Building a better exam: Using educational theory to standardize assessment construction and review processes

Kimberly Daugherty, PharmD, PhD, BCPS
kdaugherty@sullivan.edu

Sarah Raake, PharmD, MSEd, BCACP, LDE
sraake@sullivan.edu

Ben Stephens, M.Ed.
bstephens@sullivan.edu
Objectives

• Discuss the differences between faculty trainings based on constructivism vs. cognitivism.

• Describe workflow best practices to streamline assessment creation and create uniformity using past student performance

• Explain job aids and additional supports provided to faculty

• Provide data on changes in creation time, blueprint consistency, exam performance, and post-item adjustments before and after initiation of the workflow
About us: Sullivan University College of Pharmacy & Health Sciences

• Private institution in Louisville, KY
• COPHS contains 2-year Master's level Physician Assistant program and 3-year Doctor of Pharmacy (PharmD) program
• Year-round, quarter-based system
• No large assessment office
• Limited checks-and-balances system within assessment creation/review process
• Faculty and course coordination self-sufficiency is a must
Constructivism states...
"Teaching and learning, especially for adults, is a process of negotiation, involving the construction an exchange of personally relevant and viable meanings"
(italics in original)
(Candy, 1991, as cited in Merriam et al., 2007, p. 293).

Generous Interpretation: Communities of Practice

Less Generous Interpretation: Wild West

Support provided: One off questions, troubleshooting, brainstorming
A course sequence hit gold

• Excellent plan for before, during, and after assessments
• How do we bring everyone back together?
How we changed

Cognitivism states...
"The problem can exist in only two states: (1) unsolved and (2) solved; there is not state of partial solution in between."
(Hergenhahn & Olson, 2005, as cited in Merriam et al., 2007, p. 285).

• Began a library of job aids
• Ongoing support
• Uniformity became the goal
• Everyone needs to take the same path (sorry Robert Frost)
Leading up to a change

2014: Implementation of electronic assessment platform

2015-2017: Consistency of coordinators achieved (specifically within course sequences)

2018: Course reliability, validity, and remediation discussions begin
Working through the change

• This process is a work in progress.
• Focus originated in major course sequences
• One sequence tried enough things to finally find a process that works
• Subtle influences on other course sequences due to word of mouth

• Currently, we are attempting to build consensus on "best practices" in item creation and assessment construction across a larger portion of the curriculum/program.
Cyclical review & planning process

- Pre-quarter pre-review
- During-quarter review
- Planning for future use
- Post-assessment review
Cyclical review & planning process

- Coordinator driven
- Faculty driven updates
- Coordinator driven structural revisions
- Peer-review of items from coordinator group

Pre-quarter pre-review

During-quarter review

Planning for future use

Post-assessment review/

- Coordinator driven with faculty input on adjustments
Coordinator guidance vs faculty ownness

• Coordinator “pre-chooses” what items can be re-used again and “hides” ones that are not deemed appropriate

• Coordinator provides faculty with tools to improve item writing
  • Internal comments on individual items
  • General job aids on item creation (stylistic considerations, P/P reminders, etc.)
  • Clear expectations set on what additional items are needed and focus of items (objectives not assessed, needs for final vs interim exams, etc.)
Maintaining intentional focus

• Instructional objective mapping
• Consistent (and correct) content/objective mapping
• Inclusion of solid rationale (for missed items)
• Stylistic best-practices
Looking backward- What has improved?

• Exam creation time
  • Increase in creation time initially
  • Overtime exam creation time has decreased by an average of 30%

• Blueprint consistency

• Performance of the exam (KR20)
  • Pre-implementation average: 0.76
  • Post-implementation average: 0.78

• Post item adjustments
  • Number of adjustments pre-implementation: average of 4-5 questions per assessment given full credit or made bonus
  • Number of adjustments post-implementation: 1-2 question given full credit or made bonus
First Impressions on the data

<table>
<thead>
<tr>
<th>Year</th>
<th>Exam 1</th>
<th>Exam 2</th>
<th>Exam 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>0.75</td>
<td>0.65</td>
<td>0.6</td>
</tr>
<tr>
<td>2019</td>
<td>0.6</td>
<td>0.7</td>
<td>0.76</td>
</tr>
<tr>
<td>2020</td>
<td>0.78</td>
<td>0.74</td>
<td>0.71</td>
</tr>
</tbody>
</table>
Looking forward: What continues to change

• Overarching education for all faculty (live and on-demand development sessions) covering a multitude of topics including assessment data evaluation, educational theory, etc.

• Getting buy-in from various stakeholders
Resources matter!

• Workflow tools and job aids are critical for success
• Include resources & templates for all steps of process including:
  • Pre-planning (at coordinator & faculty level)
  • Assessment blueprinting
  • Exam question review (how to adjust items)
• Make these resources very easily accessible
Link: https://libguides.sullivan.edu/sucophstechsupport

SU COPHS Faculty & Staff Technology Support: Home

Box 1

Let's get started. Please download the handbook below.

- Faculty Technology Handbook
Resource example: Structured review of assessment items

ASSESSMENT ITEM BEST PRACTICES

Please use this checklist as a reminder of assessment item best practices.

GENERAL REMINDERS
- Test comprehension and critical thinking, not just recall
- Use simple sentence structure and precise wording
- Place most words in the item stem
- Don’t look in the item stem
- Avoid being tricky; you are only tricking yourself
- Avoid negatives and avoid double negatives at all costs
- Keep the number of options consistent between items (i.e., the correct answer + 3 distractors)
- Keep all answer options parallel
- Avoid F-T, T-F, or “none of the above”
  - Choice that comply with A&AAB should never be only one option and never all options
  - Multiple-choice items (A, B, C, etc.) are not permissible
- Make all options plausible; all distractors should be chosen at least once
- Limit S&A items to a bare minimum
- Please name the item using the following nomenclature:
  - Lec
tu
er.
  - Topic
  - Instructional
  - Objective
  - Descriptor
  - Ex.: Lec10.01_GC05_DQ5_2AMAstep

ASSESSMENT ITEM CHECKLIST

☐ Is the item clear and concise?
☐ Is the item clinically accurate?
☐ Is the item relevant to the topic and at a minimal competence level of instruction?
☐ Is the item applicable to a novice or a generalist?
☐ Did you include all pertinent information in the question, including drug name, dose, route, frequency, duration, etc.?
☐ Does the item have a performance history? If so, what does it show? Can that information be used to make revisions?
☐ Are there any internal comments that can guide revisions?
☐ Is everything mapped in ExamSoft correctly?
  - Author’s name
  - Curriculum topic (ACPE Appendix I)
  - Bloom’s taxonomy level
  - Programmatic outcome (I, 1.1, 2.1, 3.2, etc.)
  - Is the instructional objective mapped in the nomenclature?
  - Is the instructional objective mapped correctly?
  - Are there an appropriate number of items for each instructional objective covered?
  - Is the item statistically appropriate?
☐ Have you proofed spelling and grammar in both stem and options?
☐ Is there a rationale included with appropriate level of detail (not referencing a specific slide)?
☐ Is the item in the appropriate content folder?
Examsoft - Assessment Level Posting Options

Please check off each step of this process. Then you may give this form to your proctor.

- Create a title
- Choose a folder
- MAX points > 0 *
- Check Percentage
- Check Raw Score
- Check security options
- Set time limit
- Note the universal resume code (keep it secured - no students)
- Backward navigation

Naming Convention:
Course# or Title _ Description

This is to help stay organized. No student will see it yet.

Caution: Avoid Max Pts: 0
Most faculty set the weight when the question was made.
Default weights is typical

Most common setup

Security Options:
- Secure
- Randomize Sequence
- Randomize At Choose
- Time Limit: 66 minutes
- Universal Resume Code B982375
- Options to Enable

Mostly common setup
Resource example:
Structured review of assessment items
Resource example: Adjusting items

- Follow your programs’ PIP on adjustments
- Now, look at items that have a difficulty of ≤50% and point-biserial < 0.2.
- Now, look at items that are a difficulty between 50-69% and pure-biserial < 0.2.
- You have a very difficult assessment. Consider _______
- It appears the assessment is not skewed in difficulty. We can move on to the next step.
- Does your assessment have >10% of questions that fall into this category?
- These questions discriminated well, but are difficult.
- Review the questions that have orange or yellow difficulty and no point-biserial highlighting.
- Highlight any item with a difficulty between 50 and 69% in orange.
- Highlight any item with a difficulty of ≤50% in pinkish.
- Now, look at the point biserial.
- Run a psychometric report
- Look at the K-R20
- Highlighting is almost done...
Small wins occur early

- Confidence in course performance improves
- Consistency increases within course and between courses in sequence
- Faculty and coordinators learn
Barriers to broader implementation

• Initial investment of time and resources (especially first year or two of implementation)
• Pre-work required before course begins (reviewing items, etc.)
• Buy-in is crucial- from both faculty and coordinators
• More work/authority at coordinator level
• Concern regarding loss of “academic freedom”
• Lack of education on best-practices (on educational theory and general educational best practices)
How is this process useful to both Faculty, Students, and College Administration?
Faculty

• Less student questions about exam questions
• More consistent in grading across exams as well as courses
• More data to guide course remediation design
  • Strengths and Opportunities reports
  • Objective's spreadsheet/mapping information
  • Assessment blueprinting
• More support of faculty grading decisions if student appeals
Students

• More data to prepare for comprehensive finals
  • Strengths and Opportunities reports
  • Objective's spreadsheet/mapping information
  • Assessment blueprinting

• Data to help students study for course remediation exams (if needed)
• Students can further cement knowledge through review of the exam question rationales
Administration

• Administrative data is always available for appeal situations...
  • Valid psychometric analysis
  • Consistent item review process
• Programmatic assessment data is "clean" and consistent
• Less student complaints on student surveys as students feel they are getting clear feedback
Key takeaways

• No "one size fits all" approach (slight variations between course sequences and departments)
• Need clear communication
• Need Administrative support to ensure the rules that are put in place are followed. Consistency is important!
Building a better exam: Using educational theory to standardize assessment construction and review processes

Kimberly Daugherty, PharmD, PhD, BCPS
kdaugherty@sullivan.edu

Sarah Raake, PharmD, MSEd, BCACP, LDE
sraake@sullivan.edu

Ben Stephens, M.Ed.
bstephens@sullivan.edu