2016 ICE EXCHANGE

ICE — Institute for Credentialing Excellence™

NOVEMBER 8–11

The Broadmoor | Colorado Springs

Managing Change in Certification: Ensuring Relevance in Skills Assessment

Speakers

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

Manny Straehle, Ph.D., GISF

Founder and President

Assessment, Education, and Research Experts(AERE)

manny@aerexperts.com

2016 ICE EXCHANGE

NOVEMBER 7–11
The Broadmoor | Colorado Springs

Managing Change in Exam Development and Maintenance: Microsoft's Journey

Liberty J. Munson, PhD, Principal Psychometrician
Microsoft Learning Experiences

Learning for the Modern Era
Enabling the next generation of innovators, entrepreneurs and developers

2016 ICE EXCHANGE

1 Build Content with Strategic Partners

edX | MIT | Massachusetts Institute of Technology

THE LINUX FOUNDATION | Harvard

Microsoft Partner
Learning

2 Extend our Platform through Open Source

OPENedX

3 Go where the Learners are

edX | Microsoft Partner Learning | Hour of CODE | YouTube | pluralsight | UNIVERSITY OF MARYLAND

MVA | Channel 9 | LinkedIn | Microsoft Imagine

Creative Coding through Games and Apps	Machine Learning	Learning Credits
Harvard/MSFT CS50.AP	Microsoft Azure Fundamentals	Badges
C# for Absolute Beginners	Programming in C#	Microsoft Professional Degree
Hour of Code	MOC On Demand	
Minecraft		
Spark **Excitement**	Drive Deep **Engagement**	Enable Capability **Mastery**
Affinity	Adoption	Advocacy

4 Priorities

Microsoft Azure | Office 365

Windows

5 Goals and Outcomes

⬆ Total reach

⬆ New/Competitive Developers reached

⬆ Skilled technologists building on Microsoft platforms and services

⬆ Preference for Microsoft platforms & services

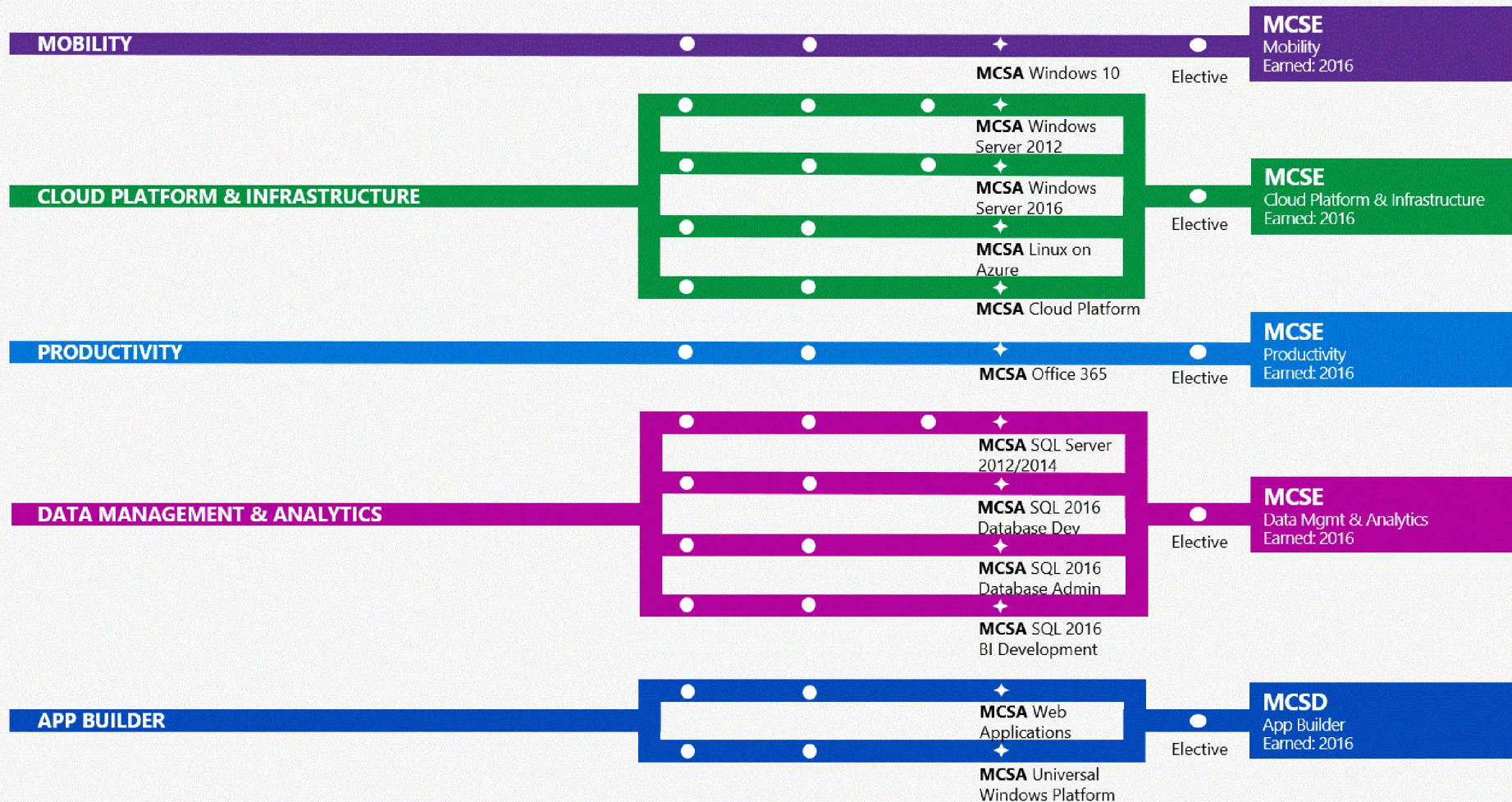| Empower Students & Educators	Activate Cloud & Productivity Partners	Enable Customers & Partners at Scale	Drive Usage, Deployment & Consumption

An Overview of MS Certification Exams

- Computer administered

- Test center deliveries or online proctoring

- Global distribution

- Ongoing delivery

- Variety of item types (e.g., multiple choice, drag and drop, active screen, hot area, case studies, labs, text entry, code analysis, etc.)

Microsoft Learning Experiences

2016 ICE EXCHANGE

Drive adoption of Microsoft technologies through training and certification

MOBILITY
MCSA Windows 10 — Elective
MCSE Mobility — Earned: 2016

CLOUD PLATFORM & INFRASTRUCTURE
- MCSA Windows Server 2012
- MCSA Windows Server 2016
- MCSA Linux on Azure
- MCSA Cloud Platform
- Elective
MCSE Cloud Platform & Infrastructure — Earned: 2016

PRODUCTIVITY
MCSA Office 365 — Elective
MCSE Productivity — Earned: 2016

DATA MANAGEMENT & ANALYTICS
- MCSA SQL Server 2012/2014
- MCSA SQL 2016 Database Dev
- MCSA SQL 2016 Database Admin
- MCSA SQL 2016 BI Development
- Elective
MCSE Data Mgmt & Analytics — Earned: 2016

APP BUILDER
- MCSA Web Applications
- MCSA Universal Windows Platform
- Elective
MCSD App Builder — Earned: 2016

Microsoft Announces NEW Certification Structure

- The five new expert certifications are:
  - » **MCSE: Cloud Platform and Infrastructure** - focusing on skills validation for Windows Server and Microsoft Azure
  - » **MCSE: Mobility** - focusing on skills validation for Windows Client and Enterprise Mobility Suite
  - » **MCSE: Data Management and Analysis** - focusing on skills validation for both on-premises and cloud-based Microsoft data products and services
  - » **MCSE: Productivity** - focusing on skills validation for Office 365, SharePoint, Exchange, and Skype for Business
  - » **MCSD: App Builder** - focusing on skills validation for Web and Mobile app development

- To earn each of these credentials:
  - » Earn Microsoft Certified Solutions Associate (MCSA) certification and
  - » Pass a single additional exam from a list of electives associated with the corresponding Center of Excellence

- Certifications will include achievement date that signifies candidates investment in staying up to date on the technology
  - » Every year, you will have the opportunity to re-earn the certification by passing an additional exam from the list of electives

Development Challenges

Goal:
- Ensure competence in use of Microsoft products
- Drive adoption and sales
- Certify as many people as possible as soon as our technologies become available

Implications:
- Exam in market around the same time as general availability of the software
- Development begins on "vaporware"

Vaporware
- Features and functionality of early versions are often in flux
- Features may not be available but the product group plans to include them in the final release, and we may need to develop items for them

Exam Development Process

Exam Development and Item Writing Strategies

- Mitigation efforts include:
  - » Test blueprint focuses on KSAs less likely to become obsolete

- Given that technology is constantly changing, we would like to better understand how likely each of the tasks included on this exam are likely to change over the next two years. Use the following rating scale to indicate **how likely the task is likely to change over time**.
  - » Likely to change in 3 months or less
  - » Likely to change in 3 months to 6 months
  - » Likely to change in 6 months to 1 year
  - » Likely to change in 1 to 2 years
  - » Unlikely to change in the next 2 years

Exam Development and Item Writing Strategies

- Mitigation efforts include:
  » Test blueprint focuses on KSAs less likely to become obsolete
- SMEs rate objectives on importance for success, frequency, and potential for obsolescence
  » Item writing focuses on fundamental concepts rather than specifications
    – Fundamental concepts are more stable
    – Function, purpose, application/use, intent
    – Minimize application or version specific content
    – Must resist being on the cutting edge in this context

Designing & Developing Exams on Vaporware

- Close partnership with business group
- Identifying "right" SMEs
  - » Have early access to software
- Changes communicated as early as possible
  - » "Feature complete" version shared with SMEs who "scrub" content domain and items to ensure relevance and technical accuracy
- Beta candidate comments
  - » Comments about technical accuracy, clarity, relevance, etc. reviewed with SMEs
  - » Occasionally comment on the long term viability of items that may become obsolete over time if rumored changes are implemented
- One final review
  - » Items are scrubbed before publication to ensure accuracy and relevance
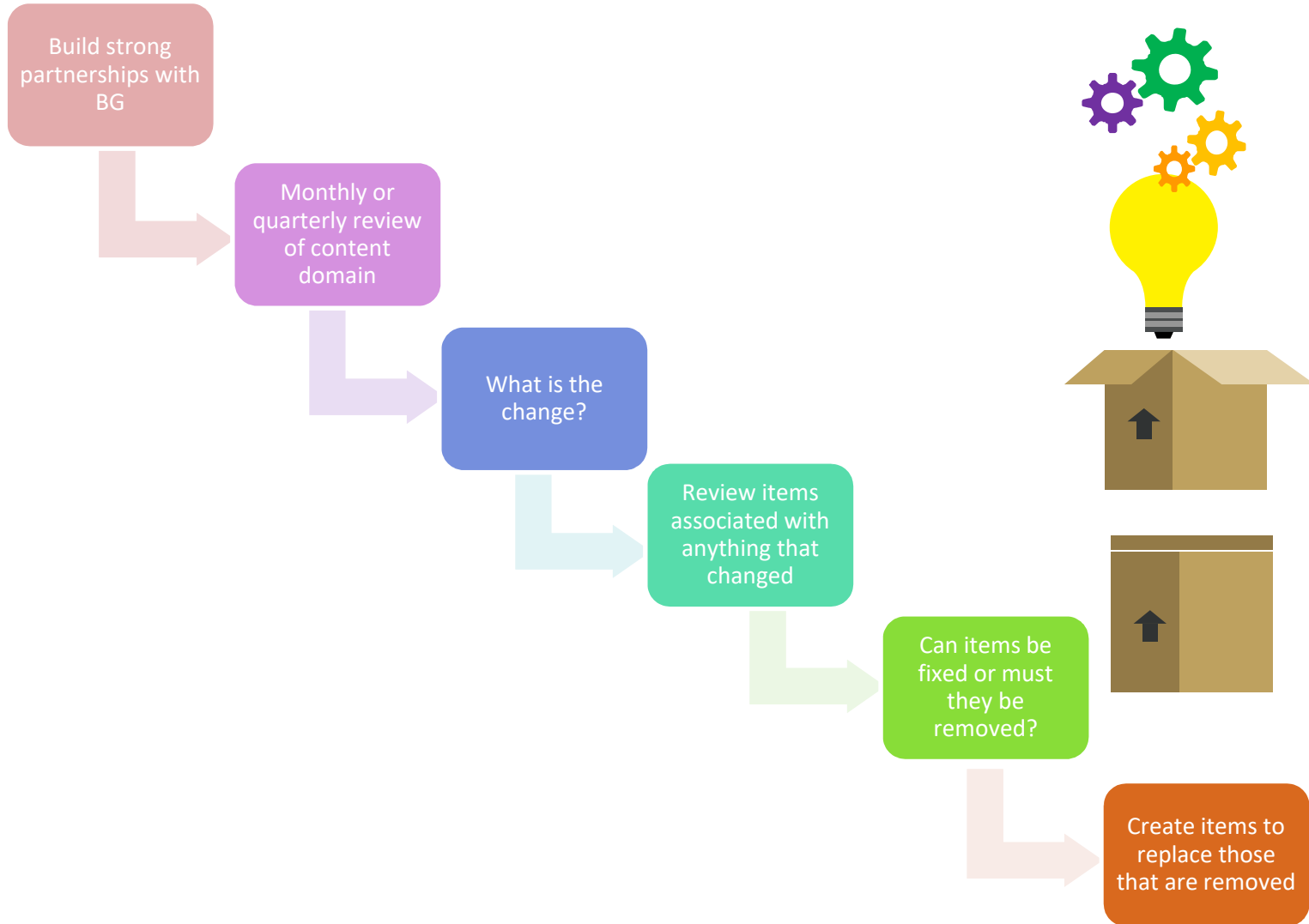
Maintenance Challenges

Goal:

- As Microsoft migrates more software to the cloud and leverages its flexibility, move away from major version releases of our software
  - Azure, Office 365
- A version-less future

Implications:

- Changes and improvements will be "evolutions" (small but constant change) rather than "revolutions" (big changes that are usually the hallmarks of the next version)
- Changes happen quickly and seamlessly
- Can be minor (e.g., feature names) to more substantial (e.g., user interface, functionality)
- Any change could effect exam content, skills assessed, and the validity of our certification process

Keeping up with Pace of Change

Build strong partnerships with BG

Monthly or quarterly review of content domain

What is the change?

Review items associated with anything that changed

Can items be fixed or must they be removed?

Create items to replace those that are removed

Content Domain Reviews

Impact	Description
No impact	Objective not impacted
Minor	Objective minimally impacted (e.g., name changes) · *Items require simple wording changes*
Moderate	Objective somewhat impacted; new functionality or tools have been introduced or enhanced, but fundamental skills assessed were not affected · *Some items may require easily made edits and revisions by SMEs*
Major	Objective is significantly impacted by new release; new features, functions, enhancements have changed the fundamental skills assessed; features, functions have been deprecated · *Items require significant changes to ensure technical accuracy; many items must be removed and replaced because the feature has been deprecated*
Obsolete	Objective no longer applies · *Items need to be removed from exam or moved to other objectives if possible*

Best Practices and Recommendations

- Build strong partnerships with those who have early insight into changes

- Plan regular reviews and updates

- Ongoing content development

  » Retainer model

  » Never stop updating even if those updates are "out of date"

- Listen to test-takers

- Are exams the best way to measure these skills? What might be a better approach?
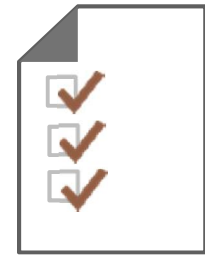
Exploring the Assessment of Clinical Decision-Making Skills: Nursing Licensure

Ada Woo, PhD, Director of Measurement and Testing
National Council of State Boards of Nursing (NCSBN)

NCLEX-RN and NCLEX-PN

- The NCLEX-RN and NCLEX-PN are the initial licensure test for nurses in all U.S. states and territories.
- Since 2015, the NCLEX-RN is also used as entry-to-practice test for ten Canadian provinces and territories.
- The NCLEX is a variable length CAT exam, ranging from 60 to 250 questions.
- Currently, six item types are included in the NCLEX.
  - » Multiple-choice
  - » Multiple response
  - » Ordered response
  - » Chart and exhibit
  - » Fill-in-the-blank calculation
  - » Hot spot
- Approximately 300,000 NCLEX are administered each year.

Triennial Practice Analysis: Methodology

- Practice Analysis Panel-Subject Matter Experts (SMEs)
  - » Nurses representing all NCLEX jurisdictions and various nursing specialties
  - » Panel members work with, supervise or are newly licensed nurses
  - » Provides feedback regarding entry-level nursing practice
- Activity Statements
  - » Creates a category structure reflective of the types of activities performed by newly licensed nurses
  - » Develops a comprehensive list of activity statements performed within each category to appear on the survey

Practice Analysis Survey

- Survey Sample
  - » 12,000 entry-level nurses
  - » Web and Paper surveys
- Responders rate each activity statement on its importance and frequency.
- Raters were asked to consider each activity as it relates to the practice of safe entry-level nursing and decreasing client complications.

Importance of Nursing Clinical Judgment

- Nurses made up the largest proportion of healthcare workers in the U.S.

- Total number of Registered Nurse and Practice Nurse licenses in the U.S. and its territories
  - » 5,483,181

- Many adverse events in the hospitals may be prevented if good clinical judgment has been exercised.

- Aging population and increased patient acuity

- Newly licensed nurses are required to make increasingly complex decisions about patients.

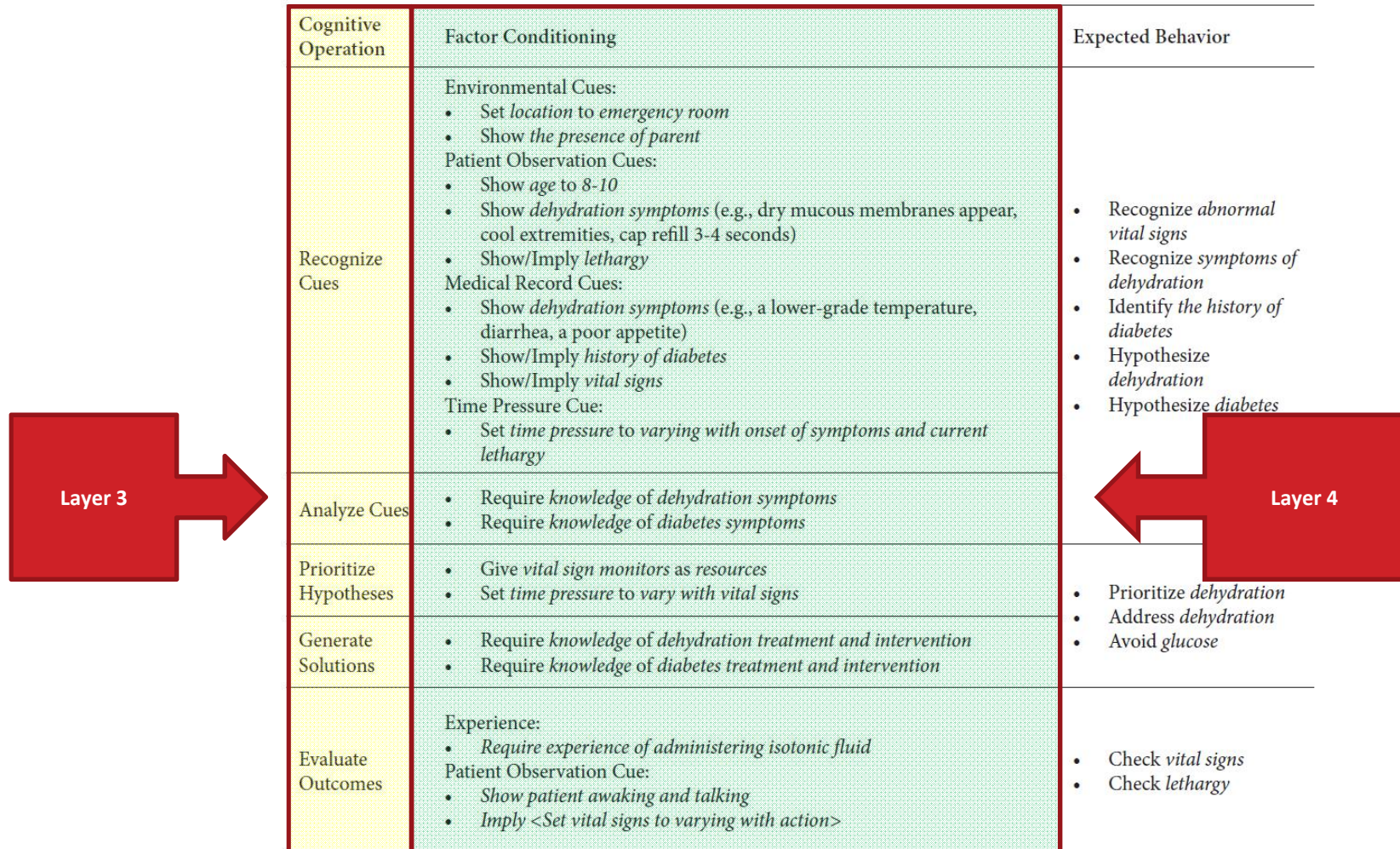# Operational Definition of Nursing Clinical Judgment

- Nursing clinical judgment is an iterative decision making process that uses nursing knowledge to observe and assess presenting situations, identify a prioritized client concern, and generate the best possible evidence-based solutions in order to deliver safe client care.

Assessment Model
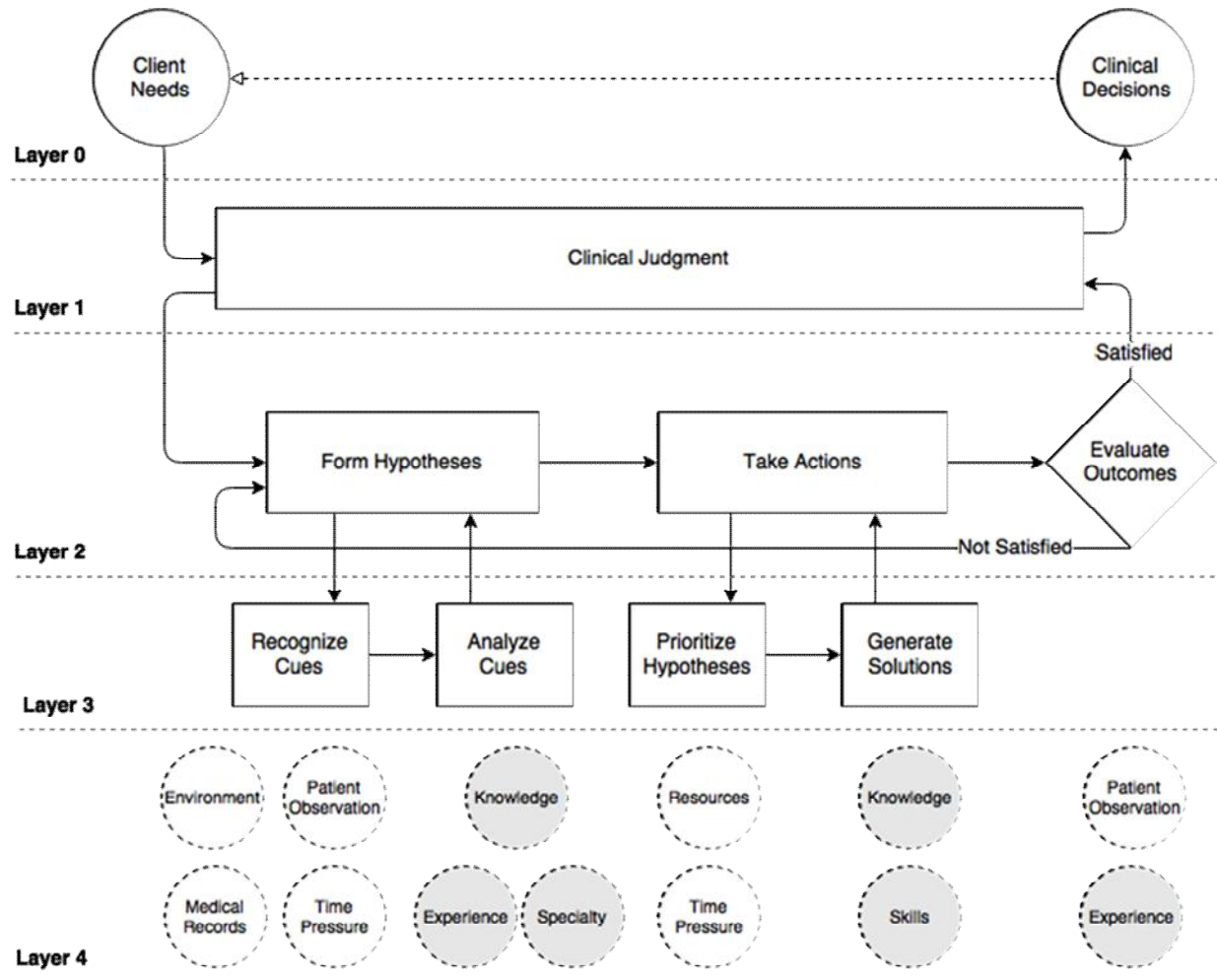
Dickison, P., Luo, X., Kim, D., Woo, A., Muntean, W., & Bergstrom, B. (2016). Assessing higher-order cognitive constructs by using an information-processing framework. *Journal of Applied Testing Technology*, *17*(1), 1-19. Retrieved from http://www.jattjournal.com/index.php/atp/article/view/89187/67797.

Task Model

Cognitive Operation	Factor Conditioning	Expected Behavior
Recognize Cues	**Environmental Cues:** • Set *location* to *emergency room* • Show *the presence of parent* **Patient Observation Cues:** • Show *age* to *8-10* • Show *dehydration symptoms* (e.g., dry mucous membranes appear, cool extremities, cap refill 3-4 seconds) • Show/Imply *lethargy* **Medical Record Cues:** • Show *dehydration symptoms* (e.g., a lower-grade temperature, diarrhea, a poor appetite) • Show/Imply *history of diabetes* • Show/Imply *vital signs* **Time Pressure Cue:** • Set *time pressure* to *varying with onset of symptoms and current lethargy*	• Recognize *abnormal vital signs* • Recognize *symptoms of dehydration* • Identify *the history of diabetes* • Hypothesize *dehydration* • Hypothesize *diabetes*
Analyze Cues	• Require *knowledge* of *dehydration symptoms* • Require *knowledge* of *diabetes symptoms*	
Prioritize Hypotheses	• Give *vital sign monitors* as *resources* • Set *time pressure* to *vary with vital signs*	• Prioritize *dehydration* • Address *dehydration* • Avoid *glucose*
Generate Solutions	• Require *knowledge* of *dehydration treatment and intervention* • Require *knowledge* of *diabetes treatment and intervention*	
Evaluate Outcomes	**Experience:** • *Require experience of administering isotonic fluid* **Patient Observation Cue:** • *Show patient awaking and talking* • *Imply <Set vital signs to varying with action>*	• Check *vital signs* • Check *lethargy*

Layer 3

Layer 4

Hypothetical Task Model: Pediatrics

Cognitive Operation	Factor Conditioning	Expected Behavior
Recognize Cues	Environmental Cues: ▪ Set *location* to *emergency room* ▪ Show *the presence of parent* Patient Observation Cues: ▪ Show *age* to *8-10* ▪ Show *dehydration symptoms* (e.g., dry mucous membranes appear, cool extremities, cap refill 3-4 seconds) ▪ Show/Imply *lethargy* Medical Record Cues: ▪ Show *dehydration symptoms* (e.g., a lower-grade temperature, diarrhea, a poor appetite) ▪ Show/Imply *history of diabetes* ▪ Show/Imply *vital signs* Time Pressure Cue: ▪ Set *time pressure* to *varying with onset of symptoms and current lethargy*	▪ Recognize *abnormal vital signs* ▪ Recognize *symptoms of dehydration* ▪ Identify *the history of diabetes* ▪ Hypothesize *dehydration* ▪ Hypothesize *diabetes*
Analyze Cues	▪ Require *knowledge* of *dehydration symptoms* ▪ Require *knowledge* of *diabetes symptoms*	
Prioritize Hypotheses	▪ Give *vital sign monitors* as *resources* ▪ Set *time pressure* to *vary with vital signs*	▪ Prioritize *dehydration* ▪ Address *dehydration* ▪ Avoid *glucose*
Generate Solutions	▪ Require *knowledge* of *dehydration treatment and intervention* ▪ Require *knowledge* of *diabetes treatment and intervention*	
Evaluate Outcomes	Experience: ▪ *Require experience of administering isotonic fluid* Patient Observation Cue: ▪ *Show patient awaking and talking* ▪ *Imply <Set vital signs to varying with action>*	▪ Check *vital signs* ▪ Check *lethargy*

Sample Item

An 8-year-old client with a history of diabetes presents to the emergency room with his mother, who reports that the child has not been feeling well for the last two days. She states he has a low-grade temperature, diarrhea, and a poor appetite. Today, the child reports he is feeling dizzy and that his head hurts. The mother also reports that he is refusing to eat or drink anything. Client vital signs upon arrival are pulse–162 beats/minute, respirations–26 breaths/minute, blood pressure–78/42 mmHg, temperature-100.3° F orally and blood serum glucose-75mg/dL. The client is admitted to the hospital, and an intravenous line is placed with 0.9% normal saline infusing at 50mL/hr. The nurse notes that the child is responsive to questions but appears lethargic. The mucous membranes appear dry, extremities are cool, and capillary refill is 3-4 seconds.

1. Which of the following orders can the nurse anticipate?
   a. Administer an intravenous fluid bolus of isotonic fluid (Key).
   b. Offer a cola beverage.
   c. Administer acetaminophen.
   d. Administer oxygen via nasal cannula.

The nurse re-evaluates the client after two hours from the initial admission. The child is awake and talking, extremities remain cool, and capillary refill is 2-3 seconds. The client is asking to drink something. Client vital signs are pulse–152 beats/minute, respirations–22 breaths/minute, blood pressure–82/46 mmHg, temperature-100.2° F orally. Laboratory values: electrolytes, within normal limits; blood serum glucose, 80mg/dL.

2. Which of the following actions should the nurse take?
   a. Administer an intravenous fluid bolus of isotonic fluid (Key).
   b. Administer insulin.
   c. Increase the 0.9% normal saline intravenous fluid rate.
   d. Discontinue the intravenous line.

3. The nurse re-evaluates the client after four hours from the initial admission. Which of the following findings indicate that the client's treatment has been effective?
   a. blood glucose of 85mg/dL
   b. pulse of 100 beats/minute (Key)
   c. respiration rate of 20 breaths/minute
   d. oral temperature of 100° F

Conclusions

- The framework is theory-based and can be empirically validated.

- The framework is flexible enough to adapt to evolving theory/requirements.

- There is a clear separation between cognitive processes (Layer 3) and cognitive attributes (Layer 4).

- Continuous validation processes are incorporated in the framework.

- The framework can be applied to measuring any high-order cognitive constructs

The Journey

1. Reality

6. Agile Approach Across Credential Lifecycle

2. Sources of Change

Agile

5. Strategic Considerations

3. Overall Solution
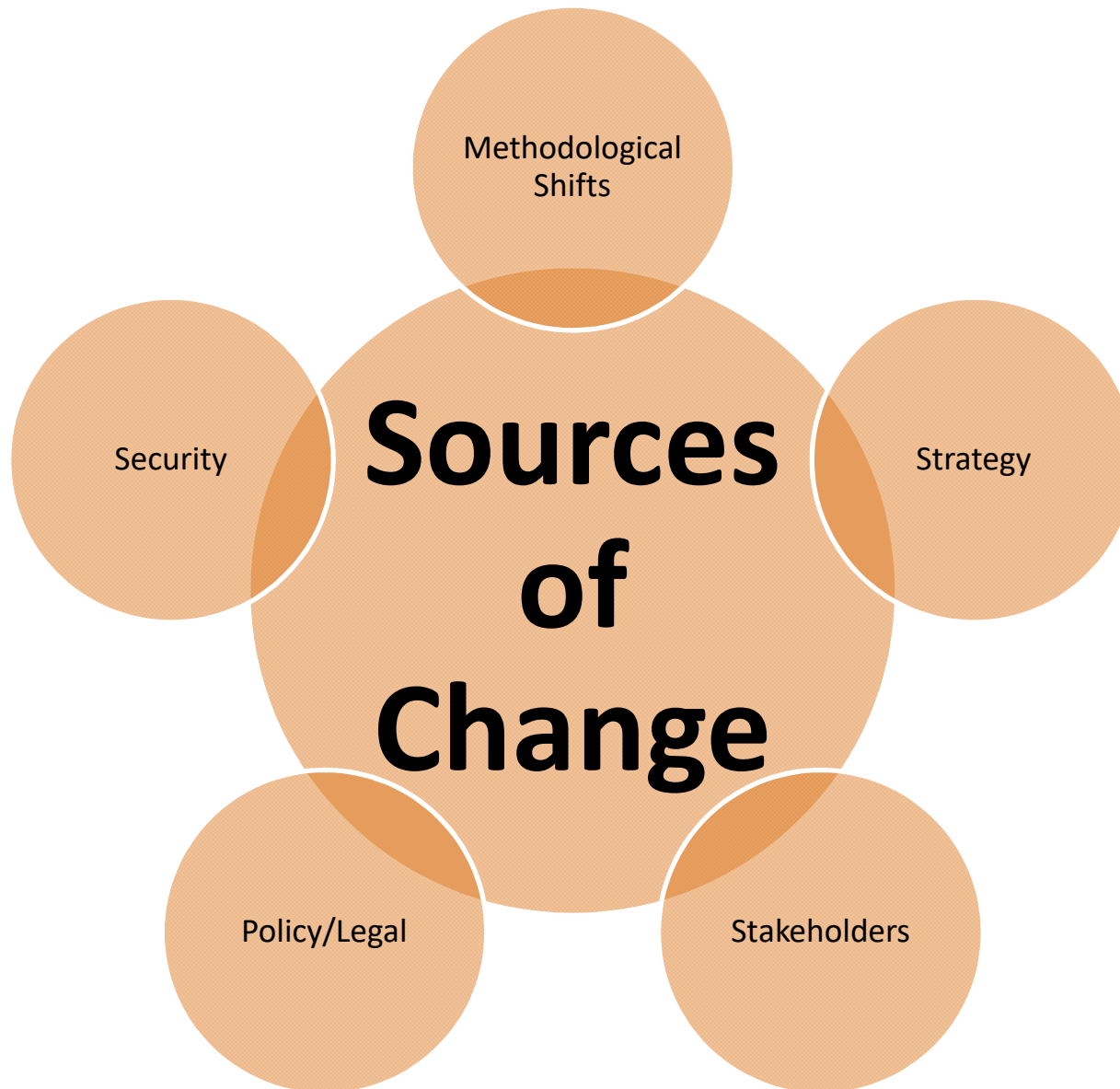
4. Psychometric Goals

Reality: Our Environment and Priorities
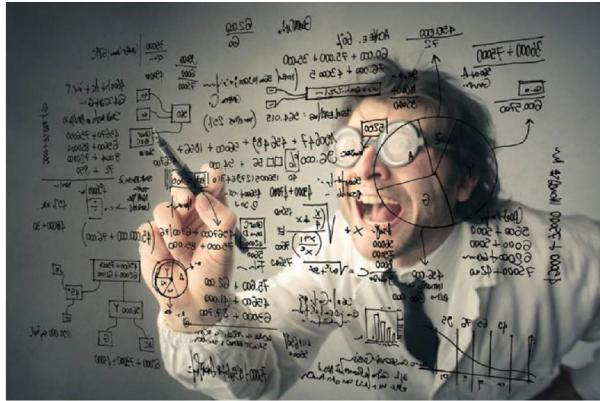
A good psychometrician or test developer will be trained to follow evidence based practices and measurement standards.
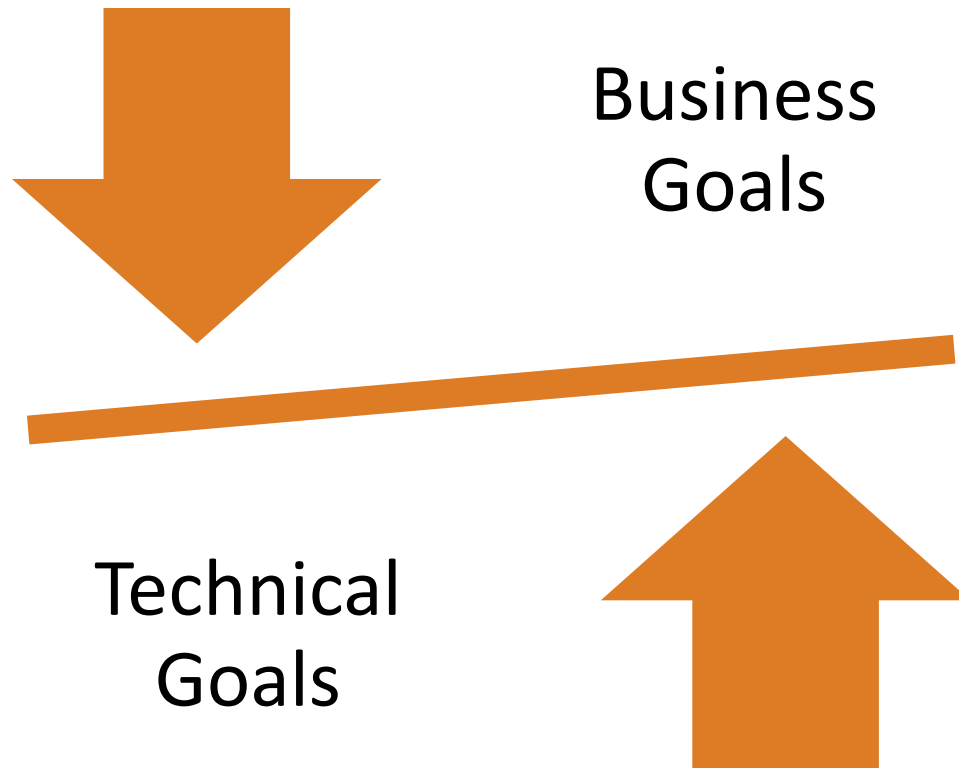
A good business executive will increase revenue while reducing costs, error, and time to complete projects.
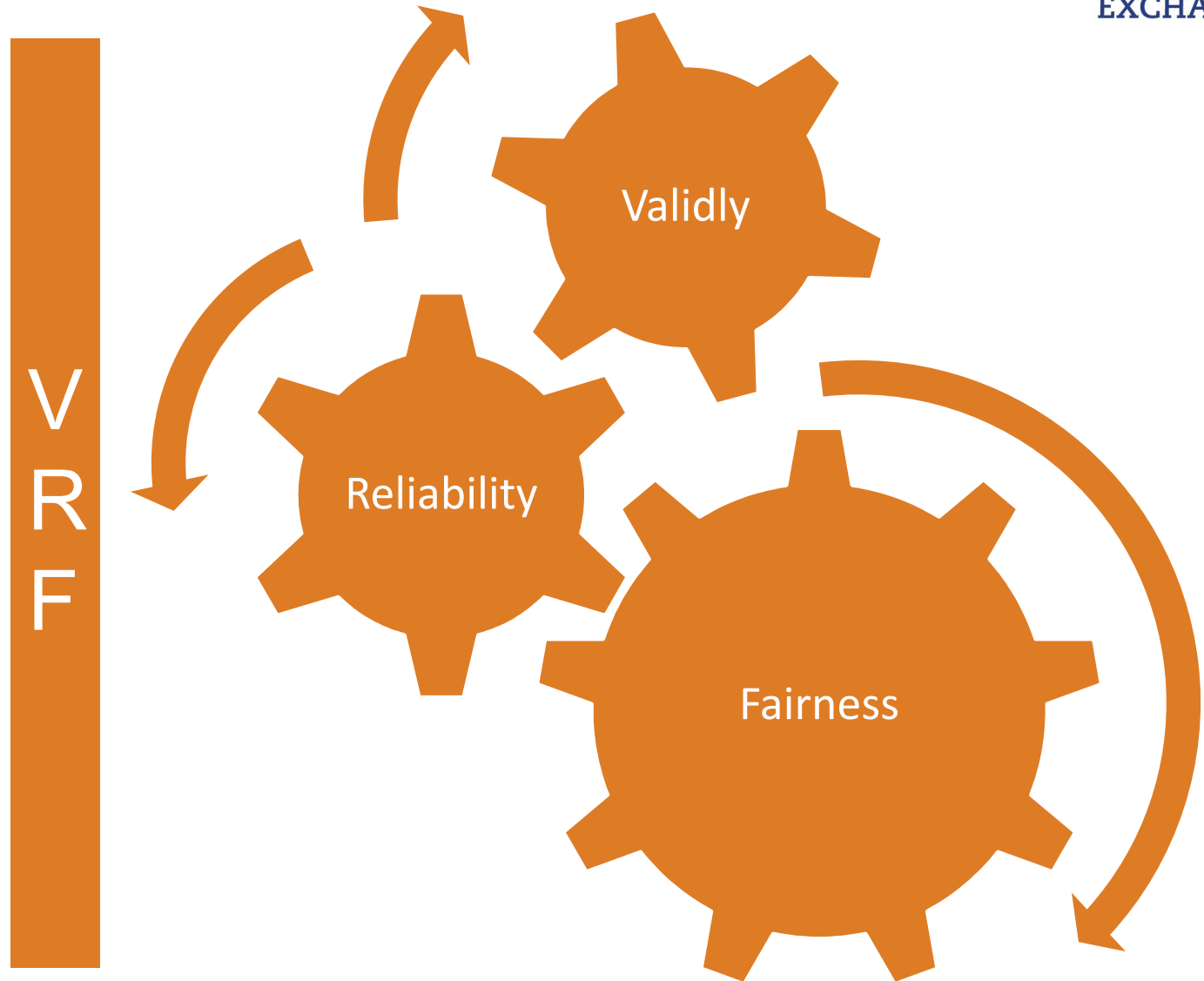
Overall Solution: Address Change Rather than Resist Change

Agile Methods May Lessen the Validity Argument to Meet Business Goals Yet It is **Not Absent** of Validity, Reliability, or Fairness.

Business Goals

Technical Goals

Psychometric Goals

0=Many Threats

Does Not Meet Accreditation Req...
Does Not Meet Testing Standards...
Violates/Does Not Use Evidence ...
Testing Experts Don't Approve Me...

VRF Sc...

that allo...

the **valic**...

organiza...

activitie...

testing a...

internat...

Minimal to No Threats

Accreditation Ready
Testing Standards/Legal Requirements
Adheres to Evidence Based Practices
...rts Approve Methods/Approaches Used

...by AERE
...ats to
...r
...various
...ices,
...S. and

Global Assessment of Functioning (GAF) Scale
(From DSM-IV-TR, p. 34.)

Consider psychological, social, and occupational functioning on a hypothetical continuum of mental health-illness. Do not include impairment in functioning due to physical (or environmental) limitations.

Code	(Note: Use intermediate codes when appropriate, e.g., 45, 68, 72.)
100 – 91	Superior functioning in a wide range of activities, life's problems never seem to get out of hand, is sought out by others because of his or her many positive qualities. No symptoms.
90 – 81	Absent or minimal symptoms (e.g., mild anxiety before an exam), good functioning in all areas, interested and involved in a wide range of activities. socially effective, generally satisfied with life, no more than everyday problems or concerns (e.g. an occasional argument with family members).
80 – 71	If symptoms are present, they are transient and expectable reactions to psychosocial stressors (e.g., difficulty concentrating after family argument); no more than slight impairment in social, occupational or school functioning (e.g., temporarily failing behind in schoolwork).
70 – 61	Some mild symptoms (e.g. depressed mood and mild insomnia) OR some difficulty in social, occupational, or school functioning (e.g., occasional truancy, or theft within the household), but generally functioning pretty well, has some meaningful interpersonal relationships.
60 – 51	Moderate symptoms (e.g., flat affect and circumstantial speech, occasional panic attacks) OR moderate difficulty in social, occupational, or school functioning (e.g.. few friends, conflicts with peers or co-workers).
50 – 41	Serious symptoms (e.g.. suicidal ideation, severe obsessional rituals, frequent shoplifting) OR any serious impairment in social, occupational, or school functioning (e.g., no friends, unable to keep a job).
40 – 31	Some impairment in reality testing or communication (e.g., speech is at times illogical, obscure, or irrelevant) OR major impairment in several areas, such as work or school, family relations, judgment, thinking, or mood (e.g., depressed man avoids friends, neglects family, and is unable to work; child frequently beats up younger children, is defiant at home, and is failing at school).
30 – 21	Behavior is considerably influenced by delusions or hallucinations OR serious impairment in communication or judgment (e.g., sometimes incoherent, acts grossly inappropriately, suicidal preoccupation) OR inability to function in almost all areas (e.g., stays in bed all day; no job, home, or friends).
20 – 11	Some danger of hurting self or others (e.g., suicide attempts without clear expectation of death; frequently violent; manic excitement) OR occasionally fails to maintain minimal personal hygiene (e.g., smears feces) OR gross impairment in communication (e.g., largely incoherent or mute).
10 – 1	Persistent danger of severely hurting self or others (e.g., recurrent violence) OR persistent inability to maintain minimal personal hygiene OR serious suicidal act with clear expectation of death.
0	Inadequate information.

Strategic Considerations

- Why? Exam's Intent
  - What will it do at the end of the day?
- Target Audience
  - US/Global
- Stakes –Low, Mid, High
- Accreditation Intent
- Eligibility Requirements
- Target Competency Level
- Item Type/Assessment Type
  - Number of items (50 at absolute minimum with new item writers)
- Reference List
- Test Delivery Method
- Proctoring

Credential Development Lifecycle: Agile MCQ Based Model

1. Job Analysis	→	2. Item Writing	→	3. Review Items/Form

↓

6. Standard Setting	←	5. Item Analysis	←	4. Administer Exam Beta Items

↓

7. Equating/Scaling	→	8. Administration	→	9. Maintenance

Agility Now!

- Recruit! Recruit! Recruit!
- Write! Write! Write!
- If needed, use the same SMEs for all steps (CAUTION HERE!)
- Hello! Everything can be done virtually
- Identify References
- Draft The Candidate Handbook
- Identify Vendors by Talking to Other Organizations in Your Space
- Consider a Technical Board of 2 to 3 Psychometricians
  - » Credible
  - » Influential
  - » Knows Psychometrics, Standards, and Accreditation
  - » Reasonable
- Consider Identifying an Accreditation Specialist
- Consider Identifying a Legal Consultant

1. Job Analysis: Agile

✓ Review existing resources to identify KSAs

✓ Keep the KSAs somewhat vague to ensure a longer lasting exam blueprint (e.g., Windows instead of Windows 10)
  - ✓ The specificity can occur at the item level and the acceptable references

✓ Use a small group of SMEs to identify and draft list of KSAs. They should be
  - ✓ Influential
  - ✓ Credible
  - ✓ Reasonable

✓ Prepare recruiting methods including drop-out considerations

✓ Use a larger pilot group to review KSAs using a survey

✓ Consider not using a survey based Job Analysis

✓ Use inclusion criteria for KSAs (based on survey results) only

✓ Use mathematical weighting only

2. Item Writing

- ✓ Over recruiting SMEs is key
- ✓ Train! Educate! Coach! Peer Coaching!
- ✓ Inform item writers of strategic overview of the exam and target item writing level including cognitive levels
- ✓ Feedback on at least three of their items at the very beginning is key
  - ✓ Allows you to remove poorly performing item writers early on without ever committing them further
- ✓ Consider facet/cloze items
- ✓ Three option multiple choice instead of four. YES three!
- ✓ Write and review items in real-time (somebody writes an item and reviewers edit it for content, psychometrics, and copyediting)
- ✓ Assign a few SMEs as reference and other gap "checkers". If there is a missing reference, domain/content area assignment, option, they would complete this.
- ✓ Use a professional copy-editor
- ✓ Provide item writing guidelines including a one-page item writing guideline
- ✓ DON'T SPEND MORE THAN FIVE MINUTES TO DRAFT THAT FINAL OPTION instead reassign it

3. Item Review

✓Use your better item writers

✓Standardize the item review process – use checklists

  ✓What should item writers review in terms of content?

    ✓Can the item writers be assigned to certain areas/domains

  ✓What should a psychometric review include?

  ✓What should the copyeditor be looking for?

4. Beta Testing

✓ Recruitment for individuals to sit for the Beta exam is key

✓ Consider using various methods for SME item difficulty estimates if you cannot Beta

5. Item Analysis

- ✓ Conservative versus liberal flags often dependent on stakes

- ✓ Replace flagged items without SME review, especially difficult items

- ✓ Small SME group review items

6. Standard Setting (Passing Score)

- ✓ Use methods requiring less time (e.g., Beuk, Hofstee)
- ✓ Use a smaller version of the exam
- ✓ Angoff considerations
  - ✓ Use blueprint to draft definition of minimally qualified
    - ✓ Can assign competency levels to each of the task and/or knowledge statements instead of drafting a version
  - ✓ Draft minimally qualified statement with one SME ahead of time and then use it with larger group
  - ✓ Use holistic definition by domain/content rather than analytical definition
  - ✓ Discrepancies between SME ratings to determine whether an item in the next round should be reviewed can be smaller
 - ✓ Only accept the Angoff and no standard error adjusted cuts

8. Administration

✓Use organization's current LMS/internet based systems

✓Use an online survey solution (e.g., Survey monkey)

✓Use IBT solutions

✓Use simple remote proctoring solutions such as not using live proctoring (if necessary)

9. Maintenance of Competence/Recertification

✓ Retest is the easiest

✓ Activities

   ✓ Use the three most popular activities since it is easier to track and standardize

Conclusion: Agility and VRF Can Work Together

Business Goals

Technical Goals
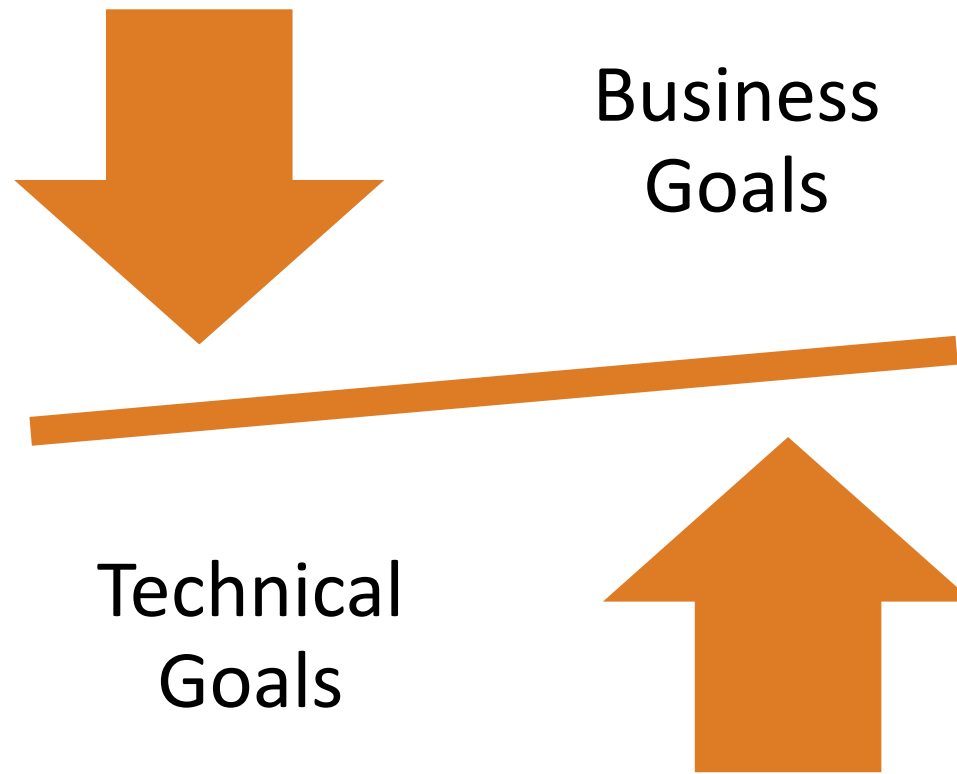
Questions

Manny Straehle, Ph.D., GISF

Founder and President
Assessment, Education, and Research Experts (AERE)

manny@aerexperts.com
www.aerexperts.com

Client-Centered, Solution-Focused, and Practicing the Practical