

2019 Education Council Symposium: Assessing Quality in Medical Physics Education

George Starkschall, PhD
Professor Emeritus, MDACC
Executive Secretary, CAMPEP

Acknowledgments

- Frada Boxer
- Ed Jackson
- Seth Stein

Disclosures

- The presenter is Executive Secretary of CAMPEP. The opinions presented are solely those of the presenter and do not necessarily reflect CAMPEP policy.

Purpose of Symposium

1. To present several definitions of the phrase “quality of education”
2. To identify and evaluate methods for assessing quality of education

Purpose of Symposium

3. To apply these methods to assess the quality of a didactic medical physics educational program – E Jackson, PhD, University of Wisconsin

Purpose of Symposium

4. To apply these methods to assess the quality of a clinical medical physics training program – K Hendrickson, PhD, University of Washington

Ultimate Goal

- Improve the quality of teaching of medical physics

What do we mean by “quality of teaching”?

What do we mean by “quality of teaching”?

- Potter Stewart approach

What do we mean by “quality of teaching”?

“I can’t define pornography, but I know it when I see it.”

- Potter Stewart
US Supreme Court Justice



What do we mean by “quality of teaching”?

- Brown M&Ms approach

What do we mean by “quality of teaching”?

- Brown M&Ms approach

Munchies

- Potato chips with assorted dips
- Nuts
- Pretzels
- ➔ M & M's (WARNING: ABSOLUTELY NO BROWN ONES)
- Twelve (12) Reese's peanut butter cups
- Twelve (12) assorted Dannon yogurt (on ice)

What do we mean by “quality of teaching”?

- Brown M&Ms approach
 - Just as the absence of brown M&Ms in the candy dish indicates the producer has read the contract, demonstration of adherence to educational standards indicates that the educational leadership is conscientious about educational quality
 - Brown M&Ms serve as a surrogate for confirmation that the producer has read the contract

...but, before we start

What are we really trying to determine?

1. Longitudinal assessment – Assessing how a program’s educational quality changes over time
2. Lateral (transverse) assessment – Comparing one program’s educational quality to that of another program

Longitudinal Assessment

- More reliable quantity to measure – removing source of systematic error
- Useful tool for Continuous Quality Improvement (CQI) program

Lateral Assessment

- Normally what we think of when we compare programs
- Is Program A better than Program B?
- Not clear as to reliability or value of lateral assessment
- Should we even be comparing programs?

On to the indicators of “quality of teaching”

1. Value added
2. Assessing outcomes
3. US News & World Report
4. Self-reporting
5. Endurance
6. Luxury and prestige
7. Conformance to requirements
8. Continuous improvement

Etc, etc

Value Added

- Measure improvement of student knowledge or skills
- Difficulties
 - Value has many dimensions
 - How to measure initial competence
 - Effects may take several years to express
 - Measurement may be complex and expensive
 - Not clear what is actually being measured

Assessing Outcomes

- Evaluate graduates on skills and capabilities acquired
- Examples:
 - ABR Part 1 pass rate – graduate programs
 - ABR Part 2 pass rate – residency programs
- ABR exam is minimal competency exam
- Only measures knowledge of clinical skills
 - What about implementation of clinical skills?
 - What about research quality?

US News & World Report

- Multidimensional
 - Program spending, faculty salaries, student/faculty ratio, class size
 - Graduation rates
 - Survey Program Directors, practicing physicists

US News & World Report

- Criticism:
 - No links to value added
 - Not a measure of what students actually learned
 - Multidimensional, with no clear way of combining “orthogonal” measures
 - Not clear how much faculty, graduates of one program really know about the quality of other programs

Self-Reporting

- Survey recent alumni/ae to assess how much benefit gained from program
 - To what extent did program significantly improve knowledge and skills

Self-Reporting

- College Results Instrument (CRI)
 - Developed by the Institute for Research on Higher Education for the National Center for Postsecondary Improvement
 - Can be adapted to medical physics education
 - Presents real-life scenarios to >5 yr graduates
 - Respondents asked to what extent they are well-prepared to work on described tasks

Processes and Participation Rates

- Identify what students actually do in training program
 - # problem sets, # tests, # treatment plans, etc
 - Correlation between activity and student learning not established

Even More Quality Metrics

- Endurance – equate longevity of program with quality
- Luxury and prestige – up-to-date facilities, research funding
- Conformance to requirements – adherence to performance standards
- Continuous improvement – rate of innovation and improvement

Perspectives on Assessment

- End of educational program (e.g., ABR pass rate)
 - Identify program deficiencies
 - Compare institutions
 - Longitudinal data may assess value added
- Direct assessment of student learning
 - Develop rubrics to evaluate performance
 - Rapid identification of needs for improvement
- Activities that imply student learning (e.g., time spent on projects, etc.)
 - Potential cause and effect relationship between behaviors and outcomes

Assessing clinical education

- ABR pass rates – Part I (graduate programs); Part 2 (residency programs)
- Issues:
 - Appears to be only quantitative metric
 - How to calculate pass rate
 - Danger of “teaching to the test” (may not be an issue)
 - Accessing information from ABR

Assessing research education

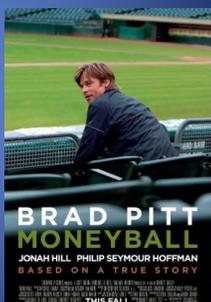
- Number of peer-reviewed publications per student in program
- Issues:
 - Validity of metric
 - Danger of “padding”
 - How far downstream to measure
 - Require self-reporting

Assessing research education

- Number of peer-reviewed publications per student in program weighted by journal impact factor
- Issues:
 - Validity of metric
 - Validity of impact factor as measure of research quality
 - Danger of “padding”
 - How far downstream to measure
 - Require self-reporting
 - Calculation may require excessive resources

Weakness of most approaches

- Most approaches to quality assessment are qualitative
- Really want to have quantitative measure of programmatic quality
- No good measures at present
- “Moneyball” is example of replacing qualitative assessment with quantitative



This meeting

- PO-GePV-E-1: “Evidence-based graduate program design with data analytics,” R Khan, T Mazur, S Mutic
 - Use machine learning to relate program characteristics to “success”
 - Success defined as ratio of number employed to number graduated for each of last 5 years

“Moneyball” works in baseball

- How can we apply it to educational programs?



Sports Illustrated, June 25, 2014

Challenges for the future

- Rapid advances in technology have resulted in new and innovative methods of providing education
- Lack of attention has been paid to alternative methods of providing education
- How can “soft” skills be assessed?
 - May be more reliable indicator of professional success (Hidden Brain, NPR Podcast, May 13, 2019)

For further reading

- DC Bennett, "Assessing quality in higher education," Liberal Education (2001)
- J Dew, "Quality issues in higher education," J Quality and Participation (2009)
- T Ransom, et al, "New Approaches to Judging Quality in Higher Education"
- SK Sokoya, "Measuring quality in graduate education," PDFs.semanticscholar.org

Continuing our discussion

- Now let's see how programs actually assess quality
- E Jackson, Univ of Wisconsin – graduate programs
- K Hendrickson, Univ of Washington – residency program
- Panel discussion

Thank you



Questions?

Text to:
713 906-7259

Email to:
drstarkschall@gmail.com
