Models of Successful Inter-Rater Reliability Programs

Kathy J. Rye, EdD, RRT, FAARC
CoARC Commissioner
Conflict of Interest

I have no real or perceived conflict of interest that relates to this presentation. Any use of brand names is not in any way meant to be an endorsement of a specific product, but to merely illustrate a point of emphasis.
Objectives

Learning objectives for this presentation:

• Review models of IRR with an emphasis on meeting accreditation requirements;
• Discover key components of successful IRR plans;
• Identify obstacles to successfully implementing the plan;
• Discuss methods for improvement;
• Questions & Answers.
Standard III – Program Goals, Outcomes & Assessment

• 3.11 The program must develop processes that facilitate the development of inter-rater reliability among those individuals who perform student clinical evaluations.
  • Records of training participation by clinical evaluators;
  • Results of a review of student evaluations for the purpose of determining IRR.
Original Interpretive Guideline for Standard 3.11

The intent of this standard is that the program includes inter-rater reliability measures in its clinical evaluation methods. These measures should allow the opportunity to compare assessments of students done by multiple clinical instructors to determine if there are significant differences. When significant differences do occur, the program should have a plan of action for addressing these differences (e.g., evaluations of clinical faculty conducted by Director of Clinical Education, program procedures established to address issues for contesting assignment of a clinical grade, etc.)
Keep Your Eyes on the Big Picture!!!!

Credentialing Success
Attrition
Job Placement
Employer Satisfaction
Graduate Satisfaction
Keep Your Eyes on the Big Picture!!!

Don’t get hung up on the little things --
Like STATISTICS!!!!!!!!!!!!!!!!!!!!!!!!!!!
The intent of this standard is to ensure consistency in the clinical evaluation of the students. The program must demonstrate that clinical evaluators are provided with training (e.g. training manual, training workshop, or online training sessions for evaluators). This process must include a comparison of student evaluations completed by clinical instructors in order to identify variability among evaluators. **Statistical analysis can be used but is not required.** When variability is identified, the program must have a plan of action which includes remediation, timeline, and follow-up. The results of this process must be reviewed by the Director of Clinical Education or Program Director at least annually.
Goal: Consistency in Evaluation of Students

3 Salient Points

• Training
• Comparison of student evaluations for variability
• Implementation of an action plan to address any identified variability(s).
5 Key Components of the Successful IRR Plan

- Clinical Instructor/Preceptor Training;
- Valid / Reliable Evaluation Tools:
  - Skills competency evaluations,
  - Affective evaluations, &
  - Quality assurance.
- Annual comparison of clinical evaluations by raters for IRR;
- Remediation of raters as needed;
- Timeline for follow-up.
IRR Model Component #1

Clinical Instructor/Preceptor Training

- Training manual, workshop, &/or online sessions;
- Roster of training session attendees;
- Dates of initial training sessions;
- Updates to training annually or as needed to address changes in evaluation system.
Clinical Instructor / Preceptor Training

Possible Topics for Consideration

- Role of the Clinical Preceptor
- Adult Learning Styles
- Clinical Teaching Strategies
- Providing Effective Feedback & Evaluation
- Using Direct Observation Skills
- Dealing With Difficult Students
- The Preceptor As A Mentor
IRR Model Component #2:

• Clinical skills competency evaluation tool
  • Easily accessible to CI / precept
  • Includes a detailed list of all steps required for certification of competency.

• Clinical Policy stating what steps MUST be taken if the student is unable to demonstrate competency.
IRR Model Component #2:

• **Affective evaluations** that are easily accessible and include clear criteria with a narrative description of the behavioral performance desired.
• **Clinical Policy** stating what steps MUST be taken if the student is unable to demonstrate appropriate affective behaviors / attitudes.
Quality Assurance - Possible Considerations:

- Require date & signature of clinical instructor/preceptor on all evaluations.
- Require date & student signature on all evaluations.
- DCE periodically reviews all evaluations:
  - Ensure timely submission of evaluations
  - Document any follow-up or discussion with regard to areas of concern on student evaluations.
IRR Model Component #3:

Comparison of raters for IRR purposes:

• Annually at a minimum!
• Consensus estimates are acceptable.
  • 70% or greater consensus
  • + or – 1 Standard Deviation
• Kappa statistics can be done but ARE NOT REQUIRED!!!!!!!!!!!!!!!
## 7 Evaluators of 1 Student’s Affective Skills

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appearance</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Attendance</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Prepared/On Time</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Reliable</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Team Player</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Positive</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Accepts Guidance</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Interaction</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Professional</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Communication</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Time Mgmt</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Responsible</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Confident</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Scale:**
- 5: Exceptional
- 4: Above Average
- 3: Acceptable
- 2: Below Average
- 1: Unacceptable
- 0: Not Observed
<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appearance</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.7</td>
<td>0.57</td>
</tr>
<tr>
<td>Attendance</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.9</td>
<td>0.41</td>
</tr>
<tr>
<td>Prepared/On Time</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5.0</td>
<td>0.00</td>
</tr>
<tr>
<td>Reliable</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.7</td>
<td>0.57</td>
</tr>
<tr>
<td>Team Player</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.4</td>
<td>0.57</td>
</tr>
<tr>
<td>Positive</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.4</td>
<td>0.57</td>
</tr>
<tr>
<td>Accepts Guidance</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.7</td>
<td>0.57</td>
</tr>
<tr>
<td>Interaction</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.7</td>
<td>0.57</td>
</tr>
<tr>
<td>Professional</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.6</td>
<td>0.57</td>
</tr>
<tr>
<td>Communication</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4.3</td>
<td>0.82</td>
</tr>
<tr>
<td>Time Mgmt</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4.3</td>
<td>0.82</td>
</tr>
<tr>
<td>Responsible</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.4</td>
<td>0.57</td>
</tr>
<tr>
<td>Confident</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.4</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Scale:
5: Exceptional
4: Above Avg
3: Acceptable
2: Below Avg
1: Unacceptable
0: Not Observed
IRR Model Component #3:

Comparison of raters for IRR purposes:

• Kappa statistics can be done but ARE NOT REQUIRED!!!!!!!!!!!!!!!

• Online Resource for Kappa
  • www.justusrandolph.net
  • “Makes it a piece of cake.”
    Tim Kelly, Metropolitan Community College, Omaha, NE

$$K = \frac{P_{observed} - P_{chance}}{1 - P_{chance}}$$
IRR Model Component #4:

Remediation of raters as needed to improve consistency:

• Based on DCE’s analysis of the evaluations that were completed by raters.
• Based on the collective results of the students’ anonymous evaluations of the individual preceptor/clinical instructor’s instruction.
IRR Component #5:

Develop a strategy & time line for follow up:

- What data should be reviewed to effectively reinforce or re-evaluate IRR
- Is there a need to re-educate raters?
- Is there a communication issue?
- When/how should the particular issue be reexamined?
- Is further action required?
Will I Have Obstacles?

Begin by making an assessment:

• What these obstacles typically are,
• Where they come from, and
• How to remove or work around them.
## Possible Obstacles

### Internal Obstacles
- Lack of time;
- Lack of awareness;
- Attitudes about change;
- Entrenched ways;
- Unclear direction;
- Financial implications.

### External Obstacles
- Clinical Instructors;
- Preceptors;
- Affiliate Leaders;
- Market;
- Competitors.
In Conclusion

- Know the Standards!
- Get Organized!
- Develop Your IRR Plan!
  - Evaluator training
  - Annual comparison of evaluation
  - Remediation, timeline & follow-up as needed.
- You are not alone!
- Document! Document! Document!
- “Your clinical program is as delicate as an orchid and just as beautiful if you nurture it properly.”