

# Introduction to Evaluation

Nancy Skehan, MD FACP



# Objectives

- Define summative and formative assessment
- Describe the role of careful observation as an evaluation skill
- Use the  $+\Delta$  tool to organize observations and feedback
- Identify barriers to providing effective feedback
- Review strategies to mitigate barriers to feedback



# Evaluation is a judgment based on observations

- Summative
  - Judgment of **competence** at the end of a learning experience by an evaluator
  - To determine to what degree goals have been met.
- Formative
  - Ongoing, to **modify behaviors**
  - Feedback



# Observations drive evaluation

- Real time
- Specific and objective
- **Planned:** Observer knows in advance what they are looking for



# How do you plan an observation?

- Define the content of the skill (Task Analysis)
- Getting to school on time
  - Waking up
  - Getting dressed
  - Eating breakfast
  - Coat and shoes
  - Make the bus



# Organizing observations: the +/-Δ

+	Δ
<p data-bbox="682 625 1121 868">Examples of desired behavior/skill</p> <ul data-bbox="591 982 1166 1225" style="list-style-type: none"><li data-bbox="591 982 1166 1139">• Woke up to the alarm</li><li data-bbox="591 1153 1166 1225">• Brushed teeth</li></ul>	<p data-bbox="1457 625 2033 868">Examples of behavior/skill that requires <b>change</b></p> <ul data-bbox="1421 939 2063 1196" style="list-style-type: none"><li data-bbox="1421 939 2063 1025">• Forgot socks</li><li data-bbox="1421 1039 2063 1196">• Got mad about breakfast options</li></ul>

# The Medical Interview



## Greeting

Introduced self  
Good eye contact

Didn't mention student status

## Chief Complaint

Open-ended questions  
*"Describe..."*

Interrupted  
*"Pt: I need to talk about..."*  
*Student: "Yes, but when..."*

## PMH

Specific questions  
*"How much alcohol?"*

Missed associated symptoms

Good summary

Missed empathy  
*"I can't afford to be sick."*



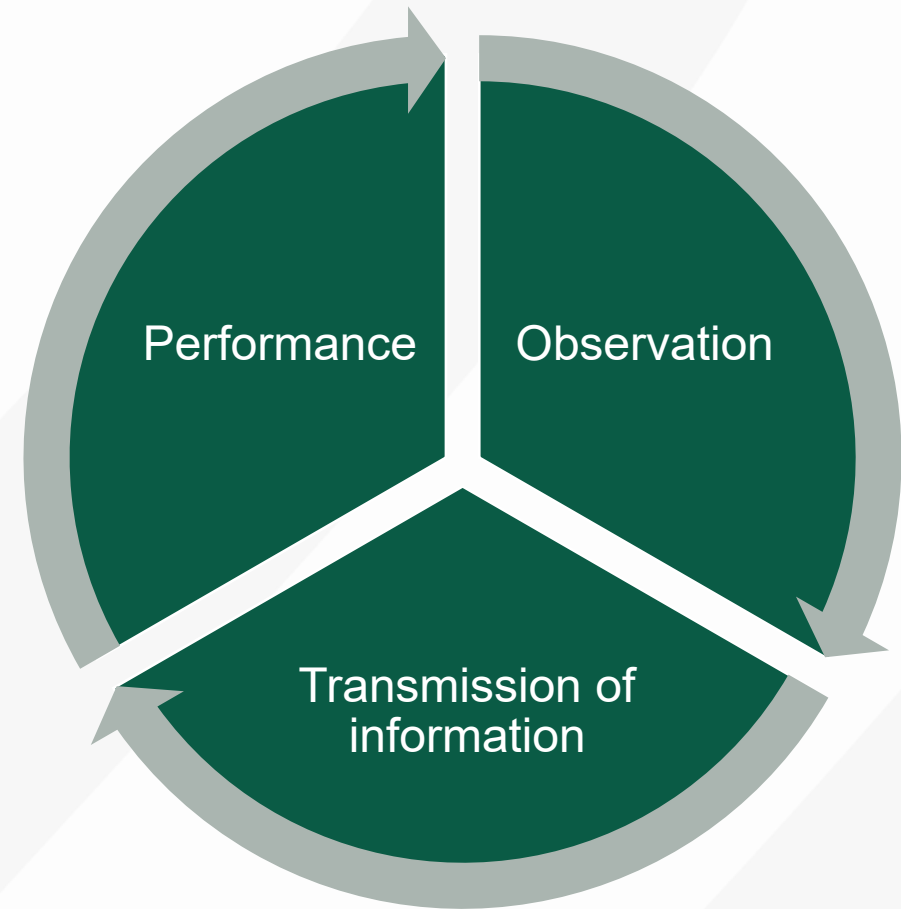
# What is Feedback

- “The return of a fraction of the output signal from one stage of a circuit...to the same or a preceding stage...tending to increase or decrease the amplification.”
- A *cycle*



# What is Feedback?

“**Specific information** about the comparison between a trainee’s **observed performance and a standard**, given with the **intent to improve** the trainee’s performance.”



# Barriers Exist to Providing Feedback





# How do we make it easier?

# Normalize it



# SET THE STANDARD



**Be specific**





**Table 4. Percentage of applicants by race/ethnicity group for whom each descriptive word was used at least once in the Medical Student Performance Evaluation.**

	Black n = 346	Hispanic n = 202	White n = 2,740	Asian n = 1,281	Multi n = 336	Other n = 109	p-value (* Alpha = .002)
<b>Word Categories</b>							
<b>Standout Words</b>							
Exceptional	50%	52%	64%	54%	64%	58%	<0.001*
Best	41%	44%	54%	49%	50%	58%	<0.001*
Outstanding	77%	84%	86%	79%	82%	88%	<0.001*
Superb	30%	32%	38%	36%	38%	38%	0.025
Stellar	7%	7%	10%	8%	9%	13%	0.067
Excellent	91%	90%	93%	93%	95%	97%	0.050
Phenomenal	3%	5%	5%	5%	5%	8%	0.213
<b>Ability</b>							
Intelligent	40%	43%	49%	50%	46%	44%	0.004
Bright	43%	44%	57%	54%	54%	52%	<0.001*
Talent	19%	18%	20%	19%	17%	15%	0.760
Brilliant	3%	1%	3%	3%	4%	2%	0.420
Competent	40%	20%	29%	27%	32%	34%	<0.001*
Smart	19%	18%	24%	23%	24%	28%	0.129
Gifted	5%	5%	6%	5%	7%	5%	0.342
<b>Grindstone</b>							
Organized	71%	74%	80%	77%	82%	79%	0.001*
Hardworking	76%	74%	77%	78%	77%	77%	0.790
Conscientious	36%	28%	32%	34%	33%	37%	0.337
Diligent	42%	32%	36%	37%	34%	31%	0.115
<b>Compassion</b>							
Caring	47%	50%	51%	49%	51%	55%	0.750
Kind	35%	32%	33%	34%	36%	42%	0.332
Empathy	36%	49%	40%	35%	38%	45%	0.003
Compassionate	56%	49%	54%	51%	51%	63%	0.480

<https://doi.org/10.1371/journal.pone.0181659.t004>

Ross DA, Boatright D, Nunez-Smith M, Jordan A, Chekroud A, et al. (2017) Differences in words used to describe racial and gender groups in Medical Student Performance Evaluations. PLOS ONE 12(8): e0181659. <https://doi.org/10.1371/journal.pone.0181659>  
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0181659>

**Table 5. Associations between race/ethnicity and adjective use controlling for USMLE Step 1 scores.**

Applicant Characteristics	Odds Ratios				
	(95% Confidence Interval)				
	Exceptional	Best	Outstanding	Bright	Competent
Race/ethnicity*					
Black	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
White	1.39 (1.09–1.75)	1.38 (1.09–1.75)	1.31 (0.99–1.74)	1.47 (1.16–1.86)	0.62 (0.49–0.79)
Hispanic	0.98 (0.69–1.40)	1.02 (0.71–1.45)	1.41 (0.89–2.22)	0.97 (0.68–1.38)	0.38 (0.26–0.58)
Asian or Pacific Islander	0.94 (0.74–1.20)	1.20 (0.94–1.53)	0.85 (0.64–1.14)	1.39 (1.09–1.77)	0.57 (0.44–0.74)
Multiracial	1.49 (1.09–2.04)	1.25 (0.92–1.70)	1.08 (0.74–1.59)	1.41 (1.04–1.91)	0.72 (0.53–0.99)
Other	1.16 (0.75–1.81)	1.75 (1.12–2.71)	1.88 (0.99–3.58)	1.31 (0.85–2.02)	0.78 (0.49–1.23)
USMLE Step 1 Score	1.02 (1.01–1.02)	1.01 (1.01–1.01)	1.02 (1.02–1.02)	1.01 (1.01–1.01)	0.99 (0.99–1.00)

\*As a category, the p-value for Race/ethnicity using a Wald test was <0.01 for each adjective

<https://doi.org/10.1371/journal.pone.0181659.t005>

Ross DA, Boatright D, Nunez-Smith M, Jordan A, Chekroud A, et al. (2017) Differences in words used to describe racial and gender groups in Medical Student Performance Evaluations. PLOS ONE 12(8): e0181659. <https://doi.org/10.1371/journal.pone.0181659>  
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0181659>

# **Make direct observations**



# Observations drive evaluation

- Real time
- Specific and objective
- **Planned:** Observer knows in advance what they are looking for





# How do you plan an observation?

- Define the content of the skill (Task Analysis)
- Getting to school on time
  - Waking up
  - Getting dressed
  - Eating breakfast
  - Coat and shoes
  - Make the bus



# Organizing observations: the +/Δ

+	Δ
<p data-bbox="639 572 1059 808">Examples of desired behavior/skill</p> <ul data-bbox="542 925 1110 1160" style="list-style-type: none"><li data-bbox="542 925 1110 1075">• Woke up to the alarm</li><li data-bbox="542 1103 1110 1160">• Brushed teeth</li></ul>	<p data-bbox="1442 572 2015 808">Examples of behavior/skill that requires <b>change</b></p> <ul data-bbox="1403 925 2040 1160" style="list-style-type: none"><li data-bbox="1403 925 2040 989">• Forgot socks</li><li data-bbox="1403 1018 2040 1160">• Got mad about breakfast options</li></ul>

# The Medical Interview



## Greeting

Introduced self  
Good eye contact

Didn't mention student status

## Chief Complaint

Open-ended questions  
*"Describe..."*

Interrupted  
*"Pt: I need to talk about..."*  
*Student: "Yes, but when..."*

## PMH

Specific questions  
*"How much alcohol?"*

Missed associated symptoms

Good summary

Missed empathy  
*"I can't afford to be sick."*

# **Make it Timely/Make it Short**

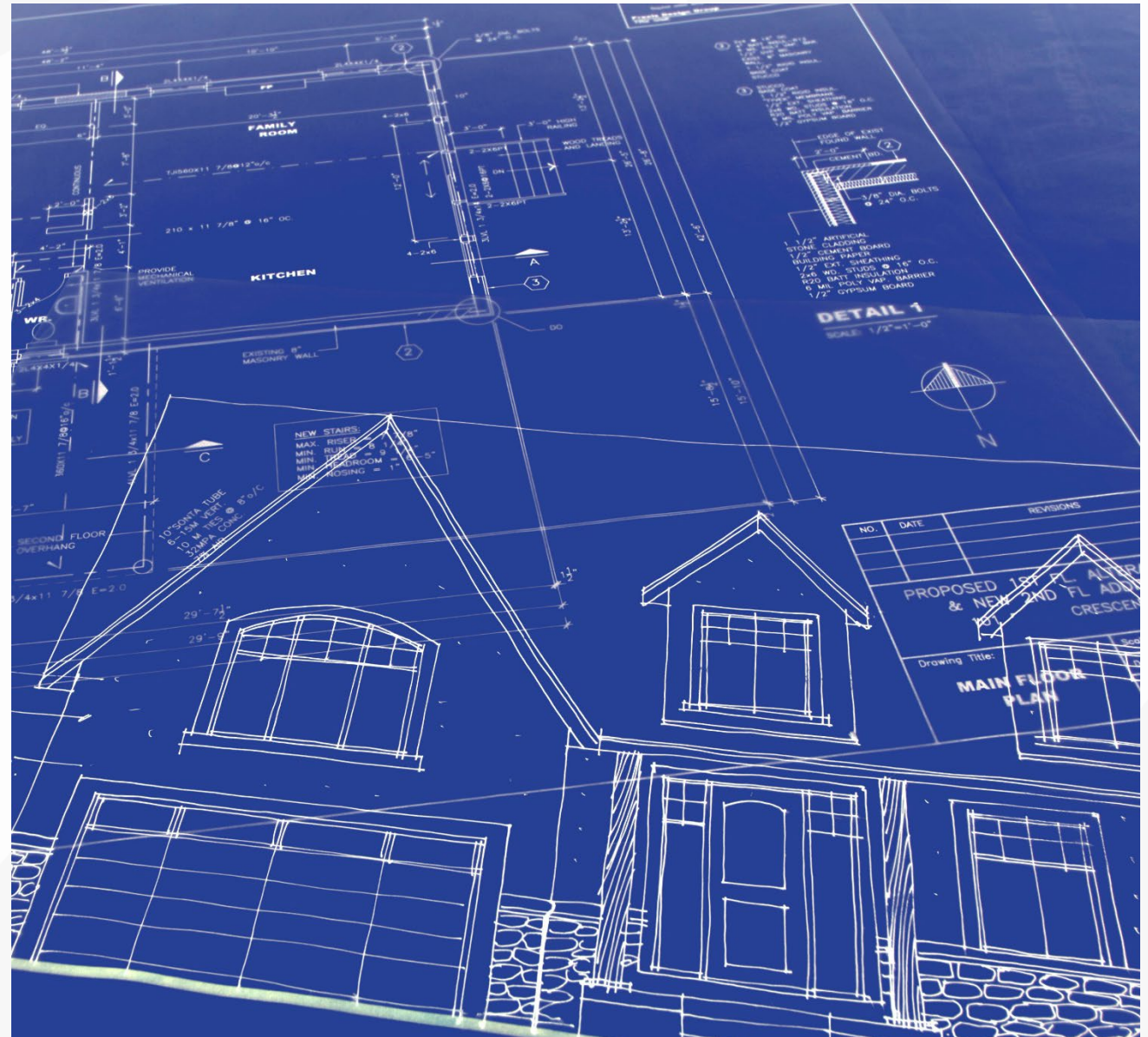




# Encourage self-reflection



**MAKE A PLAN TO  
ACHIEVE THE  
DESIRED  
PERFORMANCE**



**Consider this medical student...**







$\pm$

$\Delta$

# Culture of Feedback

- What is the culture of feedback at your institution? What works? What doesn't work?
- What can you do to influence the culture of feedback and evaluation at your institution?



# Summary

- Summative assessment is a judgment of competence at the end of a learning experience by an evaluator
- Formative assessment (feedback) is a cycle of performance, observation and transmission of information to improve performance
- Direct observations are the foundation of evaluation and good feedback (using the  $+/\Delta$ )
- Creating a strong culture of feedback can help to overcome barriers



**Thank You!**

