Learning Improvement In Higher Education

BRIAN LEVENTHAL (LEVENTBC@JMU.EDU)
JAMES MADISON UNIVERSITY
Weigh Pig, Feed Pig, Weigh Pig

Weigh Pig (Pre-Assessment)

Table 1
Average pre-test and post-test scores

<table>
<thead>
<tr>
<th>Scale (range)</th>
<th>Pre-test (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics knowledge (0-25)</td>
<td>15.60 (4.04)</td>
</tr>
<tr>
<td>Self-efficacy total (14-84)</td>
<td>47.95 (16.16)</td>
</tr>
</tbody>
</table>

*Note.* The statistics knowledge test was a 25-item, multiple-choice test. Students responded to the 14-item self-efficacy survey with a 1-6 “No confidence at all” to “Complete confidence” scale. Students responded to the mindset survey with a 1-7 “Strongly disagree” to “Strongly agree” scale.

Self-Efficacy Means Pre- Boot Camp, Summer 2017
1 = no confidence to 6 = complete confidence

- Distinguish between three measures of central tendency
- Distinguish between population parameter and sample statistic
- Explain difference between sampling distribution and population...
- Identify scale of measurement
- Identify when to use mean, median, mode
- Identify skew
- Distinguish between descriptive & inferential
- Explain standard deviation
- Interpret p-value
- Distinguish between Type I and II errors
- Interpret results of statistical procedure
- Explain standard error
- Select correct statistic
- Identify factors that influence power

The Center for Assessment and Research Studies, James Madison University
Pedagogical Development:

Think-Pair-Share (Lyman, 1981) to facilitate Problem-Based Learning (Duch, Groh & Allen, 2001) using low-stakes interteaching (Boyce & Hyline, 2002) and Direct Instruction (Engelmann & Carnine, 1982)

Program Theory:

Meta-Cognitive Checks every hour!
Rehearsal of Schema (Felder & Brent, 2016)
Repeated Retrieval of information from long-term to short-term (Mastascusa et. al. 2011)
**Table 1**

**Average pre-test and post-test scores**

<table>
<thead>
<tr>
<th>Scale (range)</th>
<th>Pre-test (SD)</th>
<th>Post-test (SD)</th>
<th>Gain (post – pre)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics knowledge (0-25)</td>
<td>15.60 (4.04)</td>
<td>17.95 (3.46)</td>
<td>2.35</td>
</tr>
<tr>
<td>Self-efficacy total (14-84)</td>
<td>47.95 (16.16)</td>
<td>63.05 (13.52)</td>
<td>15.10</td>
</tr>
</tbody>
</table>

**Note.** The statistics knowledge test was a 25-item, multiple-choice test. Students responded to the 14-item self-efficacy survey with a 1-6 “No confidence at all” to “Complete confidence” scale. Students responded to the mindset survey with a 1-7 “Strongly disagree” to “Strongly agree” scale.
Program Level Learning Improvement
Faculty Development
- Learning Cognition and Retention Methods
- Department Support

Faculty Involvement
- Faculty Involvement

Pedagogical Development
- Pedagogical Development

Assessment
- Selecting and Designing Instruments
- Reliable Scores and Valid Instruments

Program Theory
- Program-Level or Course-level interventions
- Curriculum Development
- Implementation Fidelity

Student Learning Outcomes
- Mapping of Program to outcomes

Collecting Outcomes Information
- Analyzing Results
- Reliable Scores and Valid Instruments
- Selecting and Designing Instruments
- Faculty Involvement
- Faculty Involvement
- Pedagogical Development

Curriculum Development
- Department Support
- Program Theory

Department Support
- Curriculum Development
- Program Theory

Faculty Involvement
- Faculty Involvement
- Pedagogical Development

Pedagogical Development
- Pedagogical Development

Curriculum Development
- Curriculum Development
- Program Theory

Program Theory
- Program Theory
- Curriculum Development
- Program Theory
Examples of Research Related to Different Aspects of Learning Improvement
Program Theory
Implementation Fidelity

- Why start with something novel if what we planned to do should work?

First Year Orientation:
Connecting the dots through implementation fidelity improvement

“WE SPENT YEARS CREATING this orientation programming! I can’t believe the students aren’t learning from it. I guess we have to go back to the drawing board.”

Four Possible Outcomes

- **Implementation Fidelity HIGH (+)**
  - Outcomes Assessment GOOD (+)

  Programming was implemented as designed and the learning outcomes were realized, suggesting the designed programming may be effective.

- **Implementation Fidelity LOW (-)**
  - Outcomes Assessment POOR (-)

  No inferences can be made about designed programming; designed programming wasn’t implemented. Learning outcomes weren’t realized but designed programming should not be deemed ineffective. Instead, evaluate designed programming with higher implementation fidelity.

- **Implementation Fidelity HIGH (+)**
  - Outcomes Assessment POOR (-)

  Programming was implemented as designed, but the learning outcomes were not realized. This data should be used to inform modifications to the designed programming.

- **Implementation Fidelity LOW (-)**
  - Outcomes Assessment GOOD (+)

  Programming was not implemented with high fidelity. Thus, the designed programming should not be deemed effective. That is, one should not infer the positive outcomes results were due to the designed programming.
How to Assess Implementation Fidelity?

- **Program Differentiation**
  - Detail the specific features of the program (e.g., activities, assignments, demonstrations, curriculum) that should enable students to achieve each SLO
  - Must be completed before gathering data on the four components below

- **Adherence**
  - Indicate whether the specific program features detailed during program differentiation were actually implemented as intended
  - A simple "Yes" or "No" is adequate

- **Quality**
  - For the specific features adhered to, describe how well they were implemented or delivered
  - A quality rating coupled with some descriptive text is useful

- **Exposure**
  - Indicate if the duration or amount of time actually spent on the specific program features aligns with intended duration
  - Record the number of students exposed to the programming

- **Responsiveness**
  - Observe and describe the responsiveness or engagement of students during specific program features
  - Ask students about their experience during the programming
Ethical Reasoning in Action: The Madison Collaborative

The Eight Key Questions (8KQ) framework (Madison Collaborative, 2013):

- Ask critical, open-ended questions *before deciding and acting*.
- Key Questions:
  - Fairness
  - Outcomes
  - Responsibilities
  - Character
  - Liberty
  - Empathy
  - Authority
  - Rights
Assessment of Student Learning Outcomes:

- **Ethical Reasoning Recall Test (ERRT)**
  - SLO 1: Memorization of the 8KQs

- **Ethical Reasoning Identification Test (ERIT)**
  - SLOs 2 & 3: Identifying the relationship of specific KQs to a decision/rationale scenario

- **Survey of Ethical Reasoning (SER)**
  - SLOs 6 & 7: Attitudes toward ER

- **Ethical Reasoning – Writing (ER-WR)**
  - SLO 5: Applying KQs to one's own personal, professional, and civic life

- **Ethical Reasoning-Writing 2 (ER-WR2)**
  - SLO 4: Applying KQs to a provided scenario

The Center for Assessment and Research Studies, James Madison University
SLO 5

To their own ethical situation or dilemma, students will evaluate courses of action by applying (weighing and, if necessary, balancing) the considerations raised by the 8KQ.
The Eight Key Questions (8KQ)

Outcomes

Rights

Character

Liberty

Empathy

Authority

Responsibility

Fairness

Ethical Issue
Which Apply?

Outcomes

Rights

Character

Liberty

Empathy

Authority

Responsibility

Fairness

Ethical Issue
Conduct Analyses

Ethical Issue

- Fairness
- Outcomes
- Character
- Empathy
Weigh and Decide

- Character
- Empathy
- Outcomes

Fairness
### James Madison University’s Ethical Reasoning Rubric

<table>
<thead>
<tr>
<th>Insufficient (0)</th>
<th>Marginal (1)</th>
<th>Good (2)</th>
<th>Excellent (3)</th>
<th>Extraordinary (4)</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Ethical Situation: Identifying the ethical issue in its context</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No attempt or only a superficial reference to an ethical issue.</td>
<td>No attempt or only a superficial reference to an ethical issue.</td>
<td>Ethical issues are identified and referenced.</td>
<td>Ethical issues are identified and referenced in full.</td>
<td>Meets criteria for <strong>Excellent AND...</strong></td>
<td></td>
</tr>
<tr>
<td><strong>B. Key Question Applicability: Describing which of the 8 KQs are applicable or not applicable to the situation and why</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No rationale provided for the applicability or inapplicability of any key questions to the ethical situation.</td>
<td>Provides a rationale for the applicability or inapplicability of two key questions to the ethical situation.</td>
<td>Provides a rationale for the applicability or inapplicability of six key questions to the ethical situation.</td>
<td>Provides a rationale for the applicability or inapplicability of six key questions to the ethical situation.</td>
<td>For all eight questions provides a rationale for its applicability or inapplicability to the ethical situation.</td>
<td></td>
</tr>
<tr>
<td><strong>C. Ethical Reasoning: Analyzing the individual KQs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No attempt to analyze any of the referenced key questions.</td>
<td>Analyzes two or more key questions. Includes accurate articulation of the key questions to the ethical situation.</td>
<td>Analyzes three or more key questions. Includes accurate articulation of the key questions to the ethical situation.</td>
<td>Analyzes all eight key questions. Includes accurate articulation of the key questions to the ethical situation.</td>
<td>Meets criteria for <strong>Excellent AND...</strong></td>
<td></td>
</tr>
<tr>
<td><strong>D. Ethical Reasoning: Weighing the relevant factors and deciding</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No judgment is presented or judgment presented with no rationale.</td>
<td>Uses products of analysis and provides some weighing to make a decision.</td>
<td>Conveys weighing approach using analysis products. Products include an intelligible basis for judgment.</td>
<td>Meets criteria for <strong>Excellent AND...</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Baseline, 2013**
- Mentioning the 8 KQs or equivalent terms
- References one key question

**Target, 2020**
- References all eight key questions

**Where we are now, 2017**
- Baseline:references one key question
- Target: references eight key questions
Developed by experts in ethical reasoning and assessment

5 elements developed based on theory
◦ Thought to increase hierarchically in difficulty

5 score levels

Assumptions:
1) All raters use the same thought processes to derive their scores
2) Rater thought processes match those of expert developers
3) The thought processes are in alignment with the intended interpretations and uses of scores

<table>
<thead>
<tr>
<th>Insufficient</th>
<th>Marginal</th>
<th>Good</th>
<th>Excellent</th>
<th>Extraordinary</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

A. Ethical Situation: Identifying ethical issue in its context
- No reference to decision option(s).
- Implicit reference to decision options AND/OR little context given regarding decision option(s).
- Explicit but unorganized reference to decision option(s) and context.
- Clear description of decision option(s) and context.
- Meets criteria for Excellent AND...
  ◦ Context treated with nuance
  ◦ Builds tension with organization and word choice.

B. Key Question Reference: Mentioning the 8 KQs or equivalent terms
- Reference to zero or only one key question.
- Vague references to key questions OR only two key questions referenced.
- References four key questions.
- References six key questions.
- References all eight key questions.

C. Key Question Applicability: Describing which of the 8 KQs are applicable or not applicable to the situation and why
- No rationale provided for the applicability or inapplicability of any KQs to the ethical situation.
- Provides a rationale for the applicability or inapplicability of two key questions to the ethical situation.
- Provides a rationale for the applicability or inapplicability of four key questions to the ethical situation.
- Provides a rationale for the applicability or inapplicability of six key questions to the ethical situation.
- For all eight questions provides a rationale for its applicability or inapplicability to the ethical situation.

**SPECIAL NOTE: If author identifies fewer than three applicable KQs, then Criteria "D" and "E" can be scored no higher than (1) "Marginal***

D. Ethical Reasoning: Analyzing Individual KQs
- No attempt to analyze any of the referenced key questions.
- Analysis attempted using two or more key questions. Typically incorrect assumption of the key questions to the ethical situation. Account is unclear, disorganized, or inaccurate.
- Analysis attempted using three or more key questions. Basically accurate assumption of the key questions to the ethical situation. Account is unclear or disorganized.
- Analysis attempted using three or more key questions. Accurate assumption of the key questions to the ethical situation. Account is clear and organized.
- Meets criteria for Excellent AND...
  ◦ Nuanced treatment of key questions, for example: creates subtle distinctions
  ◦ uses analogies or metaphors
  ◦ considers different issues within same key question.

**SPECIAL NOTE: If Criterion "D" is scored a 0 or 1 then Criterion "E" can be scored no higher than (1) "Marginal***

E. Ethical Reasoning: Weighing the relevant factors and deciding
- No judgment is presented or judgment presented with no rationale.
- Uses products of the analysis and provides some weighing to make a decision. Account is unclear, disorganized, or inaccurate.
- Conveys weighing approach using analysis products. Provides an intelligible basis for judgment.
- Meets criteria for Good AND...
  ◦ Logically terminates in decision that will be reached.
- Meets criteria for Excellent AND...
  ◦ Products of analysis weighted to make judgment compelling.

James Madison University's Ethical Reasoning Rubric

The Center for Assessment and Research Studies, James Madison University
Overview of DRS Studies

1) The Diagnostic Rating System: Development and evidence for use (Holzman & Curtis, 2016)
   ◦ Background on ethical reasoning at JMU
   ◦ Development of the Diagnostic Rating System (DRS)
   ◦ Pilot study results

2) The Diagnostic Rating System: A rating method to facilitate intentional programmatic changes
   ◦ Curricular advantages of the DRS compared to rubric-scoring
   ◦ Examples of how the DRS can be used to inform curriculum modification

3) IRTree models: Application to rater response processes (Myers, Holzman, Ames & Leventhal, 2017)
   ◦ Application of two measurement models for assessing the influence of raters on essay scores
   ◦ Evidence for validity of inferences from DRS scores

4) Detecting Rater Fatigue Using Bayesian Change Point Analysis (Ames, Curtis, Holzman, Ong, & Smith, 2017)
   ◦ To determine the extent to which rater fatigue is impacting performance assessment essay scores at James Madison University, and
   ◦ Investigate whether a new method of assigning scores can mitigate the fatigue effect.
### A. Ethical Situation: Identifying ethical issue in its context

<table>
<thead>
<tr>
<th>Insufficient 0</th>
<th>Marginal 1</th>
<th>Good 2</th>
<th>Excellent 3</th>
<th>Extraordinary 4</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>No reference to decision option(s).</td>
<td>Implicit reference to decision options AND/OR little context given regarding decision option(s).</td>
<td>Explicit but unorganized reference to decision option(s) and content.</td>
<td>Clear description of decision option(s) and content.</td>
<td>Meets criteria for Excellent AND...</td>
<td></td>
</tr>
</tbody>
</table>
| **Criteria:** | **Context treated with nuance**
| **Reason:** | **Builds tension with organization and word choice.** |

### B. Key Question Reference: Mentioning the 8 KQs or equivalent terms

<table>
<thead>
<tr>
<th>Reference to zero or only one key question.</th>
<th>Vague references to key questions OR only two key questions referenced.</th>
<th>References four key questions.</th>
<th>References six key questions.</th>
<th>References all eight key questions.</th>
</tr>
</thead>
</table>

### C. Key Question Applicability: Describing which of the 8 KQs are applicable or not applicable to the situation and why

| No rationale provided for the applicability or inapplicability of any KQs to the ethical situation. | Provides a rationale for the applicability or inapplicability of two key questions to the ethical situation. | Provides a rationale for the applicability or inapplicability of four key questions to the ethical situation. | Provides a rationale for the applicability or inapplicability of six key questions to the ethical situation. | For all eight questions provides a rationale for its applicability or inapplicability to the ethical situation. |

**SPECIAL NOTE:** If author identifies fewer than three applicable KQs, then Criteria “D” and “E” can be scored no higher than (1) “Marginal.”

### D. Ethical Reasoning: Analyzing individual KQs

| No attempt to analyze any of the referenced key questions. | Analysis attempted using two or more key questions. Typically incorrect ascription of the key questions to the ethical situation. Account is unclear, disorganized, or inaccurate. | Analysis attempted using three or more key questions. **Basically accurate** ascription of the key questions to the ethical situation. Account is unclear or disorganized. | Analysis attempted using three or more key questions. **Accurate** ascription of the key questions to the ethical situation. Account is clear and organized. | Meets criteria for Excellent AND...
|-----------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| **Criteria:** | Nuanced treatment of key questions, for example:
| **Reason:** | *elucidates subtle distinctions*
| **Reason:** | *uses analogies or metaphors*
| **Reason:** | *considers different issues within same key question.* |

**SPECIAL NOTE:** If Criterion “D” is scored a 0 or 1 then Criterion “E” can be scored no higher than (1) “Marginal.”

### E. Ethical Reasoning: Weighing the relevant factors and deciding

| No judgment is presented OR judgment presented with no rationales. | Uses products of the analysis and provides some weighing to make a decision. Account is mostly disorganized, or inaccurate. | Conveys weighing approach using analysis products. Provides an intelligible basis for judgment. | Meets criteria for Good AND...
|--------------------------------|---------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| **Criteria:** | **Logically terminates in decision that will be reached.**
| **Reason:** | **Products of analysis weighed to make judgment compelling.** |

---

**Ex: Rubric Element A**

The Center for Assessment and Research Studies, James Madison University
Q1) Did author reference any decision options?

- Yes – Explicitly
- Yes – Implicitly
- No

Q2) How did author present the decision options and the context surrounding them?

- Organized & Clear
- Disorganized & Unclear
- Not Present

Q3) Is the context treated with nuance and does the writing build tension with organization and word choice?

- Nuance & Tension
- Nuance OR Tension
- No
Motivation research and Intervention in Low-Stakes Testing Environments

“If we don’t know a student’s current standing, how can we speak to improvement?”

Thai Ong & Dr. Dena Pastor

- Effort-Monitoring Computer Based Test (CBT)
  - A computer based test where warning messages are presented to examinees who rapidly-guess (have incredibly fast response times) on three items in a row.

- Study 1: The warnings improve examinee behavior in low-stakes testing and ultimately, examinee performance.

- Study 2 (In Progress): Investigation of whether warnings given to all examinees, regardless of their behavior, at various points in the test are as effective at increasing effort/performance.

- Study 3 (In Future): Use of item response theory (IRT) models for low-stakes tests to identify at what point examinees appear to switch to rapid guessing behavior.
Low Stakes Testing

- Although the tests are low stakes for students, test results may be used for high-stakes decisions, such as accreditation or teacher retention decisions, which impact institutions or its professional members (No Child Left Behind Act, 2002; U.S. Department of Education, 2006).

- Examined change in test-taking effort over the course of a three-hour, five test, low-stakes testing session

Results:

- Latent growth modeling results indicated that change in test-taking effort was well-represented by a piecewise growth form, wherein effort increased from test 1 to test 4 and then decreased from test 4 to test 5.

- The degree to which examinees perceived a particular test as important was related to effort for the difficult, cognitive test but not for less difficult, noncognitive tests.

- There was significant variability in the rates of change in effort, which could be predicted from examinees’ agreeableness. Interestingly, change in test-taking effort was not related to change in perceived test importance.

Thank you!

- Learn more about JMU Assessment: https://www.jmu.edu/assessment/
- Learning Improvement by Design: https://www.jmu.edu/assessment/Visitor/Student-Learning.shtml
- Assessment Resources: https://www.jmu.edu/assessment/Visitor/AssessmentResources.shtml

E-mail: Brian Leventhal leventbc@jmu.edu