



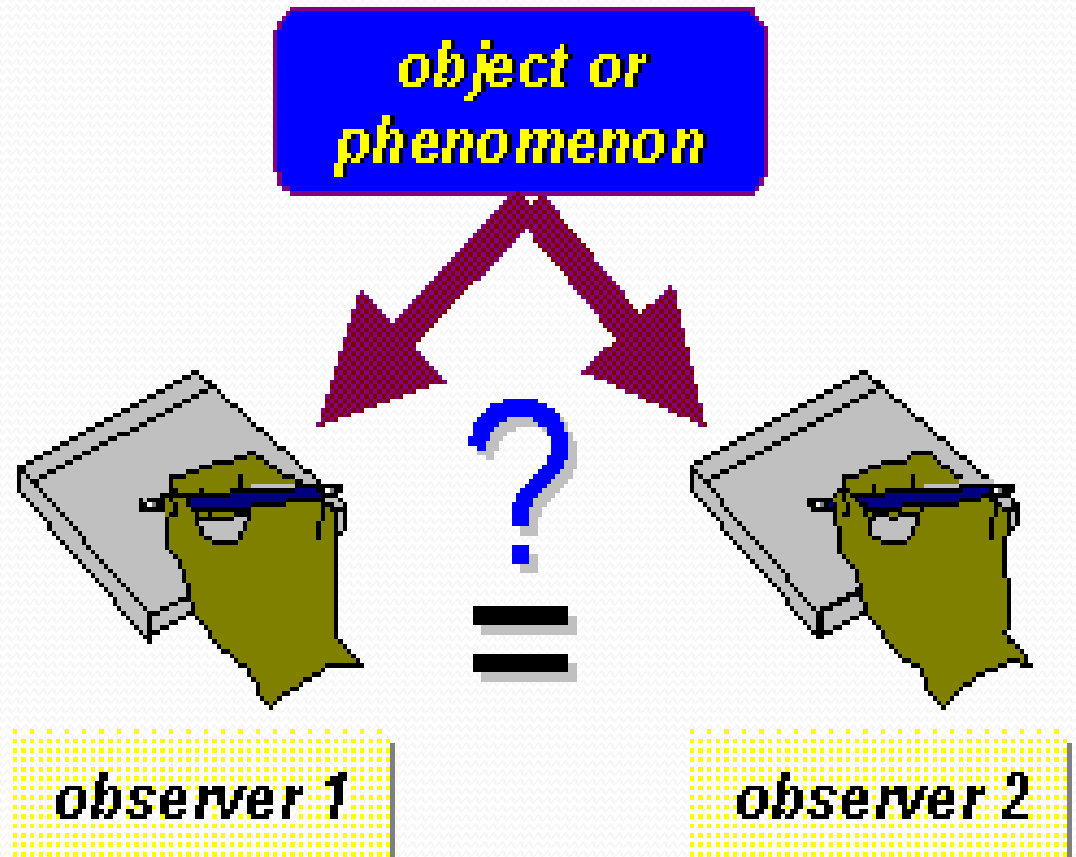
Commission on Accreditation for Respiratory Care

Incorporating Inter-Rater Reliability into Your Curriculum

**Kathy J. Rye, EdD, RRT, FAARC
Commissioner**

Objectives

- What?
- Why?
- When?
- How?



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 their 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 purpose 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

The Reality of Using Multiple Evaluators

- Differences 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.”

An Example:

2 or more RT instructors and students are discussing chest x-ray interpretation as a group.

- Individual students are required to perform a x-ray interpretation individually to certify competency.
- Instructors have a Sliding Rate Scale to rate oral responses:
 - 1 being most positive / 5 being most negative
- IRR assesses the consistency of how the rating system is used if all students are not being evaluated by the same instructor.

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:
 - I. Consensus Estimates
 - II. Consistency Estimates
 - III. Measurement Estimates

How can I ensure IRR?

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

Inter-Rater Reliability Exercise

- Review of Competency Evaluation Instrument
 - Discussion of how to use the instrument
- Evaluate Video Demonstration
- Review Competency Assessments
 - Examine differences
 - Establish consensus

Agree when good is “good enough”.

Competency Evaluation Instrument

- Review of task / skill list
 - Which of the steps are critical?
 - Does it make sense?
 - Anything missing?
 - Rate student performance?
 - **Pass (Satisfactory demonstration)**
 - **Fail (Unsatisfactory demonstration)**
 - ***Not Applicable (Task did not fit the case)***
 - ***Not Observed (Omission error)***

Video Demonstration

- Clinical Instructor Primer: Performance Evaluations by Rusty Taylor, M.Ed., R.R.T., Director of Clinical Education; Respiratory Therapy Program; Washburn University
- <http://www.washburn.edu/sas/ah/rt/clinical/CIP/>

Evaluation of Video Demonstrations

- Preceptors view streaming video presentations and submit their evaluations.
- Evaluations are assessed for IRR
 - Consistent?
 - If raters are fairly consistent on the evaluations, nothing more is needed.

Evaluation of Video Demonstrations

- Are raters inconsistent?
 - Was it due to being given poor instructions?
 - Was it due to something observed (or NOT observed) in the video presentation?
 - Are there any false assumptions made from the video?
- Retrain the raters.
- Revise the video presentation.

“The Bottom Line”

- Did the student perform the procedure well enough to not need direct supervision?
- Yes = Satisfactory
- No → Why?
 - Minor Unsatisfactory (little details)
 - Major Unsatisfactory (**BIG** details)

What Did You See?

- Satisfactory or Unsatisfactory?
- Examine Differences
- Establish Consensus

We have to agree on
“Good Enough”.

How Close Are We?

- Percent Agreement – Correlation r
- Kappa – compares 2 raters
 - Assess agreement beyond that expected by chance
- What is the maximum potential agreement between 2 raters?

Kappa

Observed agree – Expected chance
100% - Expected chance agreement

Observed agreement beyond chance
Max possible agreement beyond chance

Interpret Kappa

Kappa value	Agreement > Chance*
0	None
0.0 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost perfect

*Landis and Koch Biometrics 1977; 33: 159-74

More than 2 Raters?

- Multiple 2 x 2 tables
- Fleiss' Kappa
 - SPSS
 - SAS (Kappa.SAS macro)
 - R (<http://www.r-project.org/>)

Now What?

- Kappa statistic can be used to report the outcomes of evaluation agreement.
- Quantifies how close the group is to acceptable agreement.
- Collect evaluations & run statistics at a later date.
- Not necessary to do this step during the preceptor training session.

Record Keeping System

- Build your database of preceptor information
 - Name & Credentials
 - Qualifications
 - Date of Initial Training
 - Dates of Training Updates
 - Statistical analysis of consistency & reliability of evaluation.

Do You See What I See?

- Can you get your faculty and preceptors on the same page?
- Can you clearly communicate your expectations to your students? Your faculty? Your preceptors?
- Can you discern when your evaluation system is failing?
- Can you find someone to crunch your statistics?

Summary

- Without IRR, our hope for achieving objective measurements of student competency is diminished.
- Raters **MUST** receive training using the rubrics or instruments they will be expected to use!
- Raters should agree on the occurrence/ non-occurrence of key behaviors.
- Agreement ensures that evaluation of desirable behaviors remain consistent.
- Single observers can develop bias or become less stringent over time.

References

- Stemler SE. (2007) Comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Yale University PACE Center.
- Gallagher AG, Ritter EM, & Satava RM. (2003) Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surgical Endoscopy* 17 (10)1525-29.

Additional References of Interest

- *Shrout, P.E. & Fleiss, J.L. (1979) Intraclass Correlations: Uses in Assessing Rater Reliability. Psychological Bulletin, 2, 420-428.*
- *Field, A. P. (2005) Intraclass correlation. In Everitt, B. S. & Howell, D.C. Encyclopedia of Statistics in Behavioral Sciences . Chichester, England; Wiley.*
- *McGraw, K. O. & Wong, S. P. (1996. Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1, 30 - 46.*

Questions?

RyeKathyJ@uams.edu

OR

CoARC Executive Office

1248 Harwood Road

Bedford, TX 76021

(817) 283-2835 ext 101