



How Do You Rate What You Don't Know: The Impact of Familiarity Between Raters and Ratees on Performance Evaluations

John M. Ford
Michael D. Blair




June 26, 2006 CWH Research, Inc. ¹



Critical Role of Performance Ratings

- **Administrative Decisions**
 - Compensation
 - Promotion
 - Termination/Retention
- **Developmental Processes**
 - Training
 - Performance Management
- **Research & Development**
 - Validation

June 27, 2006 CWH Research, Inc. ²




Ratings Measure More Than Performance

- Ratee Effects (Actual Performance)
- Rater Effects
 - Individual Rater
 - Rater Perspective
- Measurement Error

Scullen, Mount, & Goff (2000)


June 27, 2006 CWH Research, Inc. ³



Why Raters Influence Ratings So Much

Opportunity to Perform
Opportunity to Observe
Attend
Clear Standard for Evaluating
Weighting Performance Dimensions
Rater Goals
Form of Measurement

June 27, 2006 CWH Research, Inc. ⁴



Importance of Opportunity to Observe

- Having the opportunity to observe ratee performance is the first step in the rater evaluation process.
 - Opportunity to observe is necessary for rater to have true performance information.
- Lack of opportunity to observe limits reliability and accuracy.
 - Reliability is the ratio of true to observed variance (Nunnally, 1978)
 - Lack of convergence
 - Low reliability limits validity
- With limited knowledge of true performance, raters are forced to rely on heuristics and biases.
 - Halo

June 27, 2006 CWH Research, Inc. ⁵




Past Research

Rothstein (1990)

- Operationalized opportunity to observe as the length of time that the ratee had been in supervisory position.
- Found a strong, nonlinear relationship between tenure and interrater reliability.
- Although the relationship between tenure and reliability was strongest during the first year, it also played an important role even after 5 years.
- Implications—it is possible to rate an employee too soon.

June 27, 2006 CWH Research, Inc. ⁶




Study

- 1,292 performance ratings collected on 425 entry-level firefighters from 3 different organizations.
 - Each ratee was evaluated by 1-6 raters.
 - 388 Supervisor Ratings
 - 904 Peer Ratings
- Ratees were evaluated on 14 different dimensions.
- Organizations were specifically asked to identify raters that were familiar with the ratee's job performance.
- Opportunity to Observe (Familiarity) was operationalized with 3 different items.

7
CWH Research, Inc.

June 27, 2006




How familiar are you with the ratee's job performance?

- 1 = Not Familiar
- 2 = Somewhat Familiar
- 3 = Familiar
- 4 = Very Familiar

8
CWH Research, Inc.

June 27, 2006




How often do you typically observe this person's actual performance?

- 1 = Never
- 2 = Yearly
- 3 = Monthly
- 4 = Weekly
- 5 = Once Per Shift

9
CWH Research, Inc.

June 27, 2006




How long have you worked with the ratee?

- 1 = Less than 1 year
- 2 = 1-3 Years
- 3 = 4-7 Years
- 4 = 8-10 Years
- 5 = > 10 Years

10
CWH Research, Inc.

June 27, 2006




Means and Correlation Coefficients

	Familiarity	Familiarity	Frequency
Familiarity	3.3		
Frequency	.26**	4.7	
Length of Exposure	.44**	.11**	1.9

** p < .01 (2-tailed), Means in Diagonals

11
CWH Research, Inc.

June 27, 2006



Halo Rating Error

- Tendency for particularly good or bad performance in one area to influence your opinion of performance in another area.
- Results in lack of variance in performance dimension ratings.
- Within Score Consistency (WSC) evaluates consistency across performance dimensions (Fox, 1981).
 - Operationalized as average variance across dimensions.
 - A "0" indicates no variance in ratings.
- Expectation—Variability would increase as familiarity increases.

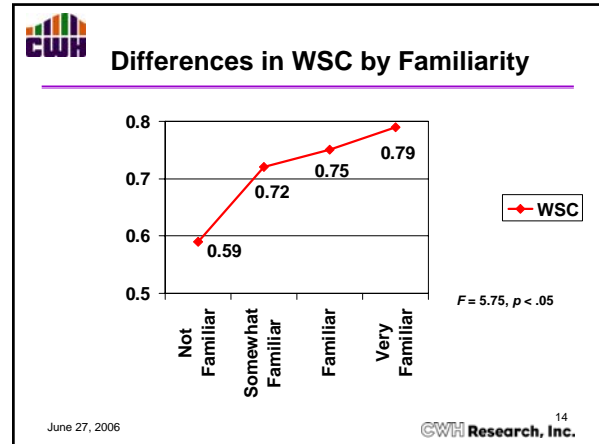
12
CWH Research, Inc.

June 27, 2006

Frequency Distribution of Familiarity Variable

	N
Not Familiar	22
Somewhat Familiar	192
Familiar	414
Very Familiar	536

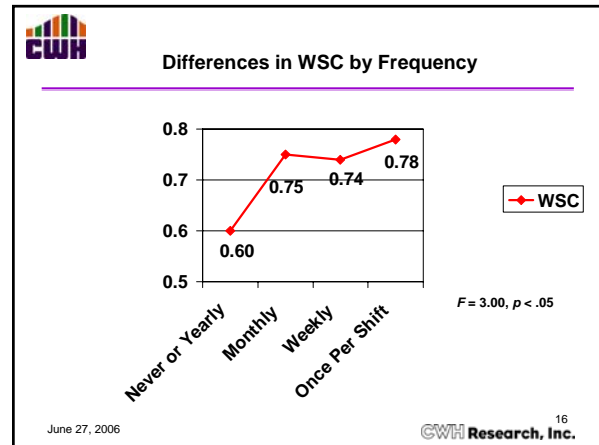
June 27, 2006 CWH Research, Inc. 13



Frequency Distribution of Frequency Variable

	N
Never	9
Yearly	5
Monthly	36
Weekly	180
Once Per Shift	821

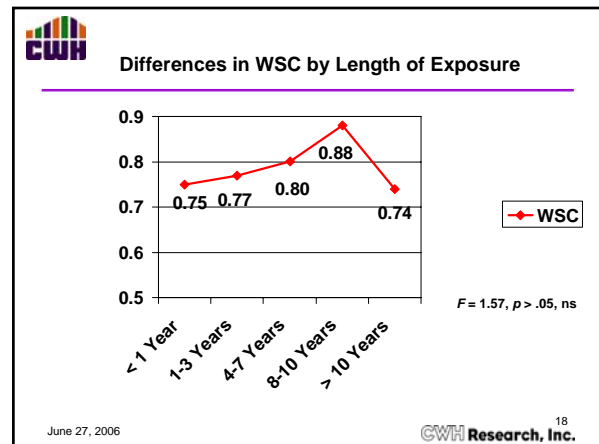
June 27, 2006 CWH Research, Inc. 15




Frequency Distribution of Length of Exposure Variable

	N
Less than 1 Year	443
1-3 Years	468
4-7 Years	136
8-10 Years	23
> 10 Years	36

June 27, 2006 CWH Research, Inc. 17






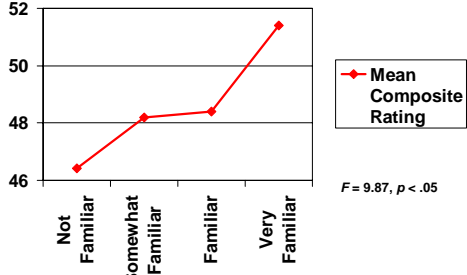
Magnitude of Mean Ratings

- Two Possibilities
 - Familiarity Leads to Lower Mean Ratings
 - Low familiarity = Leniency Rating Error
 - Familiarity Leads to Higher Mean Ratings
 - Closer relationship = Higher Affect
 - Mutually Reinforcing Relationship
- Expectation-Mean ratings would decrease as familiarity increases.

June 27, 2006 19
CWH Research, Inc.




Differences in Mean Performance Rating (t-scores) by Familiarity



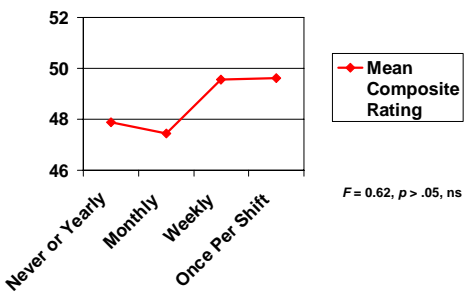
Familiarity	Mean Composite Rating
Not Familiar	~46.5
Somewhat Familiar	~48.2
Familiar	~48.5
Very Familiar	~51.5

$F = 9.87, p < .05$

June 27, 2006 20
CWH Research, Inc.




Differences in Mean Performance Rating (t-scores) by Frequency



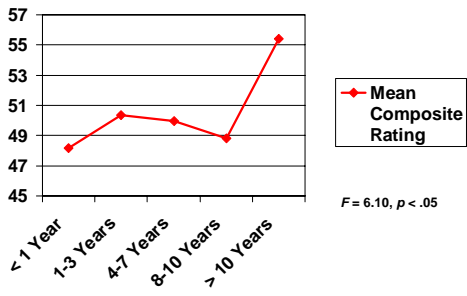
Frequency	Mean Composite Rating
Never or Yearly	~48.0
Monthly	~47.5
Weekly	~49.5
Once Per Shift	~49.5

$F = 0.62, p > .05, ns$

June 27, 2006 21
CWH Research, Inc.




Differences in Mean Performance Rating (t-scores) by Length of Exposure



Length of Exposure	Mean Composite Rating
< 1 Year	~48.5
1-3 Years	~50.0
4-7 Years	~50.0
8-10 Years	~49.0
> 10 Years	~55.5

$F = 6.10, p < .05$


June 27, 2006 22
CWH Research, Inc.



Familiarity and Accuracy

- We did not have any way to assess the accuracy of ratings.
- However, we did have scores on a validated standardized entry-level written firefighter/EMT test that matched 870 of the ratings.
- Expectation-Validity would increase as familiarity increases.

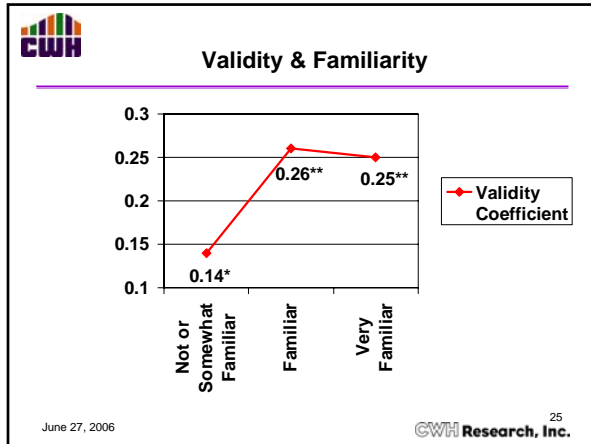
June 27, 2006 23
CWH Research, Inc.



Frequency Distribution of Familiarity Variable

Familiarity Variable	N
Not Familiar	16
Somewhat Familiar	135
Familiar	292
Very Familiar	339

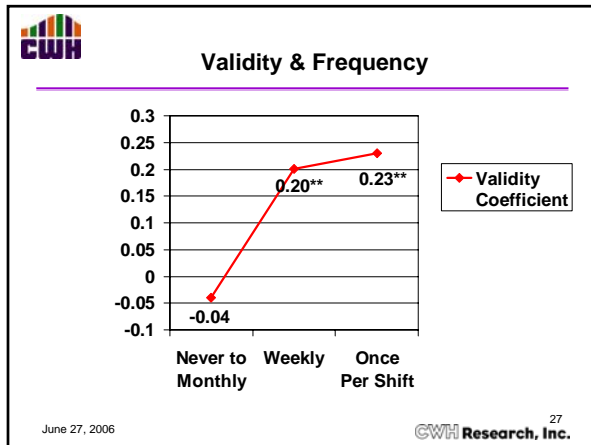
June 27, 2006 24
CWH Research, Inc.



Frequency Distribution of Frequency Variable

Frequency Variable	N
Never	5
Yearly	3
Monthly	26
Weekly	115
Once Per Shift	575

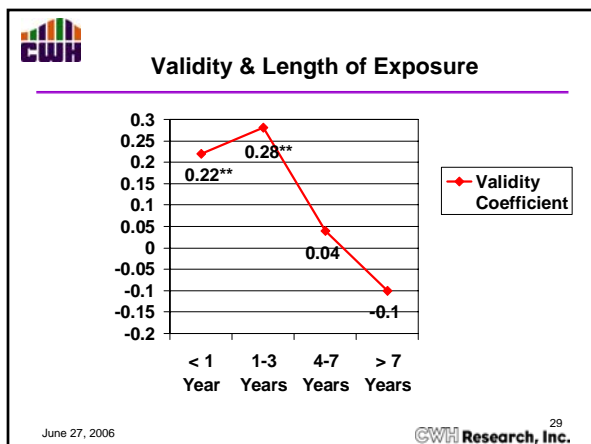
June 27, 2006 CWH Research, Inc. 26



Frequency Distribution of Length of Exposure Variable

Length of Exposure Variable	N
Less than 1 Year	321
1-3 Years	322
4-7 Years	87
8-10 Years	15
> 10 Years	16

June 27, 2006 CWH Research, Inc. 28

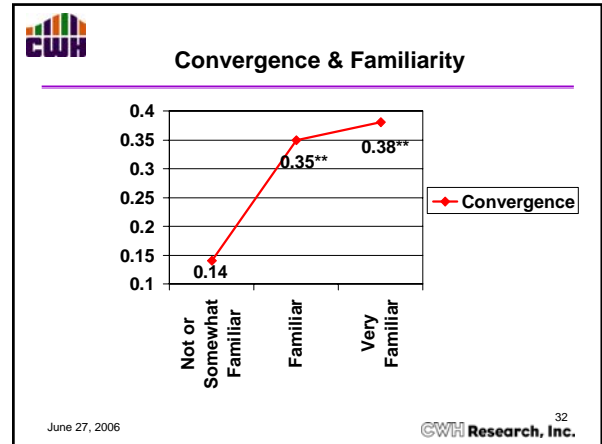


- ### Convergence
- One of the explanations often given for lack of convergence (i.e., correlations) between rater sources is differential opportunities to observe performance.
 - Increased Opportunity to Observe should be related to being exposed to a larger sample of ratee performance.
 - If both raters have observed large samples of performance, there should be greater overlap in ratings.
 - Convergence was assessed by correlating supervisor and peer ratings.
 - Expectation-When raters have similar familiarity with the ratee, convergence should increase as familiarity increases.
- June 27, 2006 CWH Research, Inc. 30

CWH **Frequency Distribution of Matched Raters on Familiarity Variable**

	N	N
Not Familiar	6	42
Somewhat Familiar	31	
Familiar	87	87
Very Familiar	182	182

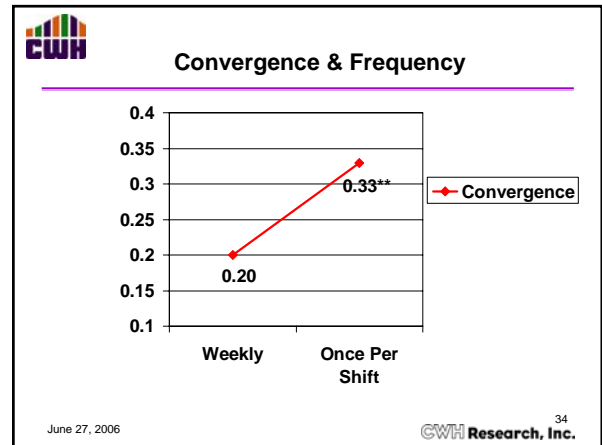
June 27, 2006 CWH Research, Inc. ³¹



CWH **Frequency Distribution of Matched Raters on Frequency Variable**

	N
Never	0
Yearly	0
Monthly	4
Weekly	28
Once Per Shift	385

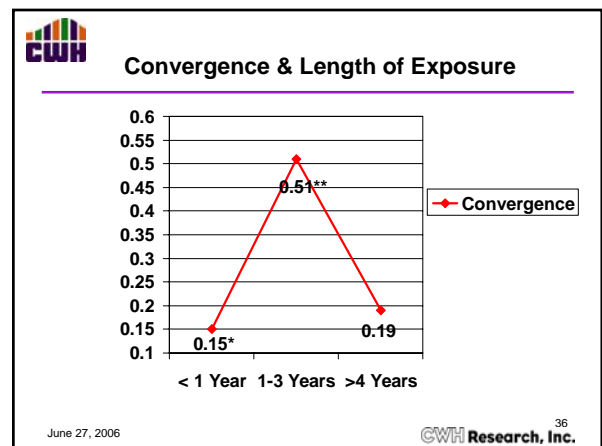
June 27, 2006 CWH Research, Inc. ³³



CWH **Frequency Distribution of Matched Raters on Length of Exposure Variable**

	N	N
Less than 1 Year	163	163
1-3 Years	142	142
4-7 Years	21	43
8-10 Years	3	
> 10 Years	1	

June 27, 2006 CWH Research, Inc. ³⁵





Implications

- Familiarity or Opportunity to Observe impacts the variability, magnitude, validity, and convergence of performance ratings.
- Efforts should be made to ensure that raters are familiar with ratee job performance before ratings are utilized.
- Although frequency of exposure and length of exposure are related to familiarity, they are not perfect proxies.
- If possible, familiarity should be assessed.
- It may be possible to rate an employee too late.
 - 1-3 years may be optimal time frame.

June 27, 2006

³⁷
CWH Research, Inc.



Limitations and Future Research

- We did not track how many raters chose not to provide ratings due to lack of familiarity.
- Small sample sizes
- Aggregation
- Could not completely evaluate accuracy of ratings.
- What does "Somewhat Familiar" mean?

June 27, 2006

³⁸
CWH Research, Inc.