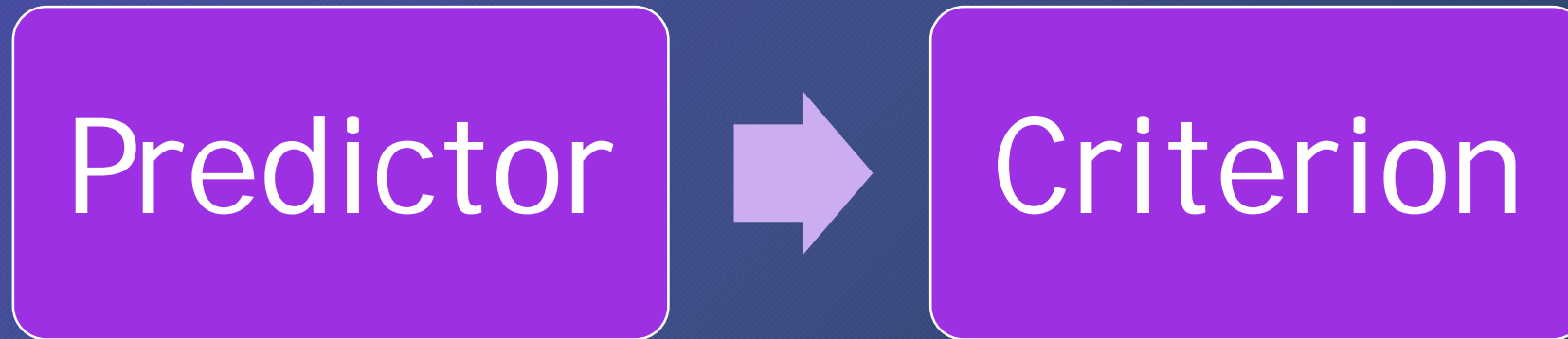


Alternative Validity Displays and Confidence Judgments: A Lens Modeling Perspective

Don Zhang, Jason Harman
Louisiana State University
zshang1@lsu.edu

Predictive Validity



Predictive validity is the degree to which a predictor – often an assessment instrument or technique – predicts a job-related criterion (e.g., job performance)

Selection Validity

“Choosing personnel selection procedures could be so simple: Grab your copy of Schmidt and Hunter (1998) and read their Table 1...”

Personnel measures	Validity (<i>r</i>)
GMA Tests ^a	.56
Integrity tests ^b	.38
Conscientiousness tests ^c	.30
Employment interviews (structured and unstructured) ^d	.35

Traditional Validity Statistics

Pearson's correlation coefficient - r

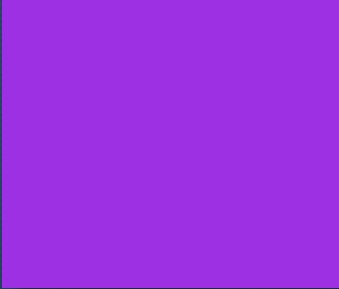
Coefficient of determination - r^2



Correlation is hard to interpret

- Most decision makers are not familiar with correlation
- Correlations lack context
- People underestimate the practical utility of correlations

Alternatives Validity Displays



Binomial effect size display (BESD)

$r = 0.2$

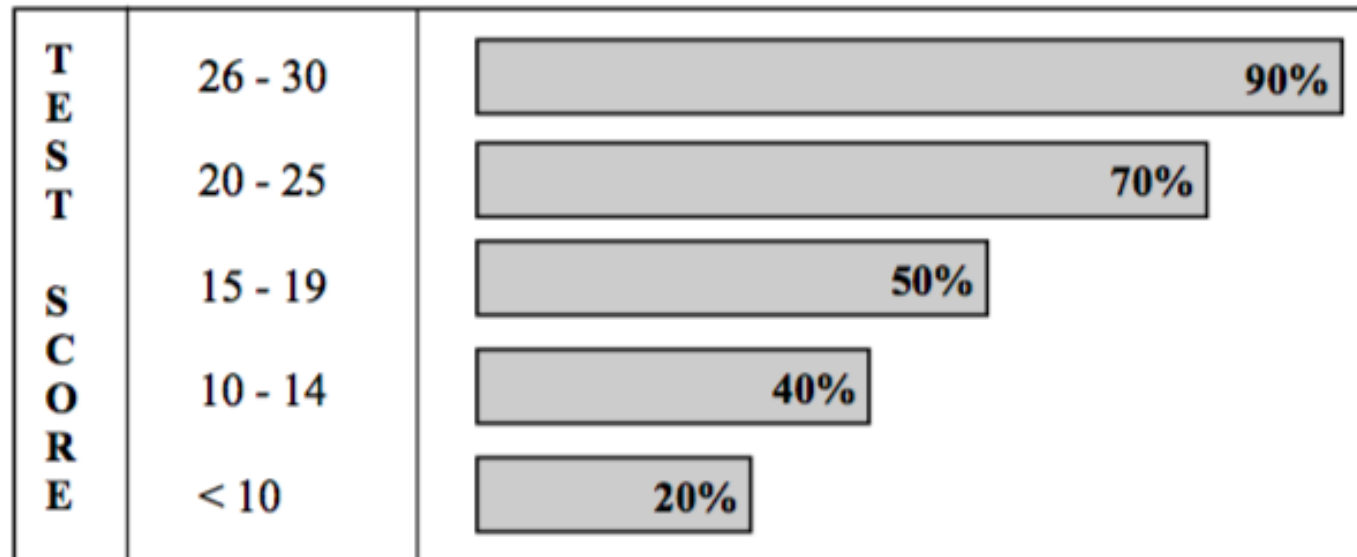
	Above 3.0 GPA	Below 3.0 GPA
Above average ACT	55	45
Below average ACT	45	55

$r = 0.5$

	Above 3.0 GPA	Below 3.0 GPA
Above average ACT	75	25
Below average ACT	25	75

Expectancy chart

Probability an Applicant Will be Successful on the Job Given a Certain Test Score



Benefits of alternative effect size displays

Alternative effect sizes are perceived as more useful and easier to understand¹

People judge selection instruments as more useful and are more willing to adopt them when effect sizes are presented with graphical alternative displays²

Limitations: past research primarily focused on subjective judgments of alternative validity statistics.

1. Brooks, M. E., Dalal, D. K., & Nolan, K. P. (2014). *The Journal of Applied Psychology*, 99(2), 332-40.

2. Zhang, Highhouse, Brooks, Zhang, *r&r*

A Lens Model Perspective

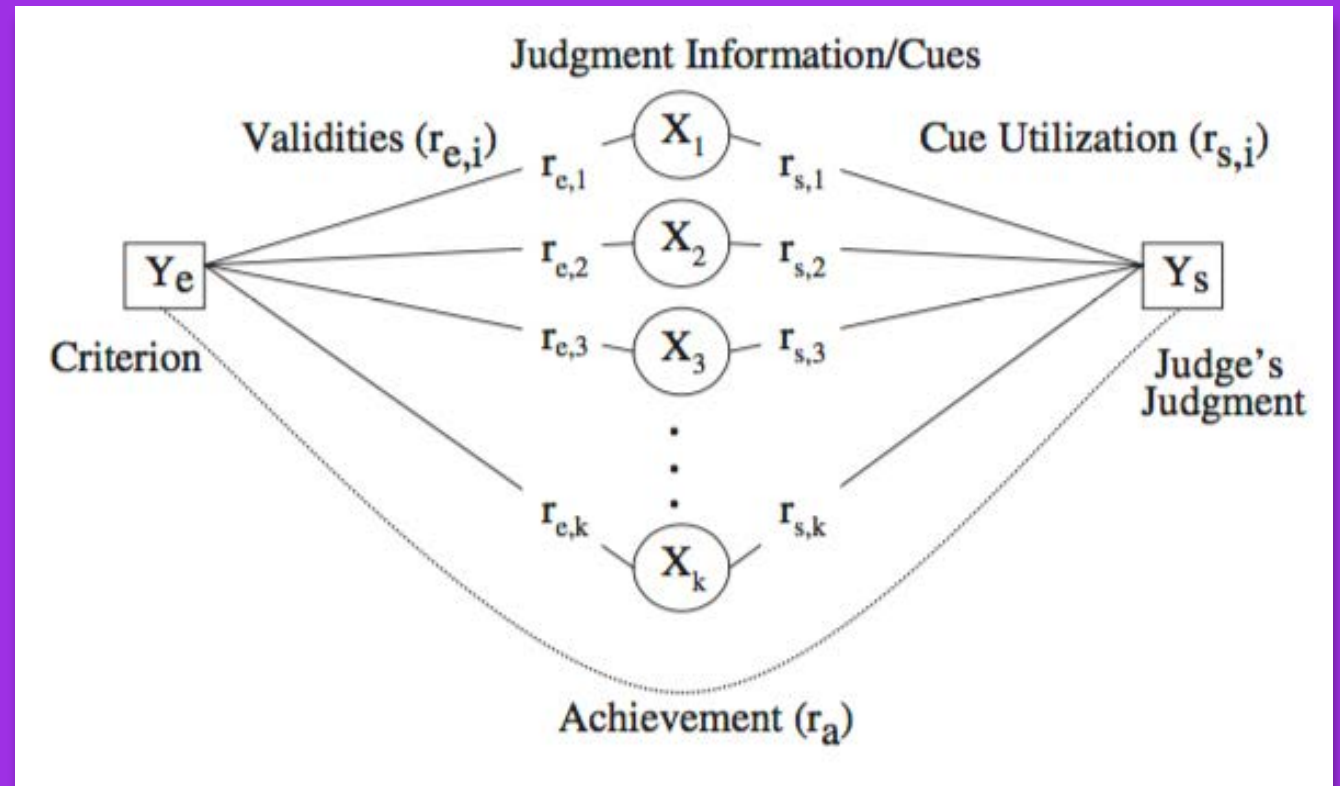
Criterion: Outcome of interest

Cues: Relevant information about target

Validities: Relationship between predictor

Cue utilization: The degree to which a cue is considered in the judgment

Achievement: Congruence between judgment and criterion



Validity information as feedback (Task Information)

The relation between cue(s) and the criterion is called Task Information¹

Presence of task information improved people's performance²

Display format (graphical vs. non-graphical) did NOT improve accuracy³

1. Balzer, W. K., Doherty, M. E., & O'Connor, R. (1989). Effects of Cognitive Feedback on Performance. *Psychological Bulletin*, 106(3), 24.
2. Balzer, W. K., Sulsky, L. M., Hammer, L. B., & Sumner, K. E. (1992). *Organizational Behavior and Human Decision Processes*, 53(1), 35-54.
3. Doherty, M. E., & Balzer, W. K. (1988). Cognitive feedback. In *Advances in psychology* (Vol. 54, pp. 163-197). Elsevier.

Does display type matter?

Cue utilization

Accuracy

Confidence Calibration

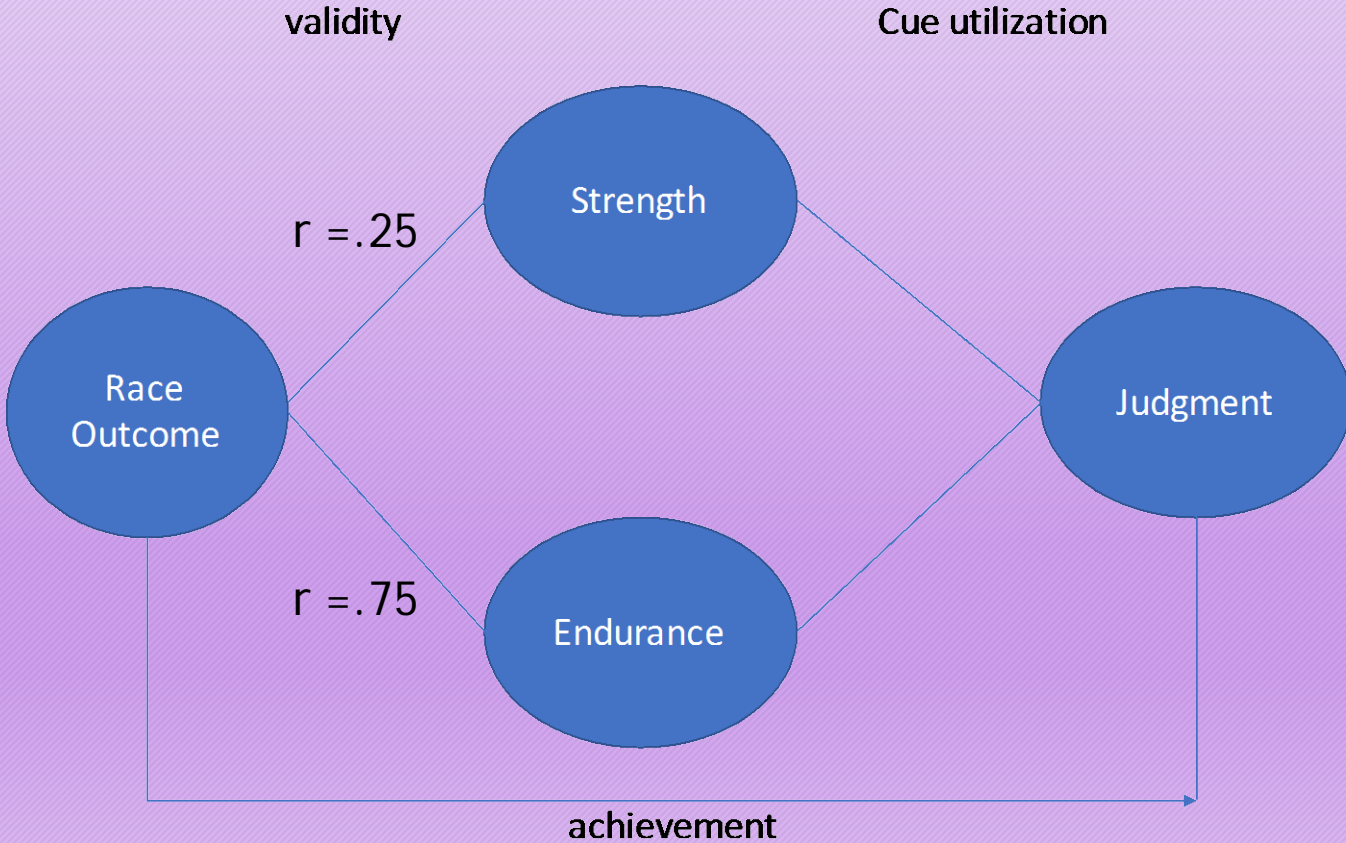
Experiment - Bicycle race predictions

Participants are presented with hypothetical bicycle race competition where cyclists are competing to qualify for the national competition.

Participants are provided information on the relationship between the physical attribute and performance on the race (i.e., validity).

Participants indicated - based on physical attributes of the bikers - how confident that the racer will qualify (0% to 100%).

Lens model
of bike
race



Experiment

Participants were only presented with ONE attribute

Trials = 90 predictions

Independent Variables (between-subj):

- Validity display
 - Correlation
 - Binomial Effect Size Display
 - Expectancy Chart
- Validity of physical attribute
 - Small: $r = .25$ (endurance)
 - Large: $r = .75$ (strength)

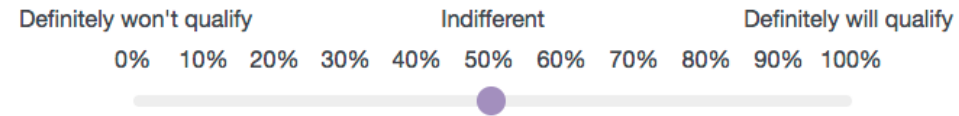
Dependent Variable: How likely will the racer qualify for the race (0~100%)

Trial: 1 out of 90

Prediction Screen

What is the probability that Omar will qualify for the race?

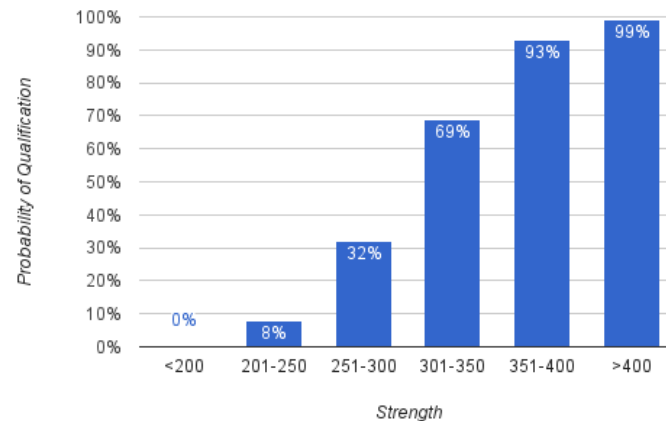
Name	Omar
Strength	229 lbs



You may consult the information below to make your judgment:

Validity Screen

Average Strength: 300 lbs (95% range: 200 ~ 400 lbs)



Correlation

You may consult the information below to make your judgment:

To summarize the information on strength:

Average strength of racers	300 lbs
95% range	200 ~ 400 lbs
Correlation of strength to race performance	$r = 0.75$

Binomial Effect Size Display

You may consult the information below to make your judgment:

Average Strength: 300 lbs (95% range: 200 ~ 400 lbs)

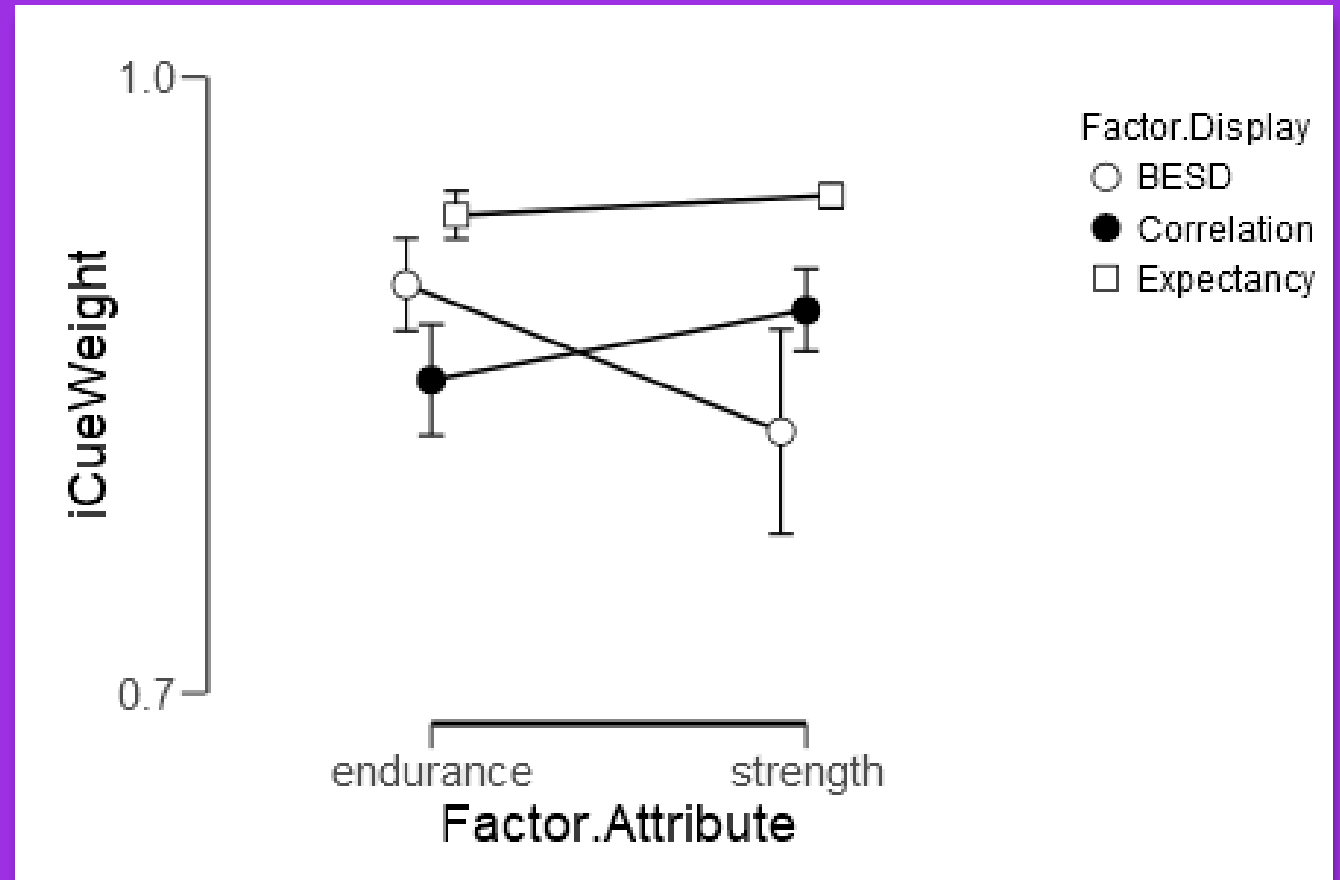
Strength	Qualified	Did not qualify
Less than 300lbs	24%	76%
Greater than 300lbs	76%	24%



RESULTS

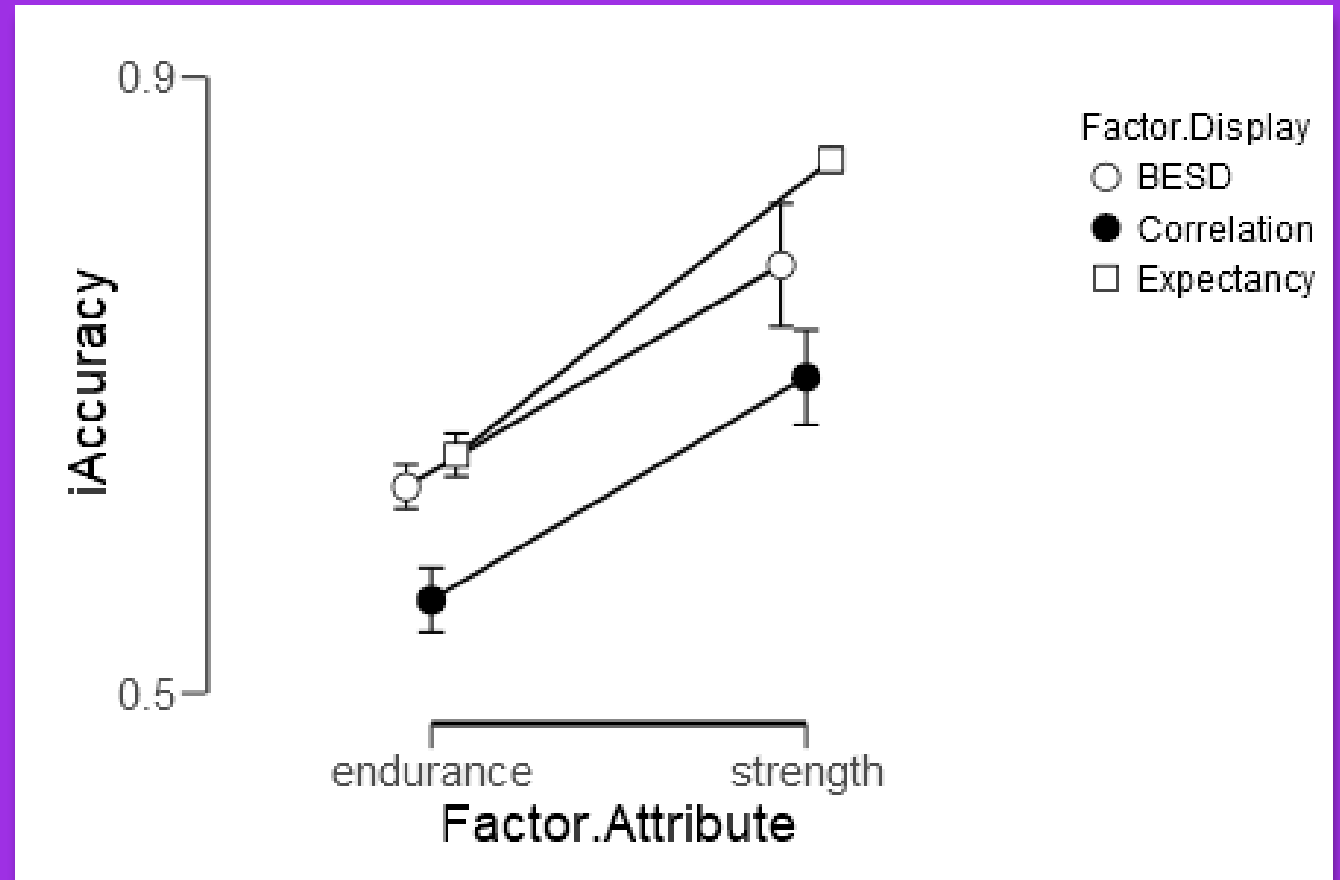
Results - Cue Utilization

	DF	MS	F
Attribute	1	.00	.00
Display	2	.05	6.01**
Interaction	2	.02	2.67



Results - Accuracy

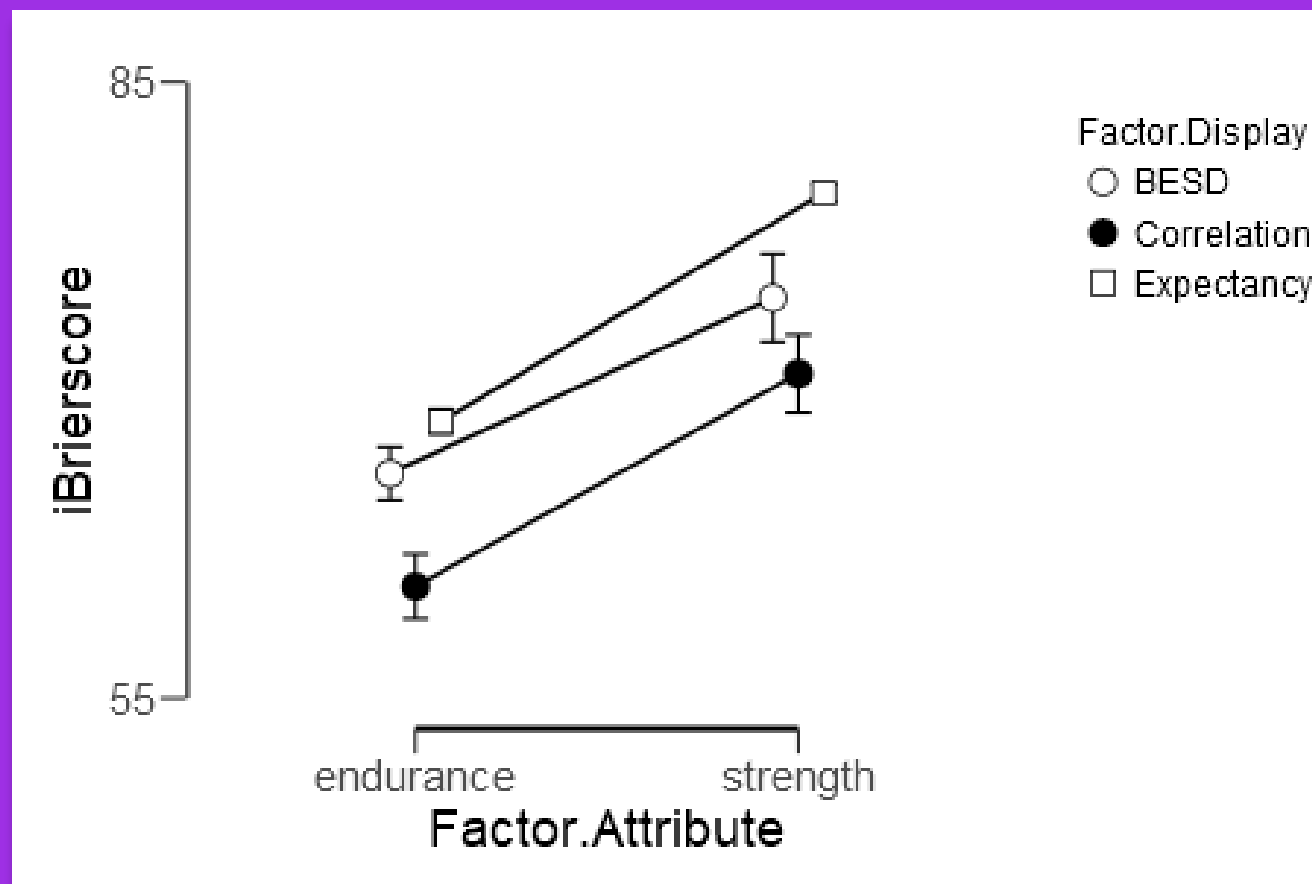
	DF	MS	F
Attribute	1	.69	83.54**
Display	2	.14	16.82**
Interaction	2	.01	.83



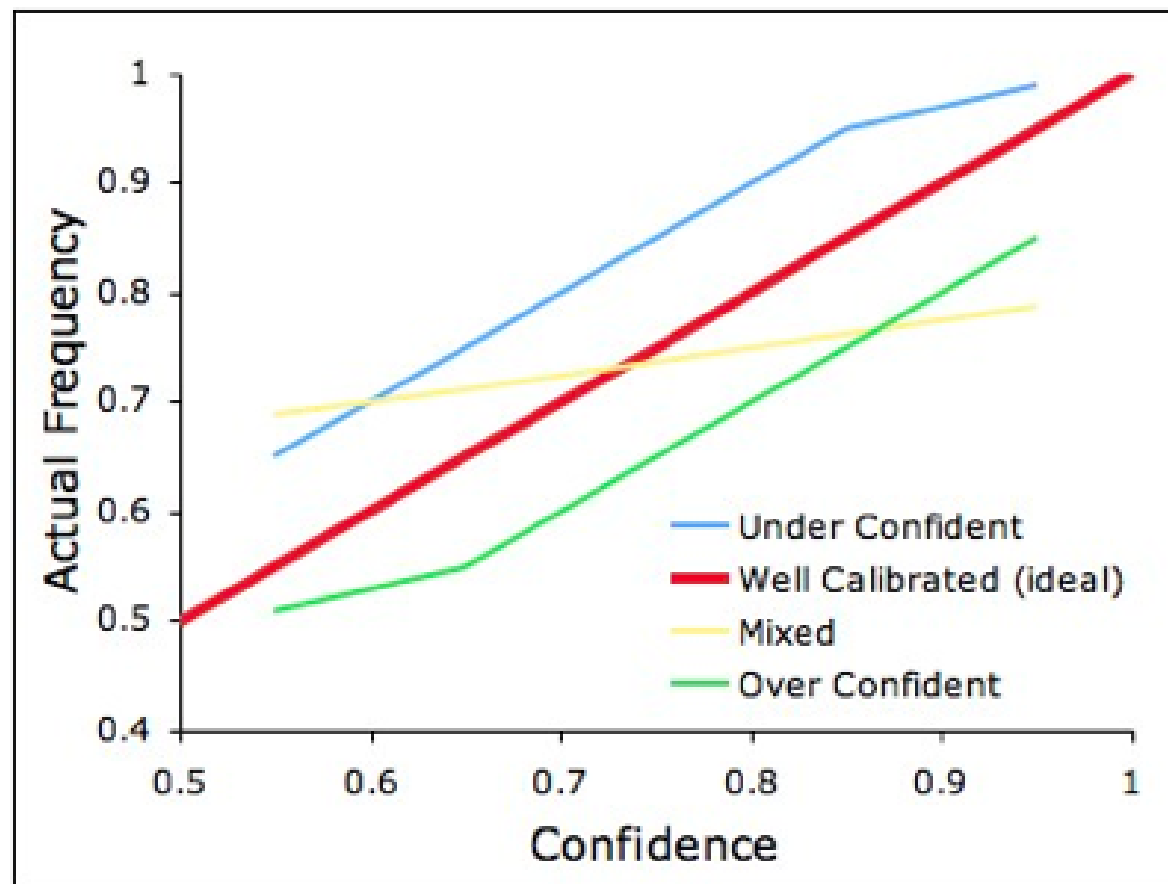
Results - Confidence Calibration

Brier score is a numeric summary of confidence calibration

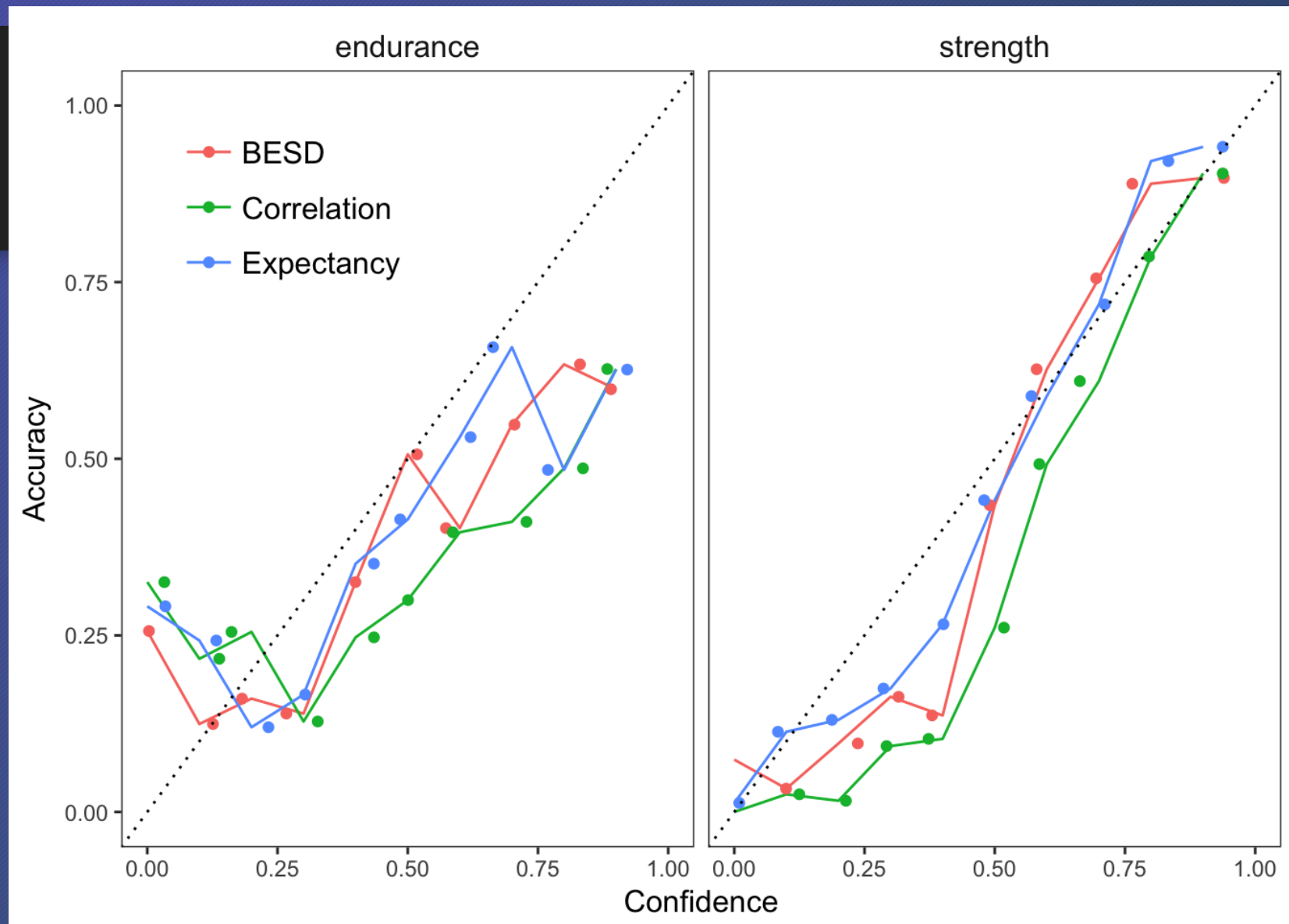
	DF	MS	F
Attribute	1	2775.5	81.11**
Display	2	696.2	20.34**
Interaction	2	14.18	.41



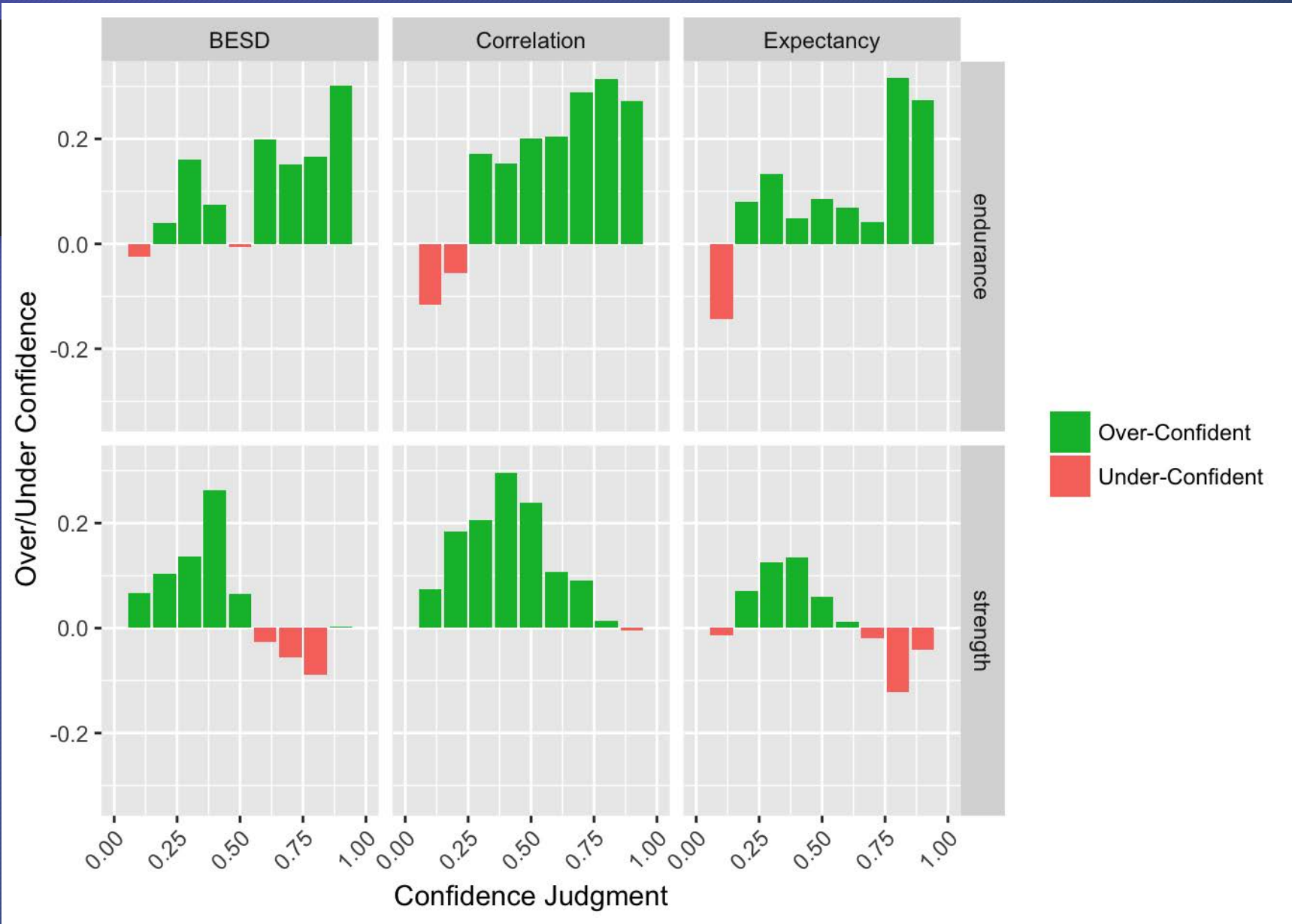
Confidence Calibration



Calibration curves



Over-Under-Confidence



Summary of Results

Cue Utilization: Expectancy > BESD/Correlation

Accuracy: Expectancy > BESD/Correlation

Confidence
Calibration: Expectancy > BESD > Correlation

Discussion

- Contrary to previous research, type of validity display in task information does matter.
- People make better judgments when validity information is presented with expectancy charts than BESD and correlation
 - People are more accurate
 - People are more calibrated
 - People are less overconfident
- Implications

Thank you