
2019
ICE
EXCHANGE
San Diego
Monday, November 18 -
Thursday, November 21, 2019
Hilton San Diego Bayfront
#ICEEXCHANGE19

2019 ICE EXCHANGE

Using Foil Tasks to Identify and Screen Careless Job Analysis Survey Respondents

Jeff Kelley, Psychometrician, Alpine Testing Solutions

Terrence Wright, Certification Director, AHIMA

AHIMA

Alpine Testing Solutions

#ICEEXCHANGE19

Job Analysis Survey Data Screening

- Should we screen for careless responses?
  - Other survey contexts
  - Self-screening in credentialing
- If so, how?
  - Intrusive methods (foil tasks)
  - Non-intrusive methods
- Does screening have an impact on results?

Intrusive Methods - Foil Tasks

- Bogus tasks
  - Clearly outside scope of practice, but not absurd
  - If endorsed, possible indication of carelessness
  - Written by the SMEs on the job analysis panels
  - Expected response is “seldom or never” (frequency) and “no risk of a negative result” (criticality)
  - “Negotiate contracts with insurance payers.”

- Screening tasks
  - “Demonstrate careful reading of this task by selecting ‘always’ for the frequency rating and ‘no risk of a negative result’ for the criticality rating.”

Non-Intrusive Methods

- Response time
  - Start to end survey time recorded by the survey delivery engine
  - Flag the fastest 5% of respondents
- Response variability
  - Flag responses with no variability across frequency or criticality scales (excluding foil tasks)
- Person misfit
  - Flag Rasch outfit mean square > 1.4 for frequency or criticality scales (excluding foil tasks)

AHIMA Job Analysis Surveys

- Certified Coding Specialist (CCS)
  - 2,086 complete responses, 7.4% response rate, 20 job tasks
- Certified Coding Specialist-Physician based (CCS-P)
  - 462 complete responses, 9.7% response rate, 24 job tasks
- Registered Health Information Technician (RHIT)
  - 2,327 complete responses, 7.5% response rate, 21 job tasks
- Incentive –$250 gift card drawing entry for fully completing the survey

Face Validity

- Survey Disclaimer
  - “You may have noticed some tasks in this survey that did not appear appropriate to the role of a coding specialist. These were included for research purposes and are not intended for inclusion on the new CCS/CCS-P/RHIT exam.”

Face Validity

- Job Analysis Panel
  - Bogus task writing
  - Concerns expressed

- Other AHIMA stakeholders
  - Concerns expressed

Face Validity – Survey Comments

- CCS (2,086 total responses)
  - Bogus task – 22 mentions
  - Screening task – 19 mentions
- CCS-P (462 total responses)
  - Bogus task – 17 mentions
  - Screening task – 3 mentions
- RHIT (2,327 total responses)
  - Bogus task – 31 mentions
  - Screening task – 6 mentions

Face Validity – Survey Comments

- Bogus task typical comments
  - Generally emphasizing the task is not in scope
  - “It is not the role of an outpatient or CCS-P coder to negotiate contracts with insurance companies.”

- Screening task typical comments
  - Some positive (25%), mostly negative (75%)
  - “Love question 3, makes sure you are paying attention.”
  - “Don't understand 4.03 - is this a trick question??”

False Positive Careless Classification

- Bogus task
  - Misread, misunderstood, random error, different interpretation
- Screening task
  - Misread, misunderstood, random error
- Response time
  - Fast reader

False Positive Careless Classification

- Response variability
  - Genuinely believe all tasks are at the same level
- Person Misfit
  - Misfit due to a truly different perspective on the job tasks

Careless Classification Decisions

- Endorse the bogus task, really fast, and no response variability?
- Trip on the screening task, really fast, and misfit?
- No variability, misfit, and endorse bogus task?

Careless Classification Decisions

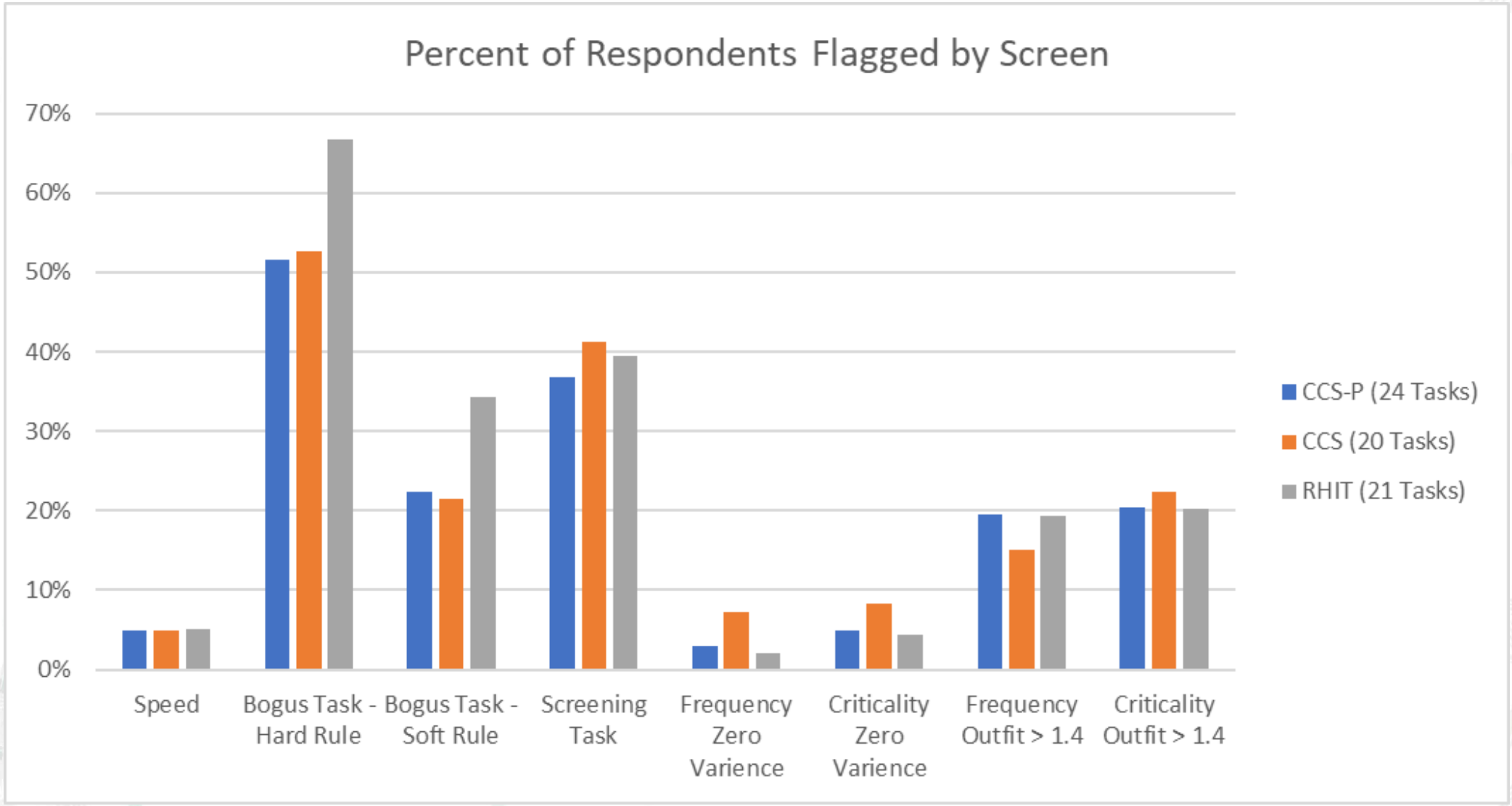
- Number of flags required?
  - Higher number – fewer false positives, more careless respondents included, larger survey sample
  - Lower number – more false positives, fewer careless respondents included, smaller survey sample
- 3 strikes!

- Experimental only. No respondents removed from actual job analysis studies.

Results

- Order effects
  - Randomly assigned both foil tasks early/late in the survey (RHIT)
  - Randomly assigned bogus/screening to early/late (CCS)
  - Negligible differences in failure rates

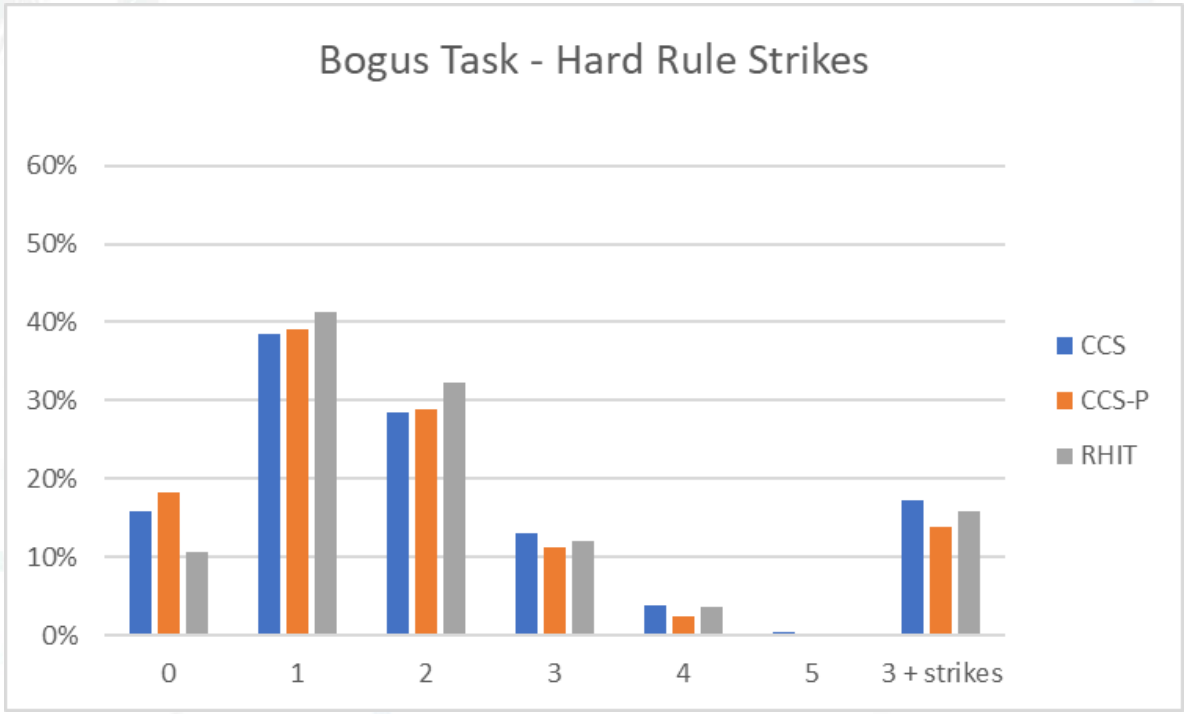
Results

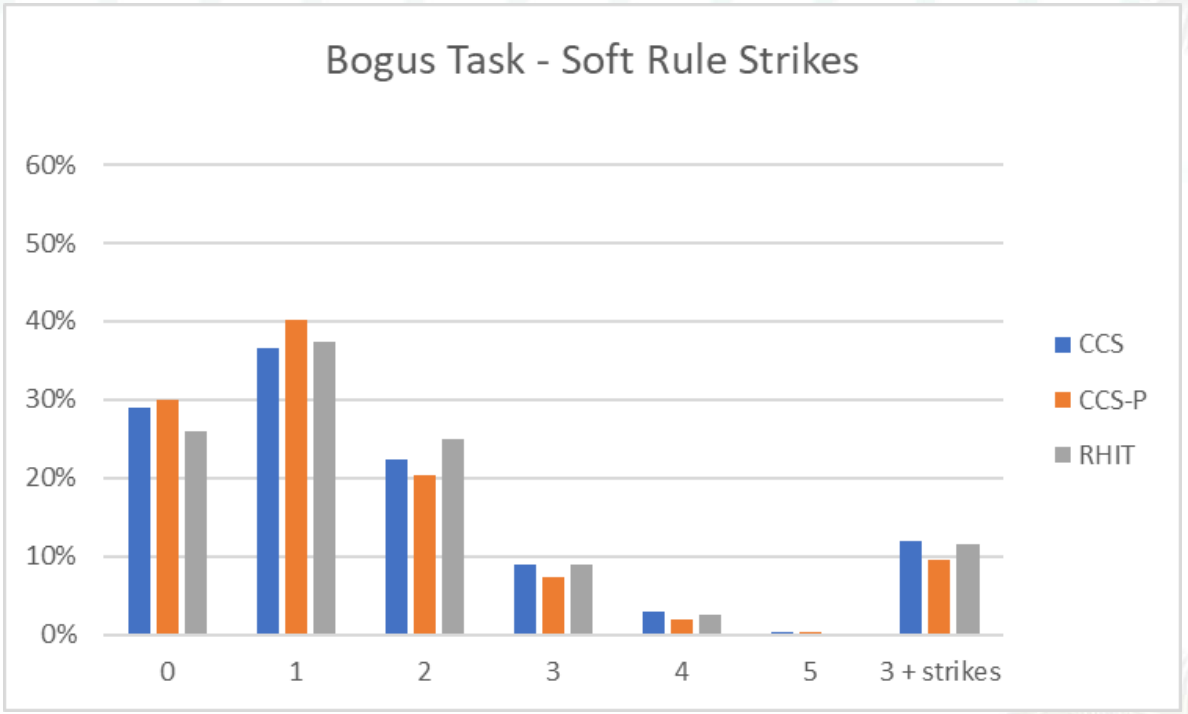

Percent of Respondents Flagged by Screen
70%
60%
50%
40%
30%
20%
10%
0%
Speed
Bogus Task - Hard Rule
Bogus Task - Soft Rule
Screening Task
Frequency Zero Varience
Criticality Zero Varience
Frequency Outfit > 1.4
Criticality Outfit > 1.4
CCS-P (24 Tasks)
CCS (20 Tasks)
RHIT (21 Tasks)

Results – Possible Strikes

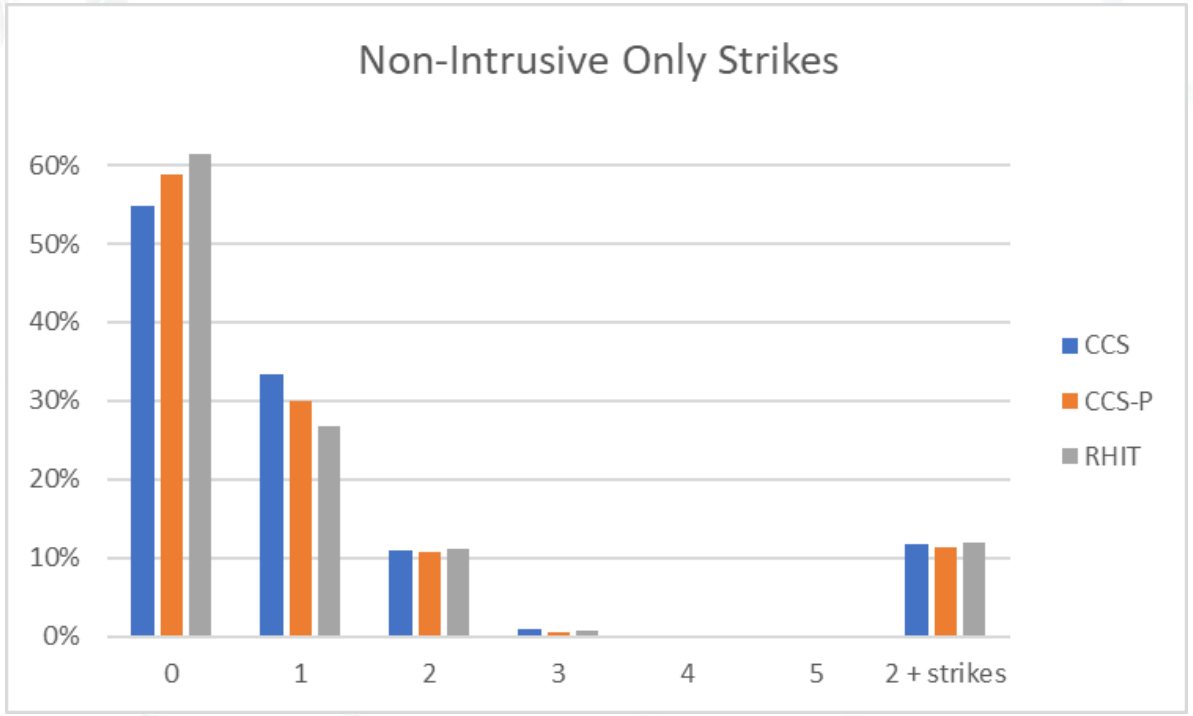
- Bogus task
  - Fail to select both “seldom or never” and “no risk” (hard rule)
  - Fail to select “seldom or never” (soft rule)
- Screening task
  - Fail to select “always” and “no risk”
- Response time
  - Fastest 5% of respondents
- Response variability
  - No variability across frequency or criticality scales (excluding foil tasks)
- Person misfit
  - Rasch outfit mean square > 1.4 for either scale (excluding foil tasks)

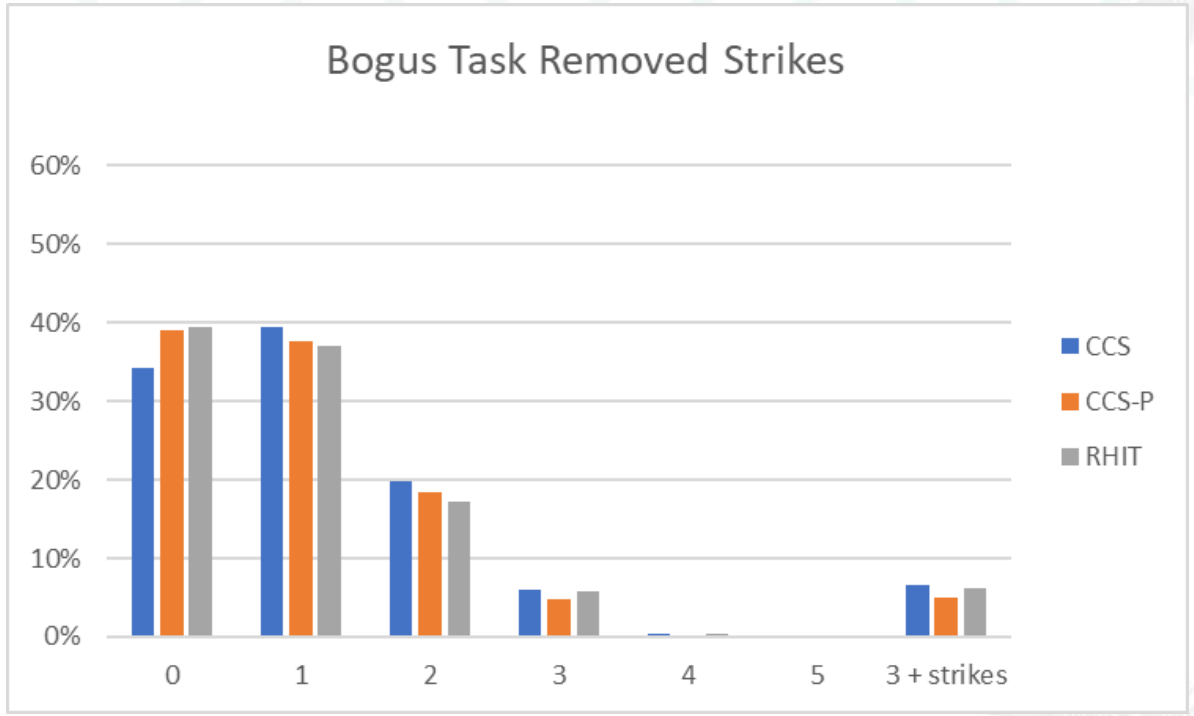
Results


Bogus Task - Hard Rule Strikes
60%
50%
40%
30%
20%
10%
0%
0
1
2
3
4
5
3 + strikes
CCS
CCS-P
RHIT


Bogus Task - Soft Rule Strikes
60%
50%
40%
30%
20%
10%
0%
0
1
2
3
4
5
3 + strikes
CCS
CCS-P
RHIT

Results


Non-Intrusive Only Strikes
60%
50%
40%
30%
20%
10%
0%
0
1
2
3
4
5
2 + strikes
CCS
CCS-P
RHIT


Bogus Task Removed Strikes
60%
50%
40%
30%
20%
10%
0%
0
1
2
3
4
5
3 + strikes
CCS
CCS-P
RHIT

Results – Average Classification Agreement

	CCS	CCS-P	RHIT
Speed	74%	75%	71%
Bogus Task Hard Rule	49%	49%	37%
Bogus Task Soft Rule	71%	70%	61%
Screening Task	58%	60%	58%
Frequency No Variance	74%	76%	72%
Criticality No Variance	74%	75%	71%
Frequency Misfit	69%	68%	63%
Criticality Misfit	62%	65%	63%

Impact – Task Weight and Mean Differences

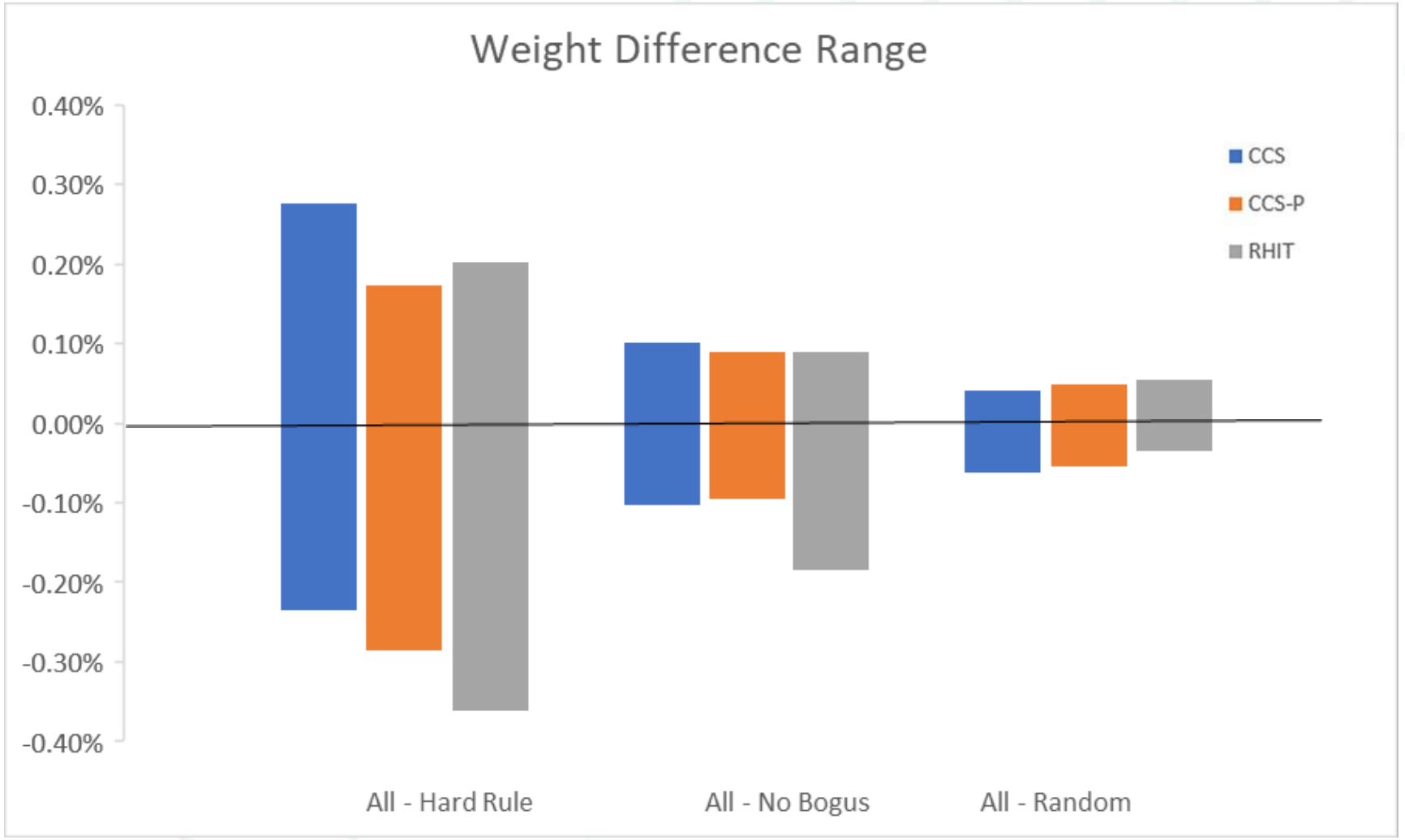
- Task weight/mean all respondents – Task weight/mean careless removed

- Task 1.01 weight difference
  - 5.57% using all respondents
  - 5.71% removing “careless” respondents
  - - 0.14% difference

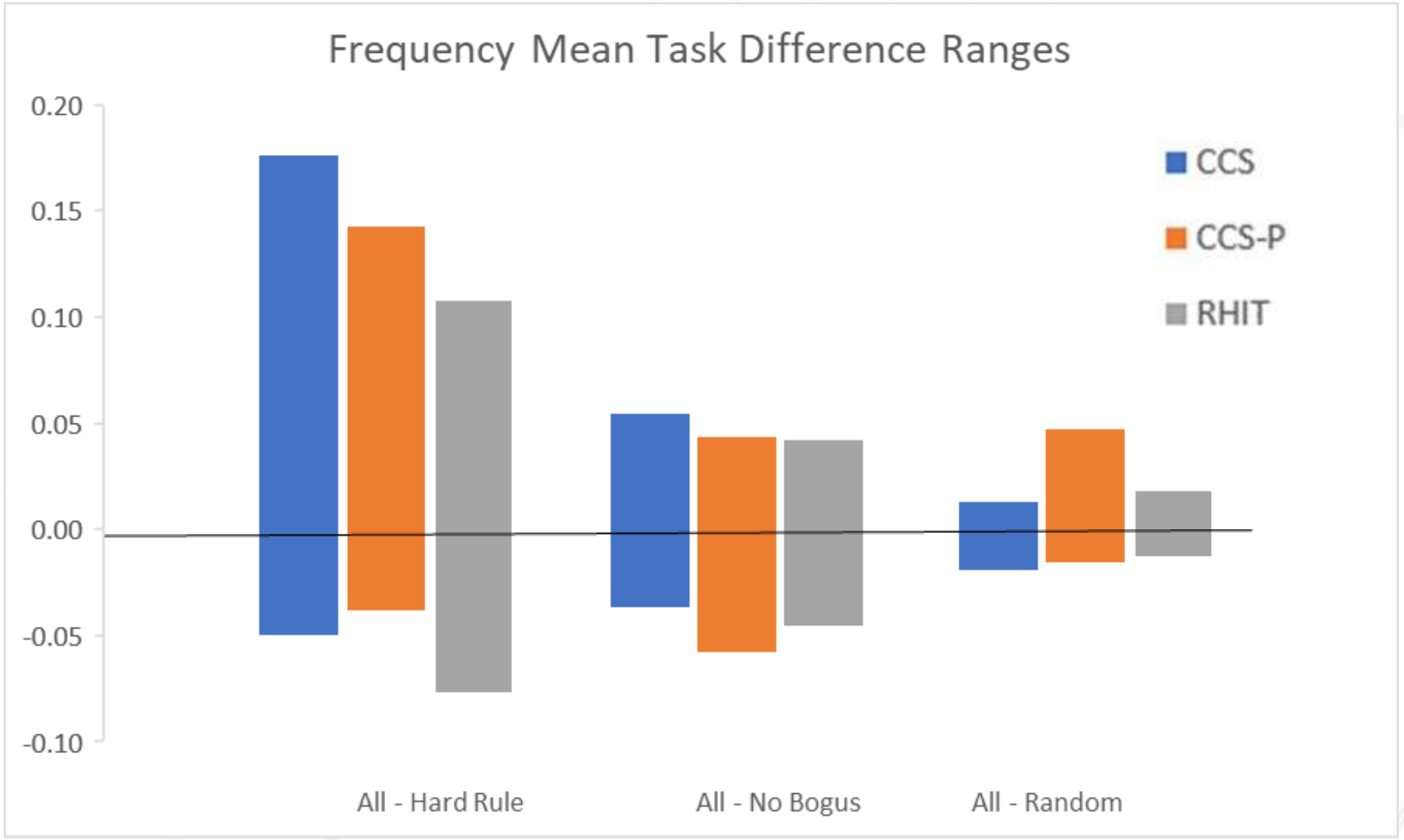
- Task 1.01 Criticality scale mean difference
  - 3.77 using all respondents
  - 3.78 removing “careless” respondents
  - - 0.01 difference

- Looked at range of differences across tasks
  - 3 strikes bogus task hard rule (maximum number removed)
  - 3 strikes no bogus task (minimum number removed)
  - Random sample equal to number removed under hard rule

Impact – Task Weight Differences



Impact – Task Mean Rating Differences



Further Investigation

- Task list length
- Better incentives
- Careless respondent demographics
- Disclaimer statement placement

Thoughts About Next Time

- Bogus tasks
- Screening tasks
- IRT person fit
- Response variability
- Response time