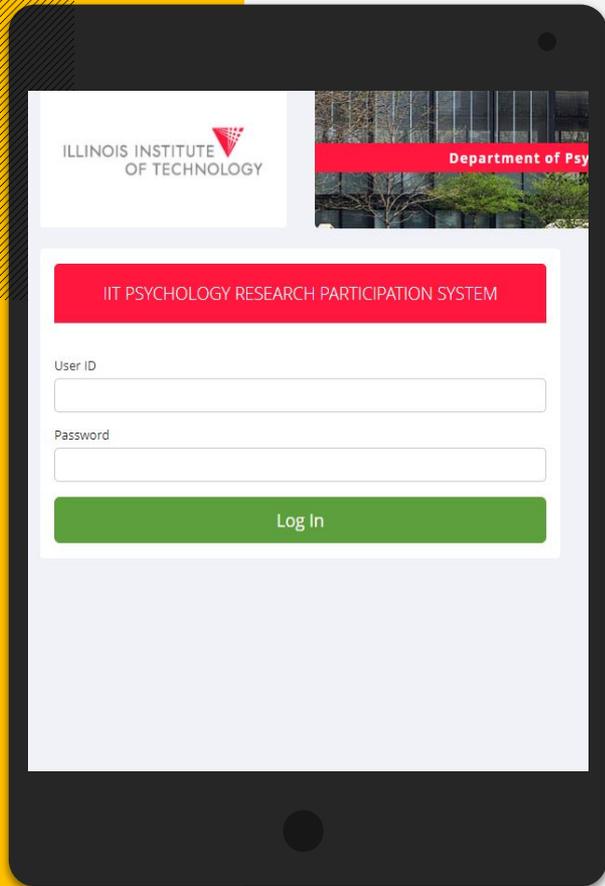




Performance Measurement

Psych 301
Instructor: Melanie Standish





Class Announcements

Extra Credit Opportunity:

<https://iit.sona-systems.com/>

Study: Classroom Experiences

Time: 15 minutes

Deadline: March 31, 2021 by 11:59pm CT



Table of contents.

01

Basic
concepts in
performance
measurement

02

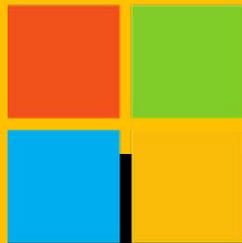
Performance
rating -
substance

03

Performance
rating -
process

04

The social
and legal
context of
performance



Microsoft



Other companies have used a forced ranking system.

Class poll!



Microsoft



Uber

YAHOO!



01: basic concepts in performance measurement

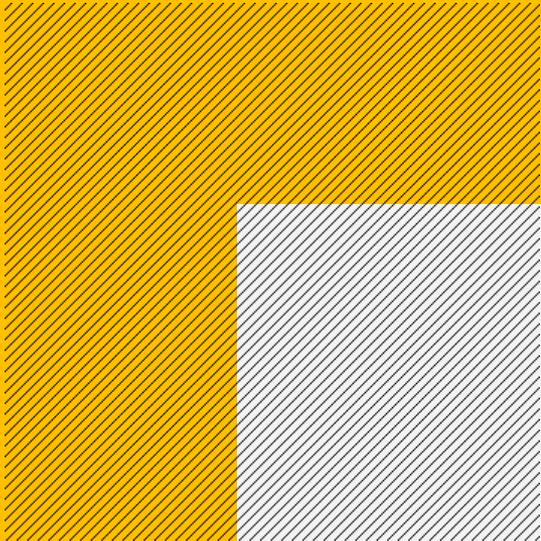


Let's think back to the Chapter 4 on Job Analysis & Performance

PERFORMANCE



“Actions or behaviors
relevant to the organization’s
goals; measured in terms of
each individual’s proficiency.”

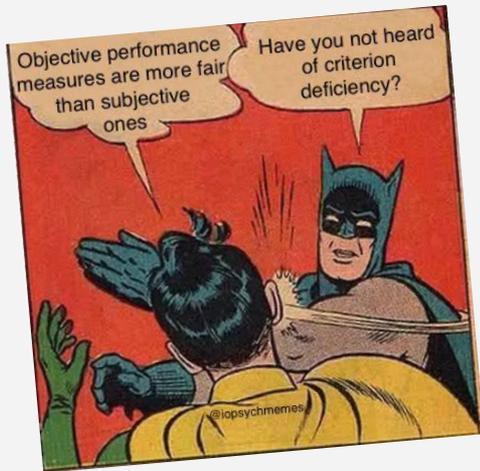


USES OF PERFORMANCE INFORMATION

- Criterion Data
- Employee development
- Motivation
- Rewards
- Transfer
- Promotion
- Termination



TYPES OF MEASURES & INTERRELATIONSHIPS



Recall from Chapter 4

- **Objective performance** measures
 - Sales data
 - Electronic performance monitoring
- **Subjective performance** measures
 - Supervisor ratings
- **Hands-on performance measure**
 - Could be classified as objective but eliminates some contaminating influences of objective measures (e.g., inadequate equipment, interrupting coworkers)



PERFORMANCE MANAGEMENT VS. PERFORMANCE APPRAISAL

- Emphasizes link between individual behavior & organizational strategies & goals
- Components
 - Definition of performance
 - Actual measurement process
 - Communication between supervisor & subordinate about individual behavior & organ. expectations



PERFORMANCE APPRAISAL VS. PERFORMANCE MANAGEMENT

PA:

- Rare (e.g., once a year)
- Initiated by HR
- Developed by HR and given to managers
- Feedback is rare (e.g., once a year)
- Nonstrategic
- Focus on evaluation

PM:

- More frequent
- Initiated by HR, manager, subordinate
- Developed by managers and subordinates
- Ongoing feedback
- Strategic
- Focus on understanding strategic performance criteria and behavior fit

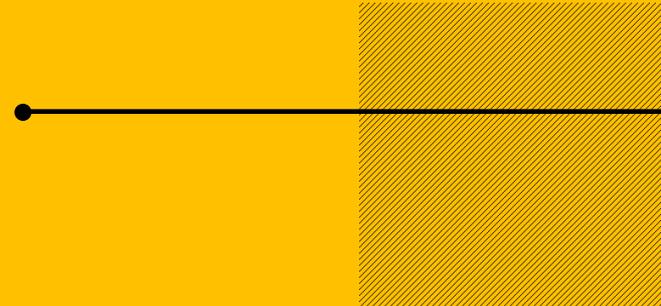


WHY WORRY ABOUT FORMAL APPRAISAL?

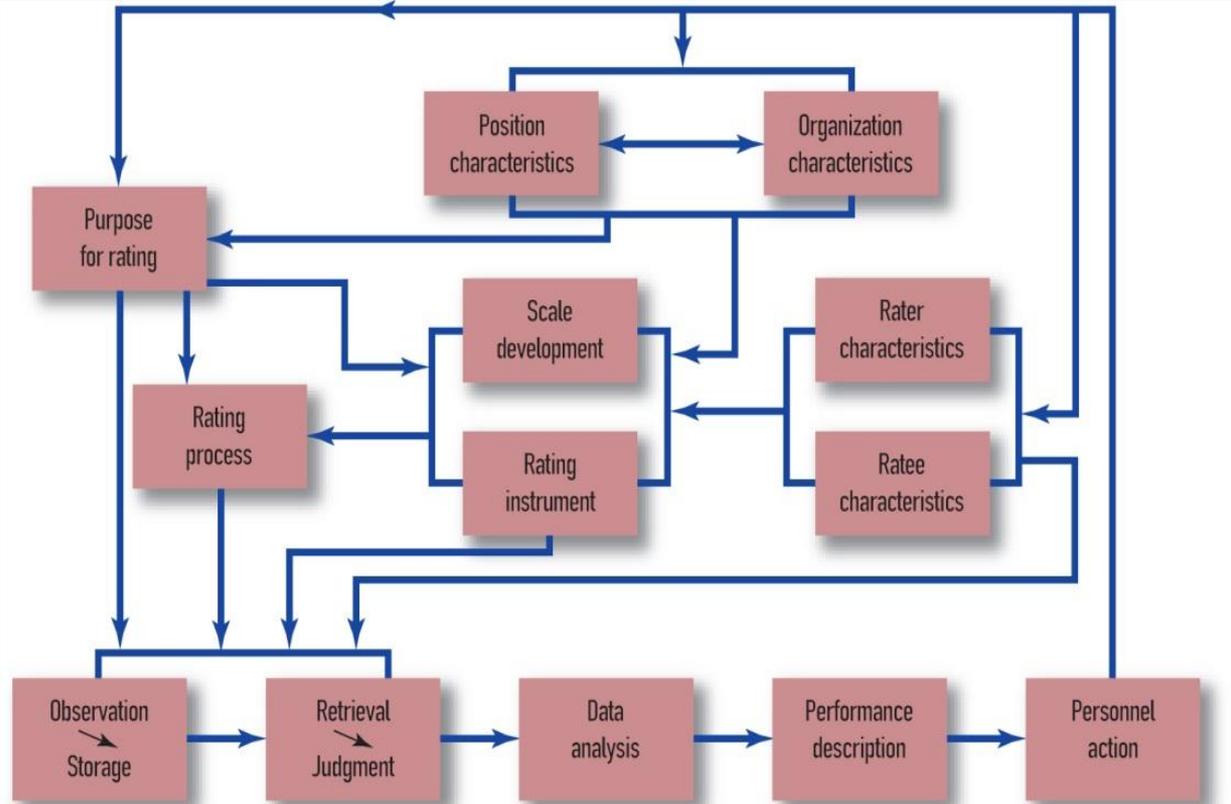
- Helps people recognize how to do their jobs better
- Identifies unknown selection and training needs
- Rewards good performance and identify failures that are “untrainable” or not cost effective
- Maintains fairness in salaries, benefits, hiring, firing
- Protects against charges of discrimination



02: performance rating substance



PROCESS MODEL OF PERFORMANCE RATING



SOURCE:
Landy & Farr (1980).





TYPES OF RATINGS

- Overall Performance
 - Influenced by
 - Task performance
 - OCBs
 - CWBs
 - Use driven by desire for simplicity, using for administrative purposes
 - Practically not useful – no conceptual meaning



TYPES OF RATINGS

- Trait Ratings = **Bad!**
 - Not actions or behaviors
 - Not defensible
- Task-Based Ratings & Critical Incidents = **Good!**
 - Typically supervisors rate tasks on scale anchored by critical incidents



EVALUATING RATING SCALES

- Behavioral definition
- Response category defined
 - e.g. what does a “4” or “satisfactory” mean?
- Response unambiguous
 - Degree to which a person interpreting or reviewing the ratings can understand what response the rater intended



RATING SCALES

- Graphic Rating Scales
 - Graphically display performance scores running from high to low
 - Most common

Graphic Rating Scale

Sloan Tool & Die, Inc.
Duluth, Minnesota

Employee Rating Scale

Employee: Javier Alva
Department: Accounting
Rated by: Irene Muharsky
Date: 3/17/01

Behavior:	Unsatisfactory	Questionable	Satisfactory	Outstanding
A. Quantity of work	1	2	3	4
B. Quality of work	1	2	3	4
C. Work initiative	1	2	3	4
D. Efficiency	1	2	3	4
E. Overall	1	2	3	4

Copyright © 2002 by South-Western

FIGURE 8-2
8-15



RATING SCALES

- Checklists & Weighted Checklists
 - List of behaviors presented to rater who places a check next to items that best (or least) describe the ratee
 - Weighted checklists assign value to each task
 - Forced-choice – requires rater to choose specified number of statements that apply to ratee
 - Good for easy generation of score; bad for feedback

TABLE 5.4

A Weighted Checklist for a College Instructor

<input type="checkbox"/>	The instructor created a classroom environment that encouraged questions and discussion (4.2).
<input type="checkbox"/>	The instructor presented material clearly (2.2).
<input type="checkbox"/>	Lectures were adequately organized (1.7).
<input type="checkbox"/>	The instructor was enthusiastic and friendly (2.7).
<input type="checkbox"/>	The instructor used examples from his/her experience or research (3.8).

Note: Effectiveness values range from 1.00 (minor contribution to effectiveness) to 5.00 (major contribution to effectiveness). The instructor's score is the average of the items checked.



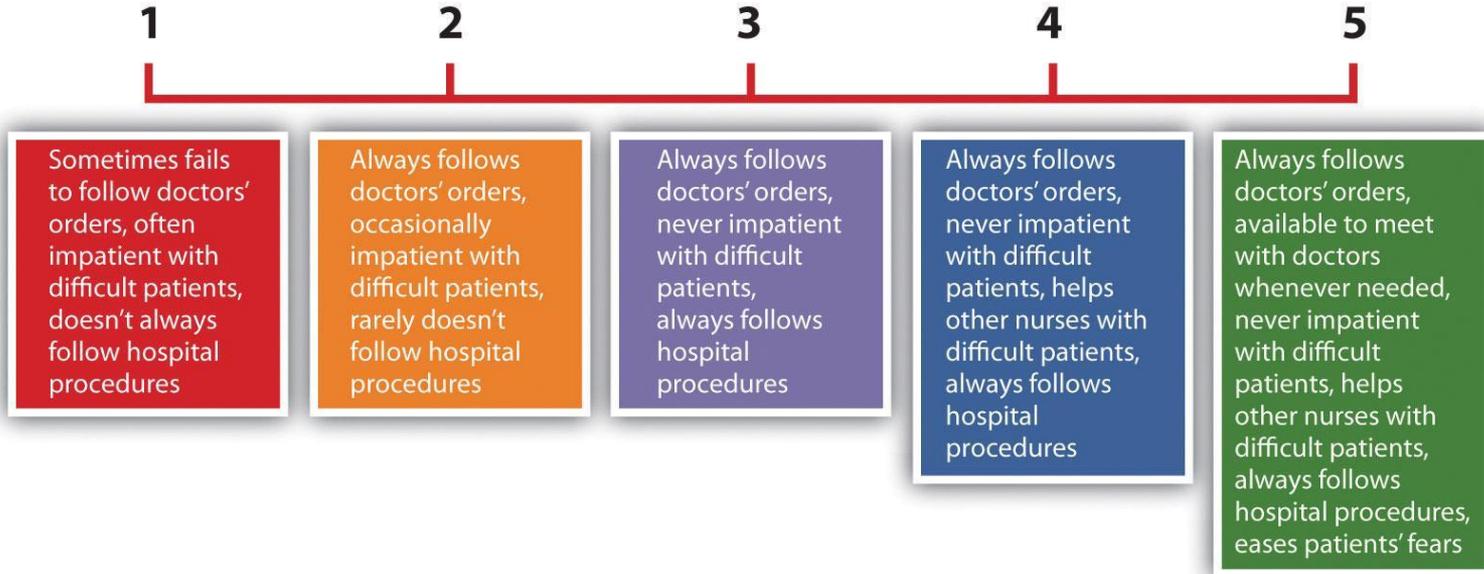
RATING SCALES

- Behaviorally Anchored Rating Scales (BARS)
 - Includes behavioral anchors describing what worker has done, or is expected to do, in a particular duty area
 - Time consuming to create
 - Increased perceptions of fairness & good for feedback

Behaviorally Anchored Rating Scale (BARS): Performance is assessed along a scale with clearly defined scale points containing examples of specific behaviors.

Example: A supervisor of a nurse indicated which scale point best describes the behavior of the nurse.

BARS for Army nurses





RATING SCALES

- Behavioral Observation Scales (BOS)
 - Response scale based on frequency (e.g., “almost never” to “almost always”)
 - Idea was that this would be more objective, and easier for the rater
 - Mixed findings regarding managers ability to recall frequencies.
 - Tziner and Kopelman (2002) – BOS was most preferred method by raters

Punch Press Operator:

1. Checks press for loose dies or other parts.

Almost never

Almost always

5

4

3

2

1

2. Cleans all machine parts with proper solvents.

Almost never

Almost always

5

4

3

2

1

3. Leaves machines when not using them.

Almost never

Almost always

5

4

3

2

1

4. Wears all safety equipment and clothing.

Almost never

Almost always

5

4

3

2

1

5. Feeds machine so as to prevent jamming or other malfunctioning.

Almost never

Almost always

5

4

3

2

1

Which scale is the best?

“The specific format of the final rating instrument does not generally have a major impact on the quality of the ratings that are obtained, given that the behavioral dimensions and the scale anchors are developed from a careful analysis of the job and its content” (Newman, Kinney & Farr, 2004, p 375)



However...

...the type of rating instrument influences both ratee and rater perceptions, such as perceived accuracy, satisfaction, confidence, fairness, and ease of use (e.g. Hedge & Teachout, 2000; Kingstrom & Bass, 1981)



COMPARISON METHODS

- Types of Comparison
 - Simple ranking
 - Employees ranked from top to bottom according to assessed proficiency
 - Paired comparison
 - Each employee in a group is compared with each other individual in the group
 - Not good for large groups – # of comparisons = $n(n-1)/2$
- Good for termination decisions; bad for feedback or performance management
- Difficult to compare across groups
 - Computer Adaptive Rating Scales





03: performance rating process





WHO CONDUCTS APPRAISAL?

- Supervisors (most common)
- Peers
- Self
- Subordinate
- Customers (if directly related to performance)
- Consultants
- Appropriate source depends on the purpose of the appraisal

**WHAT DO YOU THINK THE
CORRELATION BETWEEN
SUPERVISORS AND SELF IS?**

- 1.0
- 0.8
- 0.6
- 0.4
- 0.2
- 0.0



AGREEMENT BETWEEN SOURCES

Between...	<i>r</i>
Two supervisors	.50
Two peers	.37
Two subordinates	.30
Supervisors and peers	.34
Supervisors and subordinates	.22
Supervisor and self	.22
Peers and subordinates	.22
Peers and self	.19

360 DEGREE APPRAISAL

- Collect & provide an employee with feedback that comes from many sources
- Often used for employee development
- Issues
 - Sources of Variability
 - Event rather than a process
- Validity
 - Related to assessment center ratings
 - Related to supervisory performance evaluations



FEEDBACK

- Kluger & DeNisi's (1996) meta-analysis
 - Across 23,000 participants, the average improvement due to feedback is $d = 0.41$
 - One third of the time, feedback decreased performance
- Why?
 - Specificity
 - Accuracy
 - Timeliness
 - Negativity/Constructive Criticism

FEEDBACK

- Employee more likely to accept negative feedback if he/she believes:
 - Supervisor has sufficient “sample” of subordinate’s actual behavior
 - Supervisor & subordinate agree on subordinate’s job duties
 - Supervisor & subordinate agree on definition of good & poor performance
 - Supervisor focuses on ways to improve performance
 - “Praise-criticism-praise sandwich”

RATING DISTORTIONS (ERRORS?)

- **Halo** – rater assigns the same rating to a ratee on a series of dimensions
- **Leniency** – rater provides generally high scores across dimensions
- **Severity** – rater provides generally low scores across dimensions
- **Central Tendency** – raters marks the midpoint of a scale on a series of dimensions





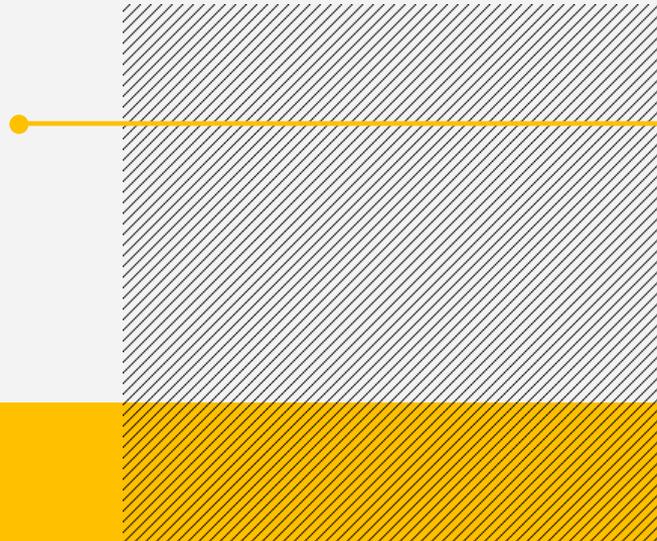
SOURCES OF RATER ERROR & THE SOCIAL CONTEXT

- Rater personality
- Rater mood
- Rater motivation
- Race/Gender/Similarity
- Managers' fear of reactions to negative feedback
- Opportunity to observe
- Goal of appraisal process
- Political factors?



MINIMIZING RATER ERROR

- Use well-defined behavioral anchors
- Train raters?





RATER ERROR (PSYCHOMETRIC) TRAINING

- Premise
 - Raters can avoid common rater errors if they are aware of them.
- Does reduce halo.
- Does not increase accuracy.
 - May actually harm accuracy



PERFORMANCE DIMENSION TRAINING

- Premise
 - Raters can better identify dimensions if they are trained on what behaviors they involve.
 - We make judgments as behaviors occur.
 - If we put judgments into categories as we observe the behavior, halo might be reduced.
- Does reduce halo
- Does not improve leniency/severity
- Does improve accuracy, slightly



FRAME OF REFERENCE TRAINING

Training based on the assumption that a rater needs a context or “frame” for providing a rating; includes (1) providing information on the multidimensional nature of performance, (2) ensuring that raters understand the meaning of anchors on the scale, (3) engaging in practice rating exercises, and (4) providing feedback on practice exercises.

- Premise
 - Raters can be more consistent if they calibrate their understanding of the scales.
- Does improve accuracy
- Does not improve halo



OBSERVATIONAL ACCURACY TRAINING

- Aim is to improve memory of job performance incidents.
 - Keep records of “key” events
 - Diary or PDA
- Does improve accuracy
- Does not reduce halo

ACCURACY VS. ERROR

- Does halo matter if the ratings are accurate?
- Is halo error?
- Are other distortions errors?



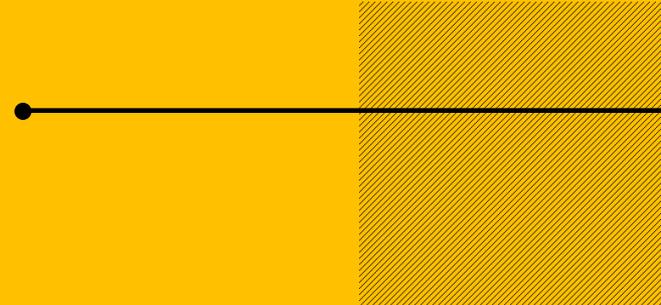


RELIABILITY & VALIDITY OF PERFORMANCE RATINGS

- Reliability
 - Currently the subject of lively debate
 - Inter-rater reliability considered poor but this isn't necessarily bad considering each rater relies on a different perspective
- Validity
 - Depends on manner by which rating scales were conceived & developed



04: social & legal context





MOTIVATION TO RATE

- PA/PM process as a means to an end, either personal or organizational
- Performance appraisal as a goal-directed activity with at least 3 stakeholders

Organizational Goals

- Between-person uses
- Within-person uses
- Systems-maintenance

Rater Goals

- Task performance
- Interpersonal
- Strategic
- Internalized

Ratee Goals

- Information gathering
- Information dissemination



GOAL CONFLICT

- When single system is used to satisfy multiple goals from different stakeholders, rater must choose which goal to satisfy before assigning a rating
- Possible solutions
 - Use multiple performance evaluation systems
 - Obtain involvement of stakeholders in developing the system
 - Reward supervisors for accurate ratings



PERFORMANCE EVALUATION & THE LAW

- Ford Motor Company & forced distribution rating system
 - Evaluators were required to place managers into performance categories based on predetermined percentages (10%, 80%, 10%)
 - Ford paid \$10.5 million to litigants
- Werner & Bolino's (1997) review of court cases
 - Analyzed 295 cases from 1980-1995
 - Judges primarily concerned with issues of fairness rather than technical characteristics of the system

BEST PRACTICES

- Standardized and uniform – based on JA
- Formally communicated
- Provide prompt notice of performance deficiencies
 - Opportunities to correct
- Employees should have access to their reviews
- Provide methods to contest
- Use multiple, diverse, unbiased raters who receive written instructions
- Require thorough, consistent documentation
- System to detect discriminatory effects
- Performance management systems

Thank you!

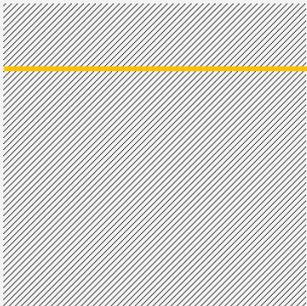
Questions? mstandis@hawk.iit.edu



Presentation Template: SlidesMania

Images: Unsplash

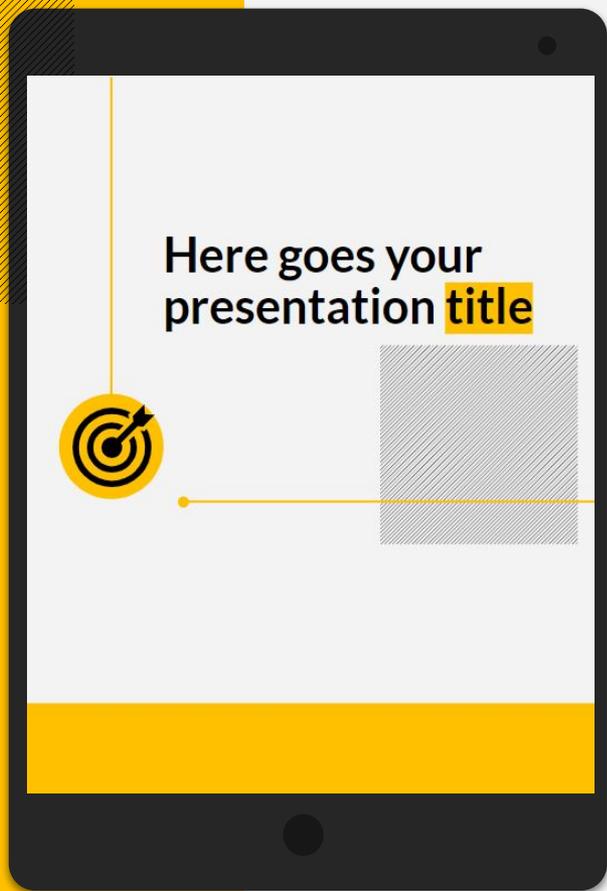
Please keep this slide or mention us and the other resources used in the footer of a slide.



Finish up your first set of slides.

Great idea!

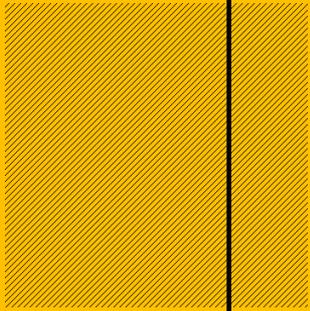




Add your **title** here.

If you are presenting a website, an internet product or an app, you can use this tablet mockup to showcase it.

Just replace the screenshot with your own.





Free themes and templates for
Google Slides or **PowerPoint**

NOT to be sold as is or modified!

Read [FAQ](#) on slidesmania.com

Sharing is caring!

