INCORPORATING INTER-RATER RELIABILITY INTO YOUR CURRICULUM

- Jerry Moss, MPA, RRT
  RT Program Director
  Metropolitan Community College
  Omaha, NE.

- Tim Kelly, BPS, RRT
  RT Program DCE
  Metropolitan Community College
  Omaha, NE.
Program Objectives

- What is Inter-Rater Reliability?
- Why is it important to use Inter-Rater Reliability?
- When is it necessary to ensure Inter-Rater Reliability?
- How can IRR be Assessed/Implemented/Used?
“Everyone is entitled to their own opinion, but not their own facts.”

Daniel Patrick Moynihan
4-term New York Senator
What is Inter-Rater Reliability (IRR)?

- The extent to which 2 or more raters agree.
- Dependent upon the raters to be consistent in their evaluation of behaviors or skills.
- Unbiased measurement of student competency.
- Addresses the consistency of the implementation of evaluation systems being utilized.
Why is IRR important?

- CoARC Accreditation Standard Requirement
  - Student Evaluation: 3.09, 3.10, 3.11
- IRR imposes some level of objectivity.
- IRR provides validation of evaluation results.
  - Certify skills competency
  - Assign the appropriate “grade”
- Increases “confidence” that people are following the guidelines in like manner.
Student Evaluation Standard 3.09

The program must conduct and document evaluations with sufficient frequency to keep students apprised of their progress toward achieving the curriculum competencies, and to allow immediate identification of learning deficiencies and the development of a means for remediation in a reasonable time frame.
Student Evaluation Standard 3.10

The program must administer evaluations uniformly and equitably to all students in the program for didactic, laboratory, and clinical education components.
Student Evaluation Standard 3.11

The program must develop processes that facilitate the development of inter-rater reliability among those individuals who perform student clinical evaluations.
3.11 Evidence of Compliance

- Records of training participation by clinical evaluators.

- Results of a review of student evaluations for the purposes of determining IRR.
When is it Necessary to Ensure IRR?

- Multiple Faculty/Instructors/Preceptors
  - Paid
  - Unpaid
- Assessments contribute to determination of competency attainment
  - Skills check-offs in laboratory/clinical settings
  - Oral presentations in the classroom setting
Who Uses IRR?

- WE DO!
- OTHERS CONSIDERING
  - ACEN-Nursing
  - CAPTE-Physical Therapy
  - LCME-Medical Education
  - CAAHEP-Anesthesia Tech., EMT/Paramedic, Exercise Technology, Medical Assisting, Neurodiagnostics, Polysom., Perfusion Tech., Sonography & Surgical Tech..
  - JRCERT-Radiography

“The pioneers take the arrows, while the settlers get the land”
The Reality of Using Multiple Evaluators

• Difference in education, experience, and values.
• Varied levels of knowledge and experience with student evaluation.
• May have no true leverage to force adherence to guidelines.
• Preceptor training “adopts” them into the program “family”
How Can I Assess IRR?

“You can observe a lot just by watchin’.”

-Yogi Berra
How Can I Assess IRR?

- 3 Categories of statistical methods for computing inter-rater reliability
  1. Consensus Estimates
  2. Consistency
  3. Measurement Estimates
How Can I Ensure IRR?

- Implement Preceptor Training annually with updates as needed.
  - Roles and Responsibilities of Preceptor
  - Direct Observation Skills
  - Effective Feedback & Evaluation
  - Inter-Rater Reliability Exercises

- Keep Records !!
Procedure for Assessing Competency (PAC)

- PAC GRADE SHEET
  - Psychomotor Skills
  - Affective Behaviors
  - Cognitive Skills – 5 Questions

- PAC Performance Elements
  - Step-By-Step
  - Each Element Matched with Acceptable Behavior
TRUST is Built on

FIRST IMPRESSIONS
Procedure for Assessing Competency (PAC)

- Review of Task/Skill Set
  - Which Steps are Critical
  - Does it Make Sense?
  - What is Missing?

- Rate Student Performance
  - 5 Independent
  - 4 Competent
  - 3 Marginal
  - 2 Unsatisfactory
  - 1 Dependent
  - NA Not Applicable
“The Bottom Line”

- Did the student perform the procedure well enough to establish competency
  - **Yes** = Good to go !!!
  - **No** = Why Not???
    - Minor Unsatisfactory (small details)
    - Major Unsatisfactory (BIG details)
Inter-Rater Reliability Exercise

- Review of Evaluation Instrument-Procedure for Assessing Competency (PAC)
- Evaluate Video Demonstration
- Review Competency Assessments
  - Establish Consensus
  - Examine Differences
Inter-Rater Reliability of a Respiratory Therapy Preceptor Training Program

- The Ohio State University Medical Center
- 12 RT Preceptors
- 5 Videos
- 23 Behaviors

Conclusion: “RT clinical preceptor training modules evaluated have a high degree of inter-rater reliability”
Video Demonstration

http://www.mccneb.edu/resp/respvid.asp
Evaluation of Video Demonstration

- Preceptors watch video & submit evaluations

- Evaluations are assessed for IRR
  - Consistent?
  - If raters consistent, nothing more need be done
Evaluation of Video Demonstration

- If Raters Inconsistent?
  - Due to poor instructions?
  - Due to something observed/not observed in video?
  - Due to false assumptions made from video?

- Retrain the Raters

- Revise the Video
How Close Are We??

- Percent Agreement
  - Correlation $r$
  - The Respiratory Care Abstract examined percentages

- Kappa - compares 2 or more raters
  - Assesses agreement beyond that expected by chance

- What is the maximum potential agreement between raters??
<table>
<thead>
<tr>
<th>Kappa Value</th>
<th>Agreement &gt; Chance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>None</td>
</tr>
<tr>
<td>0.0 – 0.20</td>
<td>Slight</td>
</tr>
<tr>
<td>0.21 – 0.40</td>
<td>Fair</td>
</tr>
<tr>
<td>0.41 – 0.60</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.61 – 0.80</td>
<td>Substantial</td>
</tr>
<tr>
<td>0.81 – 1.00</td>
<td>Almost Perfect</td>
</tr>
</tbody>
</table>
More Than 2 Raters?

- Multiple 2 x 2 Tables
- Fleiss’ Kappa
  - SPSS
  - SAS (Kappa.SAS macro)
  - R ([http://www.r-project.org/](http://www.r-project.org/))
  - MedCalc

- Easy, Quick & Free!!
  - [http://justusrandolph.net/kappa/](http://justusrandolph.net/kappa/)
# Kappa Calculator

**NO of Cases:** 16  
**NO of Categories:** 6  
**NO of Raters:** 4

Enter the number of raters who agreed that a certain case belonged in a certain category in the cells below.

<table>
<thead>
<tr>
<th>Case</th>
<th>Category 0</th>
<th>Category 1</th>
<th>Category 2</th>
<th>Category 3</th>
<th>Category 4</th>
<th>Category 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>i</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>j</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>k</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>l</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>m</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>n</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>o</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>p</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

**Calculate**

**Percent of overall agreement Po:** 0.656250  
**Fixed-marginal kappa:** 0.180443  
**Free-marginal kappa:** 0.587500
“Torture numbers and they’ll confess to anything”

Gregg Esterbrook
Now What?

- Kappa statistics can be used to report the outcomes of evaluation agreement
- Quantifies how close the group is to acceptable agreement
- Train --- Collect Data -- Run Statistics -- Retrain
Keep Records

- Build your database of preceptor information
  - Name & credentials
  - Qualifications
  - Date of Initial Training
  - Dates of Training Updates
  - Statistical Analysis of Consistency & Reliability of Evaluation
Questions to Ponder

- Are your faculty & preceptors on the same page?

- Are expectations clearly communicated to your students/faculty/preceptors?

- Is your evaluation system working?

- Who will crunch the numbers?
The Treasure at the End of the Rainbow

I was expecting something better