Assessment Methods

College of Natural Sciences and Mathematics

02/09/15
Six-step assessment process

What do we want our students to learn and/or our units to accomplish?

Develop
Student Learning/Program Performance Outcomes

Document
Assessment Activities

Plan & Execute Improvement Actions

Collect & Analyze Data

Identify
Methods & Measures

Determine
Criteria of Success

How are we doing?
How do we know?

What evidence do we need to know to determine whether we are successful?

How are we documenting the assessment AND improvement activities/results?

What changes are we making?
Are the changes working?

How do we use data to confirm/improve our practices?
Sound outcomes are...

- Learner/customer centered, not instructor/unit centered
- Aligned with division/university goals and missions
- Specific, clear and concise
- Measurable
- Discrete (no “double-barrel” statements)
- Manageable
Some vocabulary

<table>
<thead>
<tr>
<th>Direct</th>
<th>vs.</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value-added</td>
<td>vs.</td>
<td>Absolute</td>
</tr>
<tr>
<td>Embedded</td>
<td>vs.</td>
<td>“Add-on”/ External</td>
</tr>
<tr>
<td>Formative</td>
<td>vs.</td>
<td>Summative</td>
</tr>
</tbody>
</table>
Choosing the right methods

- Valid
- Reliable
- Actionable
- Sustainable
- Start simple
- Prioritize embedded measures
- Use multiple methods
- Meaningful and engaging to stakeholders

Focus on “what we want to measure,” not “what existing measure we have”
Choosing the right sample

- Relevant
- Representative
- Reasonably sized

Determined by outcome, program circumstances
Direct assessment: Examples

- Exam/Quiz: Embedded or external
- Course assignment: In class or take home
- Concept inventory/map
- Paper/Presentation
- Project/Portfolio
- Recital/Exhibition
- Peer evaluation
Indirect assessment: Examples

- Self-reflection paper (can be short!)
- Survey**: Homegrown or published
- Interview: Individual or group
- Focus group: Structured or unstructured

**Surveys are not always indirect assessment
# Embedded or External?

<table>
<thead>
<tr>
<th></th>
<th>Pro</th>
<th>Con</th>
</tr>
</thead>
</table>
| **Embedded** | • Align directly with local curriculum/outcomes  
• Provide specific, actionable results  
• Meaningful to students/faculty  
• Embedded - No extra requirement/cost | • Time to develop and coordinate  
• Less rigorous process makes validity and reliability difficult to establish  
• Often isolated results; no comparison groups |
| **External** | • Have sound psychometric properties  
• Provide comparison groups  
• Provide test administration and/or scoring | • Do not align with local curriculum  
• Do not provide specific, actionable results  
• Not meaningful and/or burdensome to students/faculty  
• Costly |
Developing tests/surveys: Validity

<table>
<thead>
<tr>
<th>Validity</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face Validity</td>
<td>Do the questions/items seem relevant and representative of what we want to measure?</td>
</tr>
<tr>
<td>Content Validity</td>
<td>Do the questions/items reflect the content domain of what we want to measure?</td>
</tr>
<tr>
<td>Construct Validity</td>
<td>Do the questions/items really measure what we want to measure?</td>
</tr>
<tr>
<td>Convergent Validity</td>
<td>Do the questions/items converge on something that it theoretically is similar to?</td>
</tr>
<tr>
<td>Discriminant Validity</td>
<td>Do the questions/items diverge from something that it theoretically is NOT similar to?</td>
</tr>
<tr>
<td>Concurrent Validity</td>
<td>Do the questions/items distinguish between groups that they should theoretically distinguish between?</td>
</tr>
<tr>
<td>Predictive Validity</td>
<td>Do the questions/items predict something it should theoretically predict?</td>
</tr>
</tbody>
</table>
## Developing tests/surveys: Reliability

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Do the questions/items produce consistent results?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal Consistency</strong></td>
<td>Consistency across questions/items</td>
</tr>
<tr>
<td><strong>Inter-rater reliability</strong></td>
<td>Consistency across raters</td>
</tr>
<tr>
<td><strong>Intra-rater reliability</strong></td>
<td>Consistency across situations</td>
</tr>
<tr>
<td><strong>Test-retest reliability</strong></td>
<td>Consistency across time</td>
</tr>
<tr>
<td><strong>Parallel form reliability</strong></td>
<td>Consistency across alternative form</td>
</tr>
</tbody>
</table>

### Measures of Reliability

- **Cronbach’s alpha**
- **Inter-item correlation**
- **Percentage of agreement**
- **Cohen’s kappa/Intra-class correlation coefficient**
- **Correlation**
# Rubric

<table>
<thead>
<tr>
<th>Rating</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior</td>
<td>The whine have multiple purposes/targets; There is full audience participation; The pitch oscillated through entire frequency range...</td>
</tr>
<tr>
<td>Strong</td>
<td>The whine has a clear target; Many people participated in the whine; Pitch of whine is very high like fingernails on a chalkboard...</td>
</tr>
<tr>
<td>Incompetent</td>
<td>The whine has no clear target; Nobody noticed the whine; The pitch is very low...</td>
</tr>
</tbody>
</table>

(Details for Score 2-4 omitted)

### Analytic

<table>
<thead>
<tr>
<th>Score</th>
<th>Purpose</th>
<th>Audience</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>The purpose of the whine is to get a group response...</td>
<td>There is full audience participation...</td>
<td>The whine was seemingly endless, carrying on for days...</td>
</tr>
<tr>
<td>5</td>
<td>There is a clear target of the whine...</td>
<td>Everybody in the area noticed your whine...</td>
<td>Whine is audible over classroom noise...</td>
</tr>
<tr>
<td></td>
<td>(Details for Score 2-4 omitted)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>The whine is not directed at a target.</td>
<td>Nobody noticed; nobody cared.</td>
<td>The whine is little more than a whimper.</td>
</tr>
</tbody>
</table>
Developing a rubric

• Identify what you are assessing
• Identify the characteristics of what you are assessing
• Describe the best work you could expect using the characteristics [TOP category]
• Describe the worst work you could expect using the characteristics [BOTTOM category]
• Decide on your scale
• Develop description of intermediate categories
• Apply the rubric, and refine it to eliminate ambiguities

…and don’t forget…

• Calibrate the rubric, and further refine it
Is rubric the answer?

• **Effective scoring tool:**
  • Assess complex performance/products effectively
  • Capture multiple criteria/characteristics
  • Clarify expectations for faculty/students
  • Facilitate criterion-referenced grading
  • Facilitate the establishment of “criteria of success”

• **Potentially time-consuming:**
  • Development
  • Calibration: Inter-rater reliability
Remember...

- Triangulation
- Criteria of success
Some ideas

Physics SLO:
Students will solve problems by applying the primary physical theories: classical mechanics, thermodynamics, wave phenomena, electricity and magnetism, and modern physics.

<table>
<thead>
<tr>
<th>Method</th>
<th>Direct</th>
<th>Indirect</th>
</tr>
</thead>
</table>
| - Concept inventory (pre/post) on classical mechanics in Phyx 340  
 - Phyx 481 student project written report | | Junior Physics major focus group |

<table>
<thead>
<tr>
<th>Criteria of success</th>
<th>Direct</th>
<th>Indirect</th>
</tr>
</thead>
</table>
| - Students on average improved 50% on the Concept Inventory  
 - 80% of students receive satisfactory or above on the report | | The focus group participants report in general that they have a firm grasp of the primary physical theories, and can apply them to solve problems |
Some ideas

Physics SLO:
Students will apply appropriate mathematical tools to solve physical problems.

<table>
<thead>
<tr>
<th>Method</th>
<th>Direct</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Final exam question in Phyx 499 -- asking students to apply a particular math tool to solve physical problems</td>
<td>Graduate survey -- items asking students’ self-perceived ability to apply math tools to solve physical problems</td>
</tr>
<tr>
<td></td>
<td>(Even better) Same/Similar final exam question in Phyx 225 and 499</td>
<td></td>
</tr>
</tbody>
</table>

| Criteria of success          | 75% of students receive 80% or higher credit on the Phyx 499 final exam question. | The average rating of survey respondents is 3.0 or higher on a 4-point scale (3 = “confident”; 4 = “very confident”) that they are able to apply math tools to solve physical problems |
|                               | Students on average improved 20% on their score from Phyx 225 to Phyx 499           |                                                                          |
Some ideas

Physics SLO:
Students will demonstrate the ability to work collaboratively to collect and interpret data and draw conclusions.

<table>
<thead>
<tr>
<th>Method</th>
<th>Direct</th>
<th>Indirect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Capstone Advanced Lab (Phyx 481) project presentation</td>
<td>Student self-reflection paper on group project (Phyx 481) reviewed by a panel of faculty</td>
</tr>
<tr>
<td></td>
<td>- Peer evaluation (Phyx 481) on collaboration ability - 2 peers per student</td>
<td></td>
</tr>
<tr>
<td>Criteria of success</td>
<td>- 75% of students receive a score of 3 or higher on a 4-pt rubric for the presentation</td>
<td>90% of students are judged by the faculty as “able to collaborate successfully”</td>
</tr>
<tr>
<td></td>
<td>- 90% of students received an average rating of “satisfactory” in terms of collaboration ability from peers</td>
<td></td>
</tr>
</tbody>
</table>
Reporting expectations

- Step 1-3: March 1
- Step 4 & 5: June 15