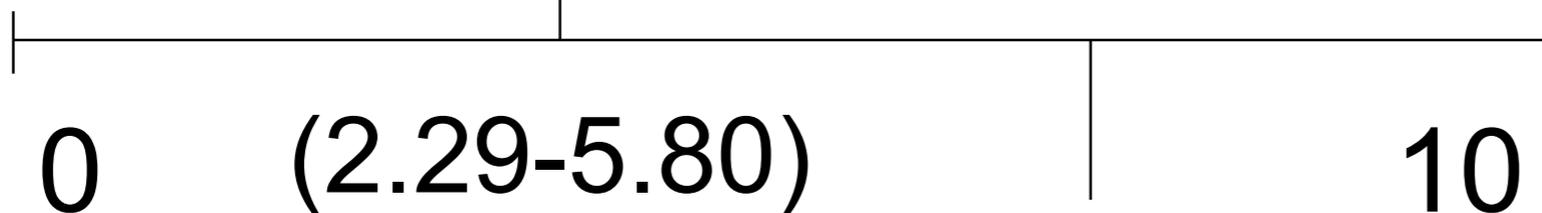
0		10	
Extremely easy	No confidence	Nothing like casework	Extremely challenging
			Extremely confident
			Exactly like casework

Results: Perceptions of Test Items

Average
perceived
difficulty

4.27

(6.39-7.44)



6.97

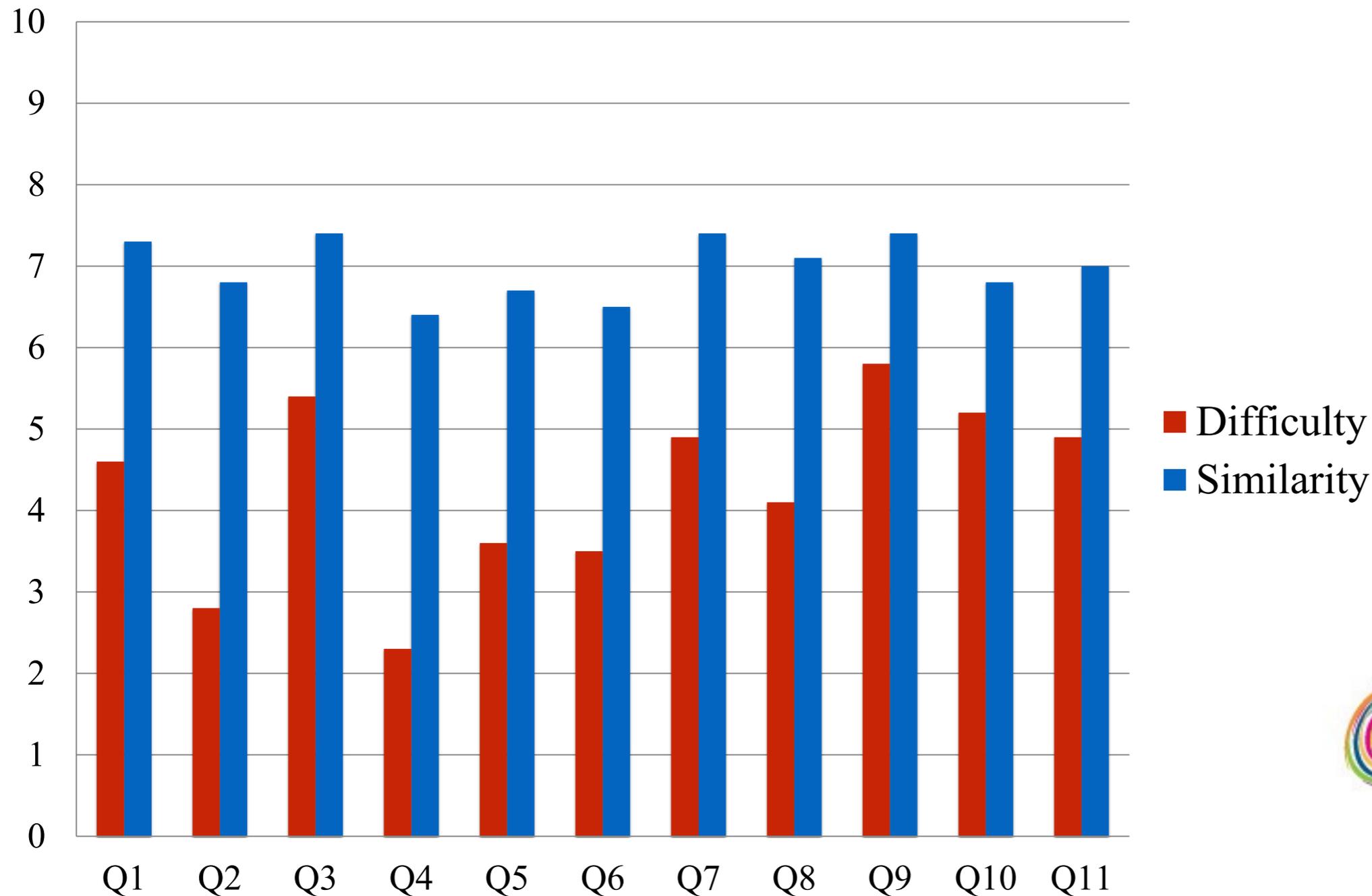
Average perceived
**similarity to
casework**

Test items perceived as more similar to casework were also perceived as more difficult

($r = .32, p < .001$)

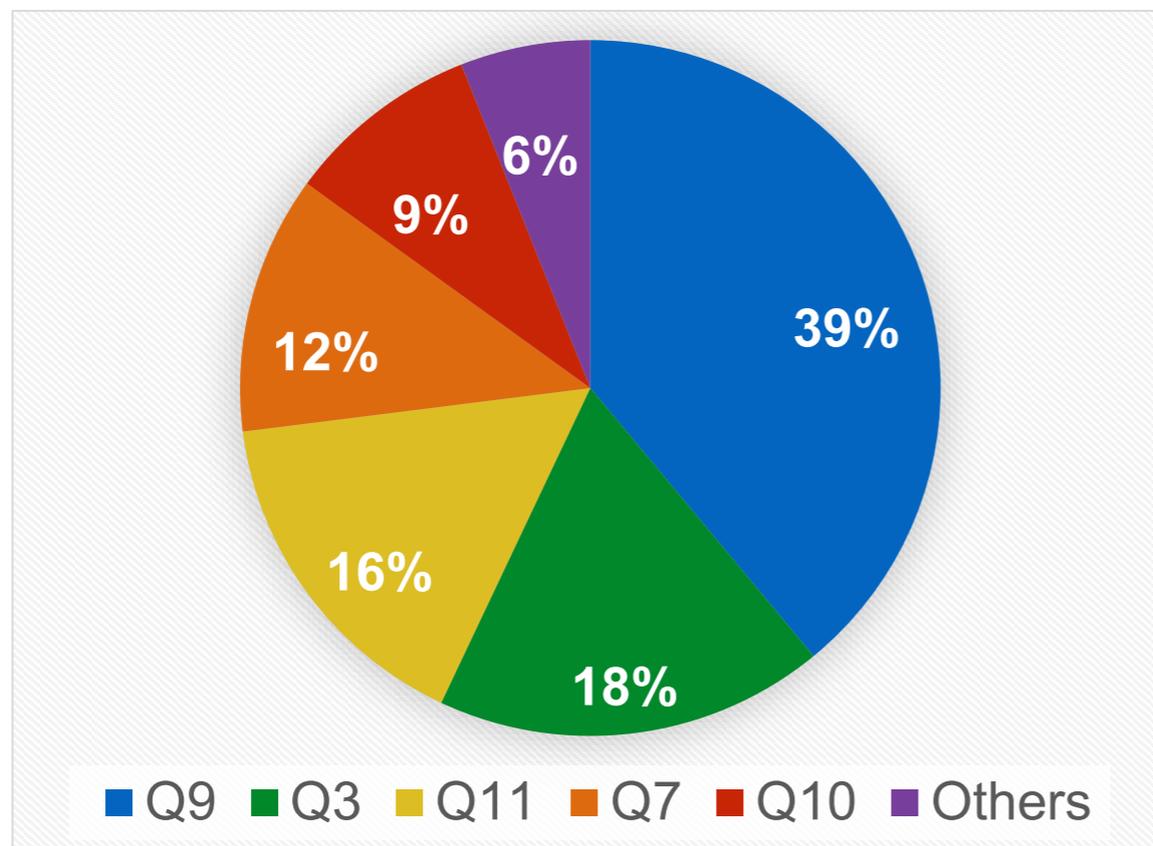
Results: Perceptions of Test Items

Perceived Difficulty of Test Items and Their Similarity to Casework

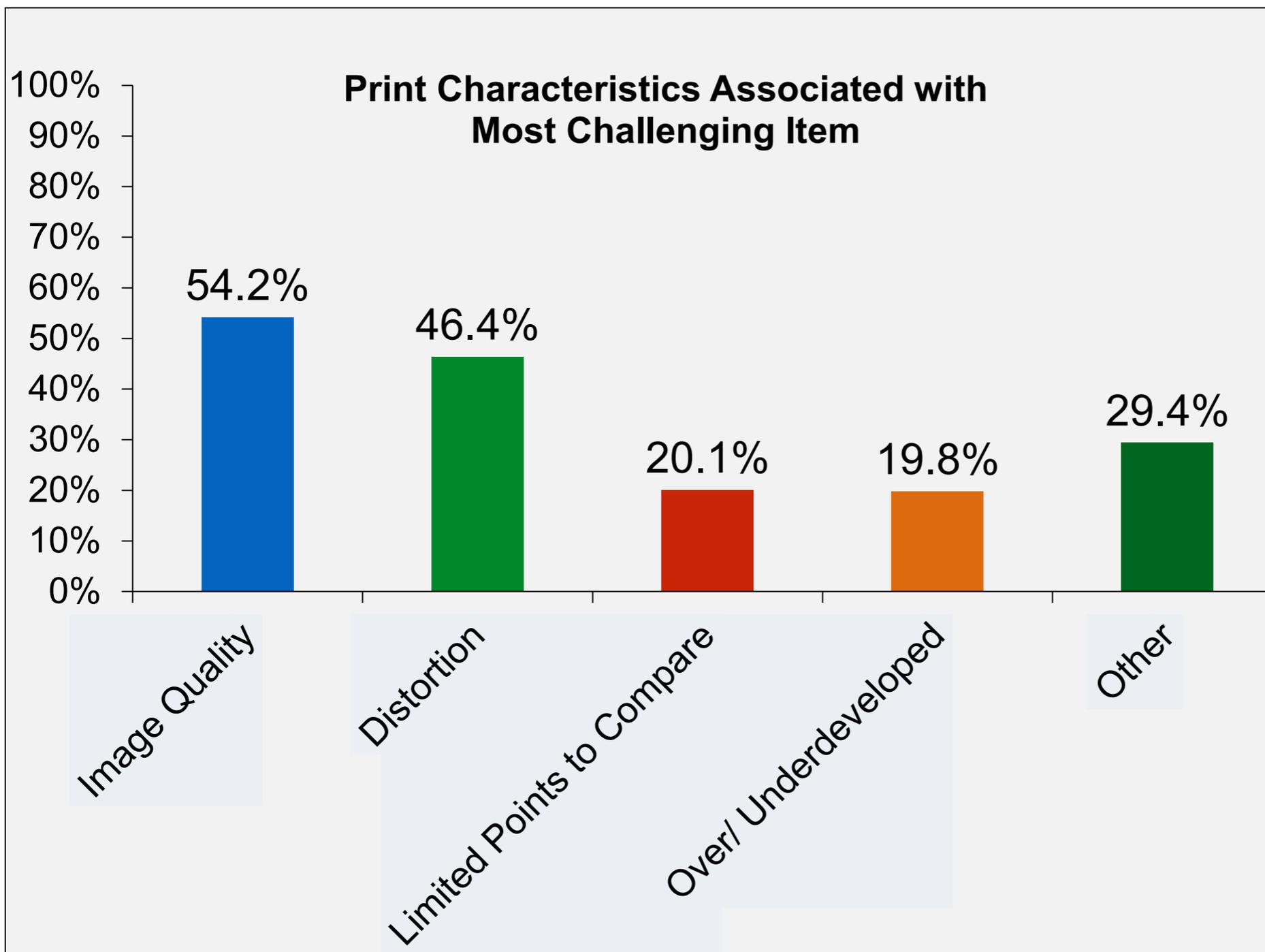


Results: Least and Most Challenging

- 80% of participants identified one of two items as **least** challenging.
- No more than 39% of participants identified any one item as **most** challenging.



Results: Reasons for Challenge



- “Other”**
- Known print quality concerns
 - Palm prints
 - Creases
 - Pressure

Quality Metrics

Objective LQMetrics

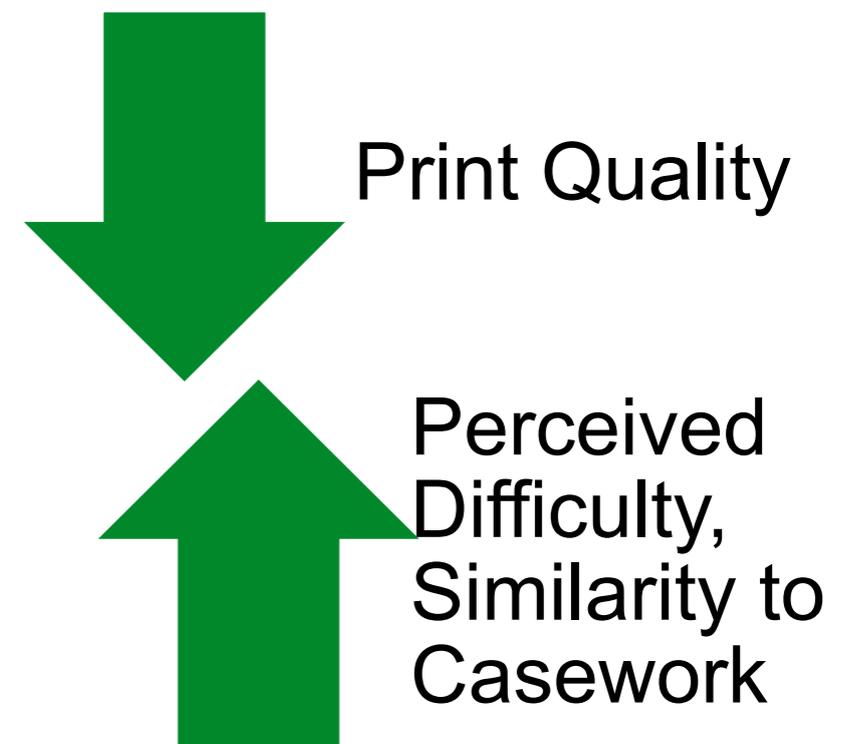
Prints	Overall Quality	VID	VCMP
Latent prints ($n = 11$)	78.64 (60 – 99)	98.55 (96 – 100)	99.55 (99 – 100)
Known fingerprints ($n = 124$)	92.41 (55 – 99)	99.64 (95 – 100)	99.97 (99 – 100)
<ul style="list-style-type: none"> Full-hand prints ($n = 44$) 	88.30 (55 – 99)	99.30 (95 – 100)	99.91 (99 – 100)
<ul style="list-style-type: none"> 10-print card ($n = 40$) 	98.80 (96 – 99)	100	100
<ul style="list-style-type: none"> Simult. 10-print ($n = 40$) 	90.55 (73 – 99)	99.68 (98 – 100)	100
Source fingerprints ($n = 22$)	93.41 (71 – 99)	99.77 (98 – 100)	99.95 (99 – 100)

Quality Metrics and Examiner Perceptions

- LQMetric scores for latent prints were **not** associated with examiner perceptions of difficulty ($r = -.21, p = .53$) or similarity to casework ($r = -.30, p = .37$)
- BUT:
- The item that was perceived as most difficult and similar to casework contained the lowest quality source prints ($M = 81.00$), etc.



General Trend:



Quality Metrics and Examiner Perceptions

- We also computed quality metric scores for each **test item** (average of the quality metric scores for each latent print with the quality metric scores of its source prints)
- This quality score was significantly correlated with:
 - Perceptions of item difficulty
 - 56% of variance explained ($r = -.75, p = .05$)
 - Perceptions of similarity to casework
 - 67% of variance explained ($r = -.82, p = .02$)
- Test items perceived as more challenging and more similar to casework contained prints of lower quality.

Results: Confidence & Performance

Least challenging item

- 96% respondents: 10/10
- Average: 9.95/10

Most challenging item

- Average: 9.4/10

One item accounted for 54% of errors (most frequently identified as most challenging)

11 out of 290 respondents made an error* on any test item.

Results: Performance

11 examiners who provided erroneous responses did not generally perceive items to be more or less difficult than those who did not provide any inconsistent responses.

Only 2 of the 11 respondents with erroneous responses identified the item for which they provided an inaccurate response as being the most challenging

Item contributing to most errors did not have lowest quality latent print

Item contributing to most errors did have much lower quality source prints (LQMetric = 71, 76, 79, & 98)

Discussion

- Participants found the proficiency test items to be relatively easy (4.27/10) and moderately similar to casework (6.97/10)
- Participants most often cited **image quality** and **distortion** as factors that made the most challenging items particularly challenging.
- Prints of high quality overall
- Participants were highly confident in their decisions
- Overall accuracy was high
- The extremely high rate of accuracy precluded evaluation of any relationships between survey responses and accuracy.

Limitations & Future Directions

- Survey associated with only 1 CTS proficiency test
- Unclear how representative sample was
- Results suggest need for broader range of prints to better reflect casework
 - Incorporate quality metrics when designing tests?
- Broader proficiency testing issues:
 - Blinding





Sharon.Kelley@virginia.edu