



LEADING the CONVERSATION

A Q&A Session with Experts to Overcome Test Development Challenges



Liberty Munson, PhD
Principal Psychometrician
& Assessment Quality Lead
Microsoft Learning Experiences
lmunson@Microsoft.com



Ada Woo, PhD
Director of Measurement and Testing
National Council of State Boards of Nursing (NCSBN)
awoo@ncsbn.org



Laura Balk
Senior Manager, Exam Development
CompTIA
LBalk@comptia.org



Manny Straehle, Ph.D., GISF
Founder and President
Assessment, Education, and Research Experts(AERE)
manny@aerexperts.com

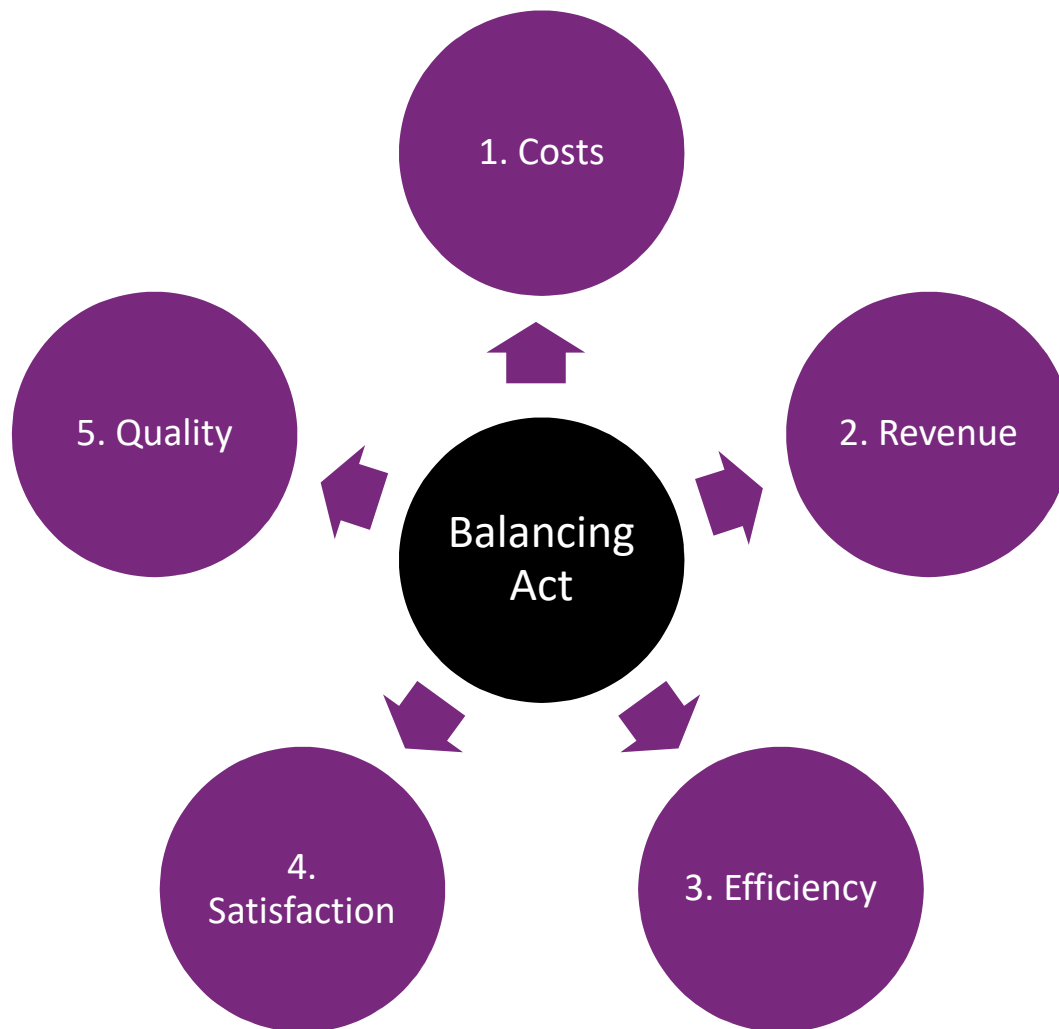
Agenda

- Ideal Product
- Key Product Considerations
- Test Development Cycle
- Q&A

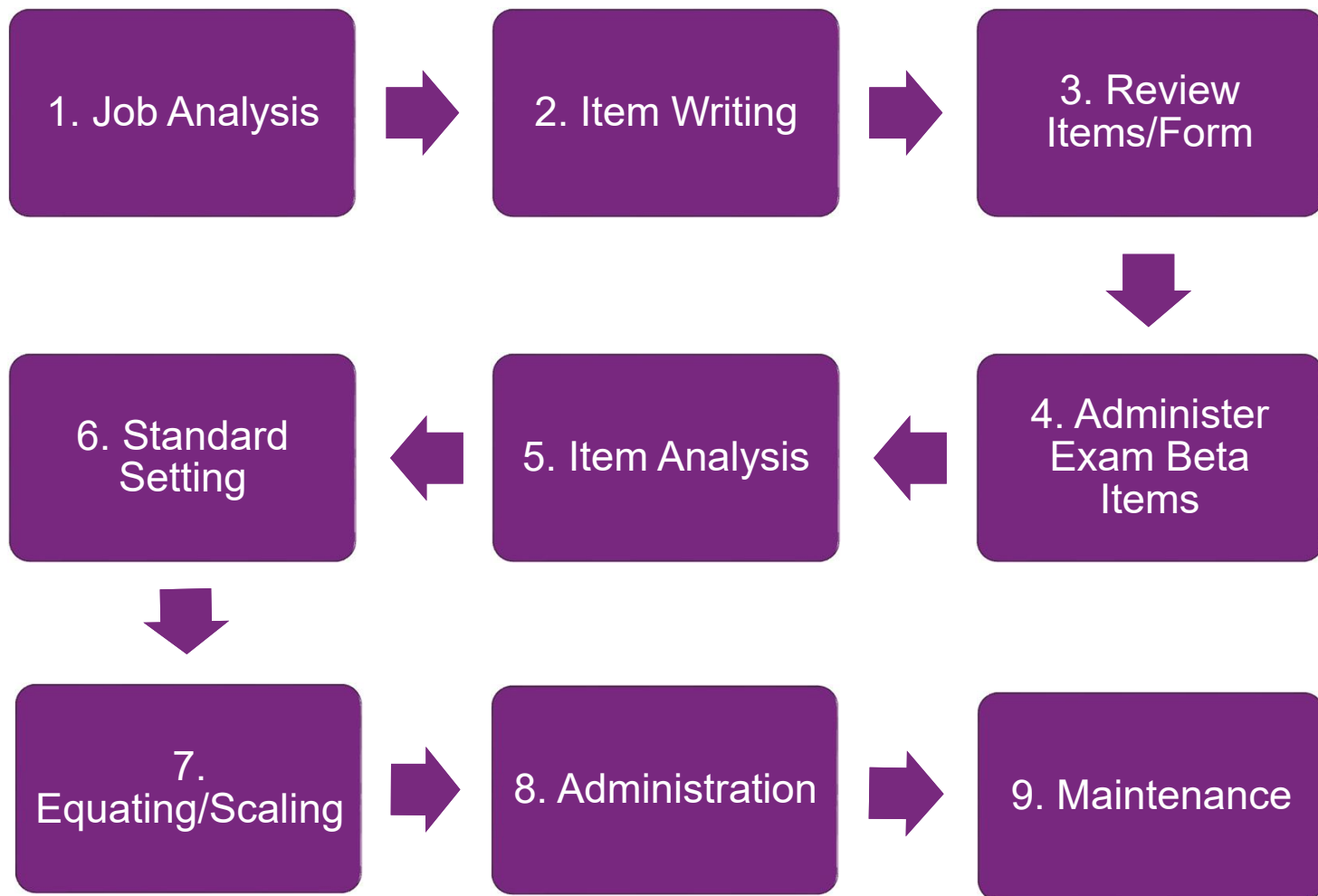
The Ideal Product



An Assessment is a Product



Test Development Lifecycle



General Questions



Q. I don't think I have enough SMEs to complete this [insert exam development step here]. What should I do?

Q. Candidates are complaining that the training materials don't align to exam content. How do I respond?

Q. Can I use the same group of SMEs for all my test development efforts (e.g., those who participate in item writing can be used for standard setting?)

Q. How many SMEs do I need on the job analysis panel?

Q. Can I do a JTA without a survey? I don't think I will have enough subject matter experts.

Q. There are fewer than 10 SMEs in the world. How do I conduct a JTA?

JTA: Process



Q. My JTA survey is too long (more than the 7 to 8 minutes that marketing survey professionals recommend). What can I do?

Q. What ratings should I use? Is importance the only one?

Q. Should I include demographic questions on my JTA survey? If so, which ones?

Q. How should I analyze the results? What approach should I take to weighting (mathematical or SME derived or both)?

Q. What about a factor analysis?

Item Development



Q. We planned to have 100 items for this exam, but we weren't able to develop that many. We will be seeking accreditation (NCCA/ANSI). What can I do?

Q. Should I have agreed upon resources/reference/body of knowledge identified before item writing?

Q. How should I select item writers?

Q. How many items can someone write in one day?

Q. What is the general rejection rate of items as they proceed through the technical review and field testing? How can I ensure that I X number of items at the end of the content development process?

Q. What documentation is needed for the item writing session?

Q. I had a few candidates who did not take the entire exam. Can I still use their results?

Q. The exam has been in beta for weeks, and candidates are starting to complain about not getting their result. I don't have enough responses to feel confident in the psychometrics. What should I do?

Item Analysis

Q. My testing program has very low annual test volume. Should I still field-test new items before using them operationally?

- What can I do to get a scored exam in market more quickly?

Q. Can I use Angoff or SME opinions if I don't have any item difficulty items? What other methods could I use(e.g., Thorndike method)?

Standard Setting



Q. I don't want my SMEs to spend 7 hours in a modified Angoff session. Is there a quicker way?

- My exam is very low stakes.
- My exam is moderate stakes.
- My exam is high stakes.

Q. SMEs don't seem to understand that the cut score is not the same as a grade. What should I do?

Q. I think my SMEs are intentionally distorting their responses, trying to inflate/deflate the cut score. What should I do?

Standard Setting



Q. I want to conduct a virtual standard setting meeting. How will that be different than running an in-person standard setting meeting?

Q. My organization is concerned that the exam cut point is incorrect. What can be done to ease concerns? Can the cut point be changed?

Q. My paper and pencil test forms were lost in shipment. Are the items considered compromised? What can I do?

Q. What should exam security include?

Q. Is a compromised item really compromised? If so, what should I do if I don't have the budget or resources to create new items?



Q. I want to transition my test from paper and pencil to computer-based. What factors do I need to consider?

Q. I want to incorporate performance-based item types. What factors do I need to consider? How do I handle ADA requests? What accessibility concerns should I have?

Q. I want to create a time-sensitive exam. How would I address any ADA requests?

Q. How do I make sure that all of my test forms are parallel? Do I need to do that?

Q. I have a global exam where localization volumes are low. What can I do to ensure equivalence across all language forms?

Q. Can I conduct a standard setting for each form instead of equating?

Q. How many common items do I need for linear equating?

Q. The content domain that is the basis of my exam is rapidly changing.

- How do I keep my exam content up to date?
- Major sections of my exam are no longer relevant. What should I do?

Q. I just completed the psychometric analysis of my exam, and all of the items in one of the content areas are performing poorly. What should I do?

Q. How often should I perform item and exam health checks? What should I include in these reviews?

Score Reporting



Q. What do I need to include in the score report?



LEADING the CONVERSATION