



Building a Certified Peer Observation Program: *Improving Teaching and Removing Biases*

Rebecca Johnston, Associate Director

Roger Runquist, Director

Carl Ohrenberg, Assistant Director

Lindsay Linsky, Assistant Director

Center for Teaching, Learning, and
Leadership



PollEv.com/catherinelin547

**“Your class is going to be formally observed today.”
What is your reaction?**

Purpose

- Examine Student Evaluations of Teaching (SETs) in extant literature
- Discuss benefits of peer observation
- Explain genesis of program development
- Present elements of the program
- Provide opportunity for questions and discussion

Activity:

<https://padlet.com/cllinsky/9v0arb4i8lqkalm8>

**List something strange, unusual,
or just plain crazy that you've
seen or heard of on *any*
student evaluation.**

Student Evaluations of Teaching

- SETs primary source of data to evaluate teaching, merit raises, and promotion/tenure
- Use of SETs as primary high-stakes measure is problematic
 - Issues include invalidity, unreliability, bias, and small sample size
- Despite this, SETs remain the primary tool

(Lince, 2017; Boring, Ottoboni, & Stark, 2016; Spooren, Brockx, & Mortelmans, 2013; Kelly, 2012; Storage, Horne, Cimpian, & Leslie, 2016; MacNell, Driscoll, & Hunt, 2015; Hornstein, 2017; Braga, Paccagnella, & Pellizzari, 2014)

Problems with SETs

- Do not legitimately assess *teaching effectiveness*
 - Measure *student opinions* of teaching effectiveness
 - Students are *not qualified* to assess teaching effectiveness
 - SETs gather collective views of student experience in a *single course* with a particular faculty
 - Not a global evaluation

(Hornstein, 2017; Arreola, 2007; Hativa, 2013; Linse, 2017)

Problems with SETs

- Many behaviors and skills are involved in teaching effectiveness
 - Knowledge and content expertise
 - Teaching methods
 - Course design and organization
 - Quality of course materials
 - Assessment instruments and methods
 - Grading practices
 - *Students not qualified to assess (not trained in course design and instruction)*

(Marsh, 2007; Svinicki & McKeachie, 2011; Berk, 2013)

Problems with SETs

- Female instructors evaluated more critically
 - Gender bias accounts for .50 on a 5-point scale
 - Female instructors evaluated more critically even when gender randomly assigned
 - Bias influences even objective measures, i.e. time to return work
- (Macnell, Driscoll and Hunt, 2015)

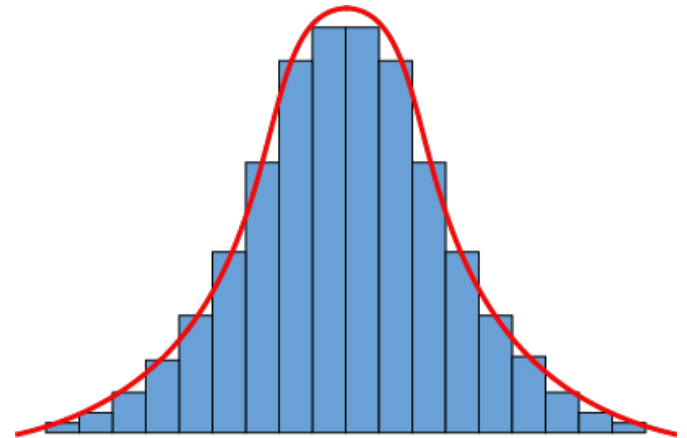
Problems with SETs

- Other documented biases:
 - Faculty rank
 - Student motivation
 - If course is required or elective
 - Anticipated grade
 - Upper vs. lower division
 - Class size
 - Academic discipline
 - Student workload

(Braskamp & Ory, 1994; Marsh & Dunkin, 1992, 1997; Centra, 2003, 2009; Hoyt & Lee, 2002a, 2002b)

Problems with SETs

- Valid evaluations cannot be made using parametric statistics
 - Data is categorical
 - Parametric analyses make assumption of symmetry in distribution not reflected in SETs samples



Problems with SETs

- Possible to have instructors very effective at teaching, but very low scores on SETs
- The reverse is also true

(Steury, et al., Auburn University)

Problems with SETs

- Use of SETs for faculty hiring, promotions, merit increase
 - Can encourage:
 - poor teaching
 - result in grade inflation
 - empower students to shape faculty behavior
- Higher ed. continues to use SETs regardless of problems

(Hamermesh & Parker, 2005)

Studies in Support

- SETs positively correlated with teaching effectiveness and student learning
 - Useful for measuring *in aggregate*
 - To compare multiple individuals across departments or programs
 - Do not distinguish among individual teachers, especially through a single class

Studies in Support

- Scores from multiple courses taught over multiple years
 - May give some indication of individual teacher effectiveness
 - Data over time
 - Trends

History of SETs

- In the 1970s, used primarily for *formative* assessment
 - How can teaching be improved?
- Since then, have become primary means of *summative* assessment
 - Single high-stakes summations
 - *Absolute* measures of quality of teaching

(Hornstein, 2017)

Recommendations

- SETs are poor indicator of individual instructor's overall effectiveness
- Should be used formatively
- Ratings should be evaluated from multiple courses across time
- If poor participation rate, should not be used
- Should be used alongside *other* forms of evaluation

Why are They Still Primary Measures?

- Easily quantifiable data
- Easily repeatable
- What are alternate forms of evaluation?

Alternate Forms of Evaluation

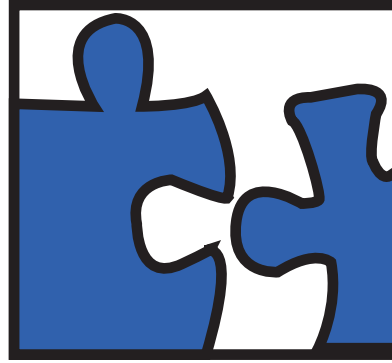
- Peer review of course material
- Peer review of instruction
- Review by expert outside evaluators
 - CTL
- Teaching scholarship
- Learning outcome measures
- Teaching portfolios

Benefits of Peer Observation

- Creates an opportunity for reflection
 - For both faculty and CPO
- Provides a data point to balance SETs
- Can be used to document teaching effectiveness
- Formative focus on faculty growth
- Trained observers focused on analysis of teaching behaviors

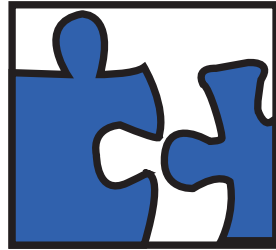
Program Development

- Our CTL has often offered to do an observation but there has been no formalized process
- A presentation by St. Leo at SoTL Commons offered insights into an effective program
- Faculty interest at UNG to have a counterpoint for SETs

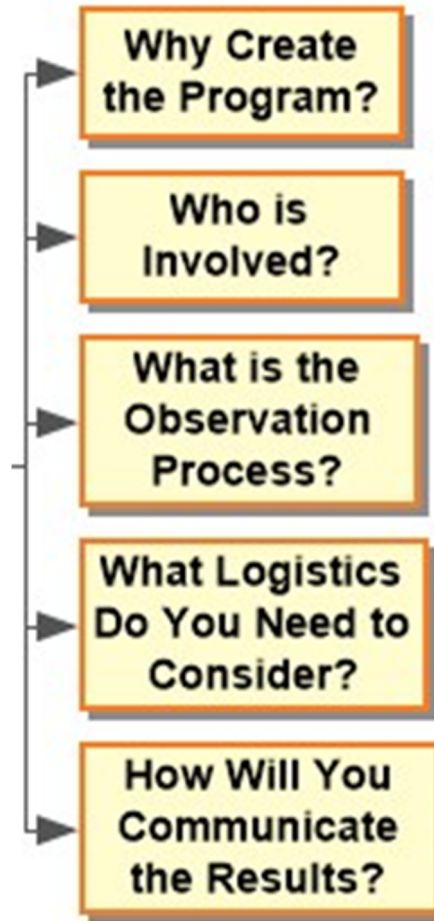


Developing a Peer Assessment Program

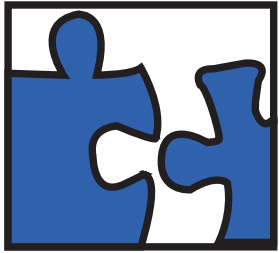
CPO Development Process



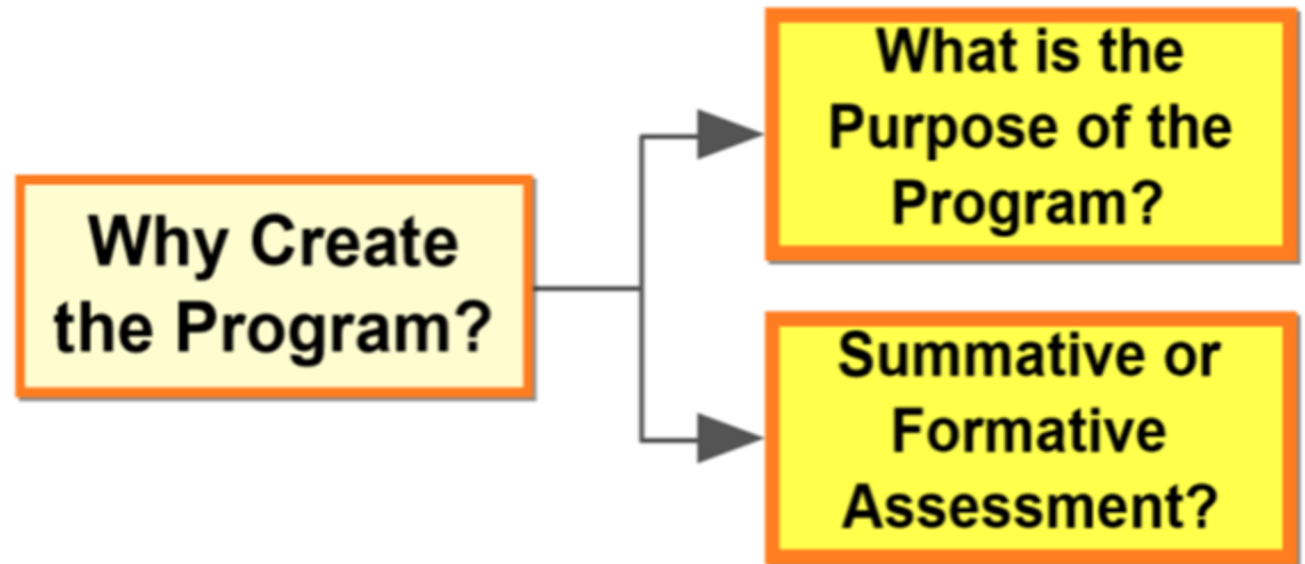
Developing a Peer Assessment Program



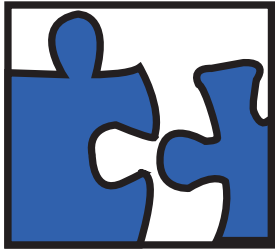
CPO Development Process



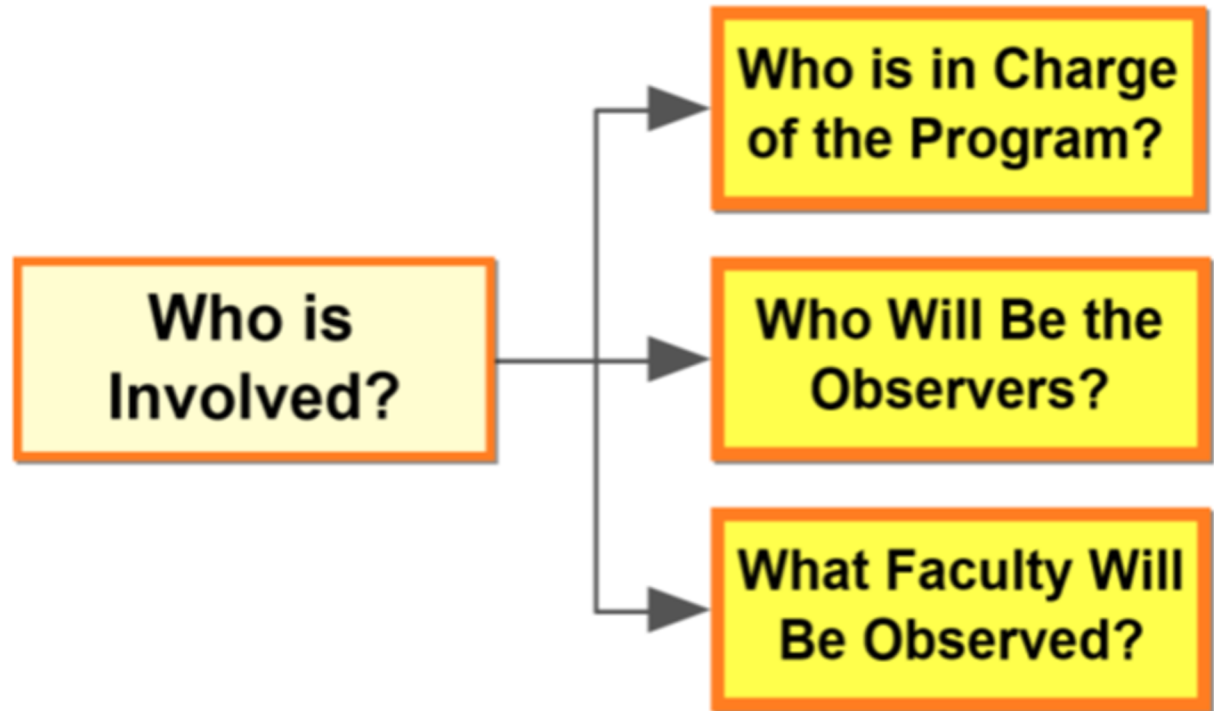
**Developing a
Peer
Assessment
Program**



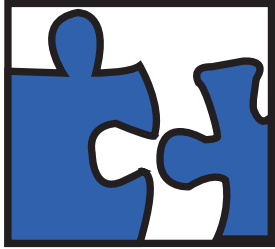
CPO Development Process



**Developing a
Peer
Assessment
Program**



CPO Development Process



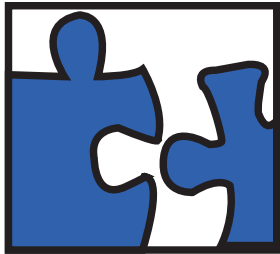
**Developing a
Peer
Assessment
Program**

**What is the
Observation
Process?**

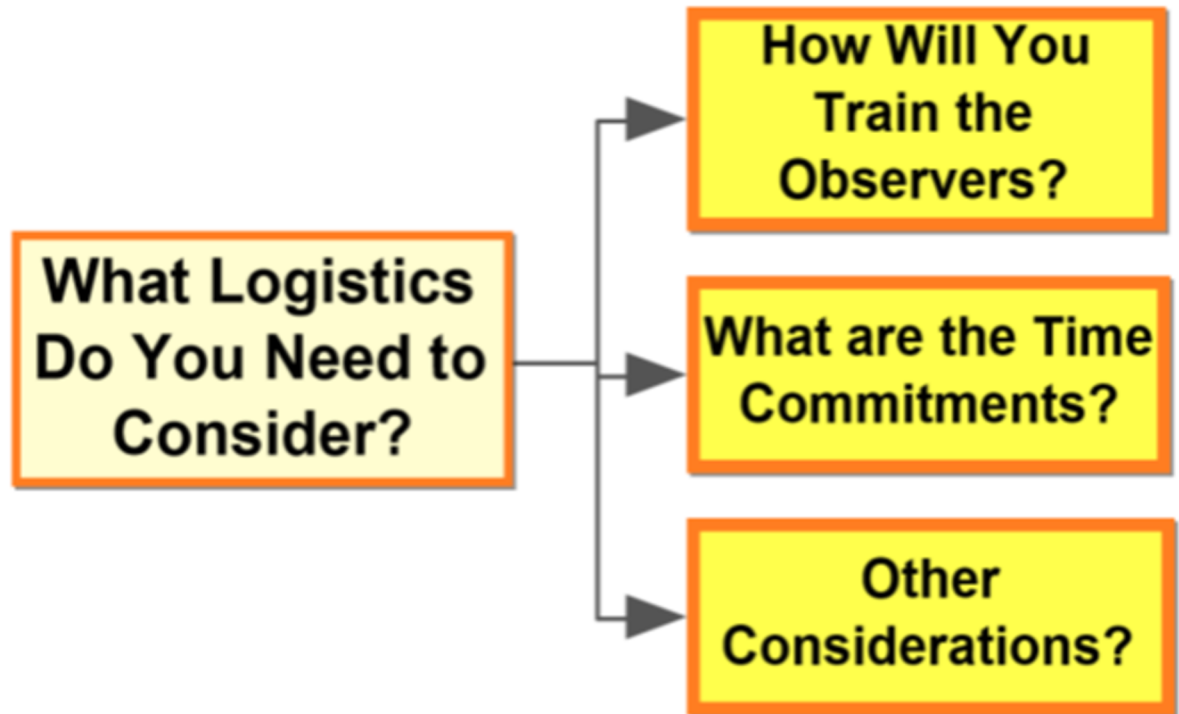
**How and When
Will the Observer
Meet With the
Faculty?**

**What Evaluation
Tool Will Be
Used?**

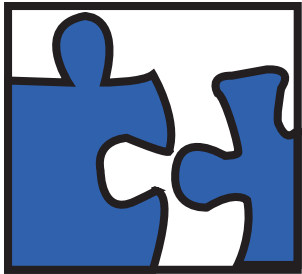
CPO Development Process



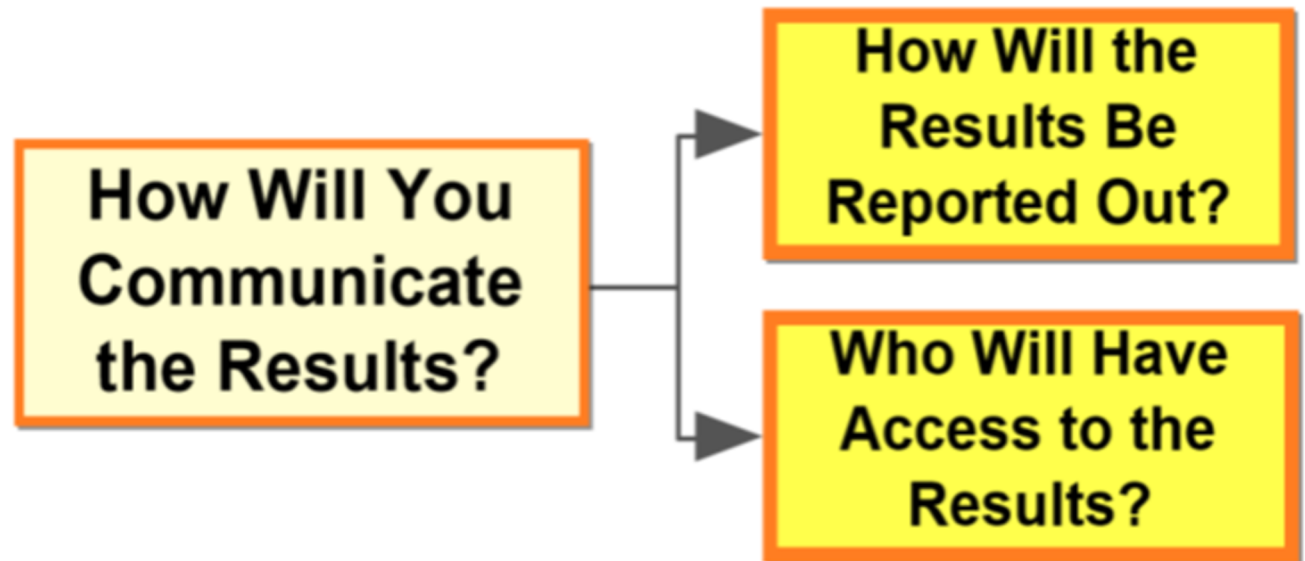
**Developing a
Peer
Assessment
Program**



CPO Development Process



**Developing a
Peer
Assessment
Program**

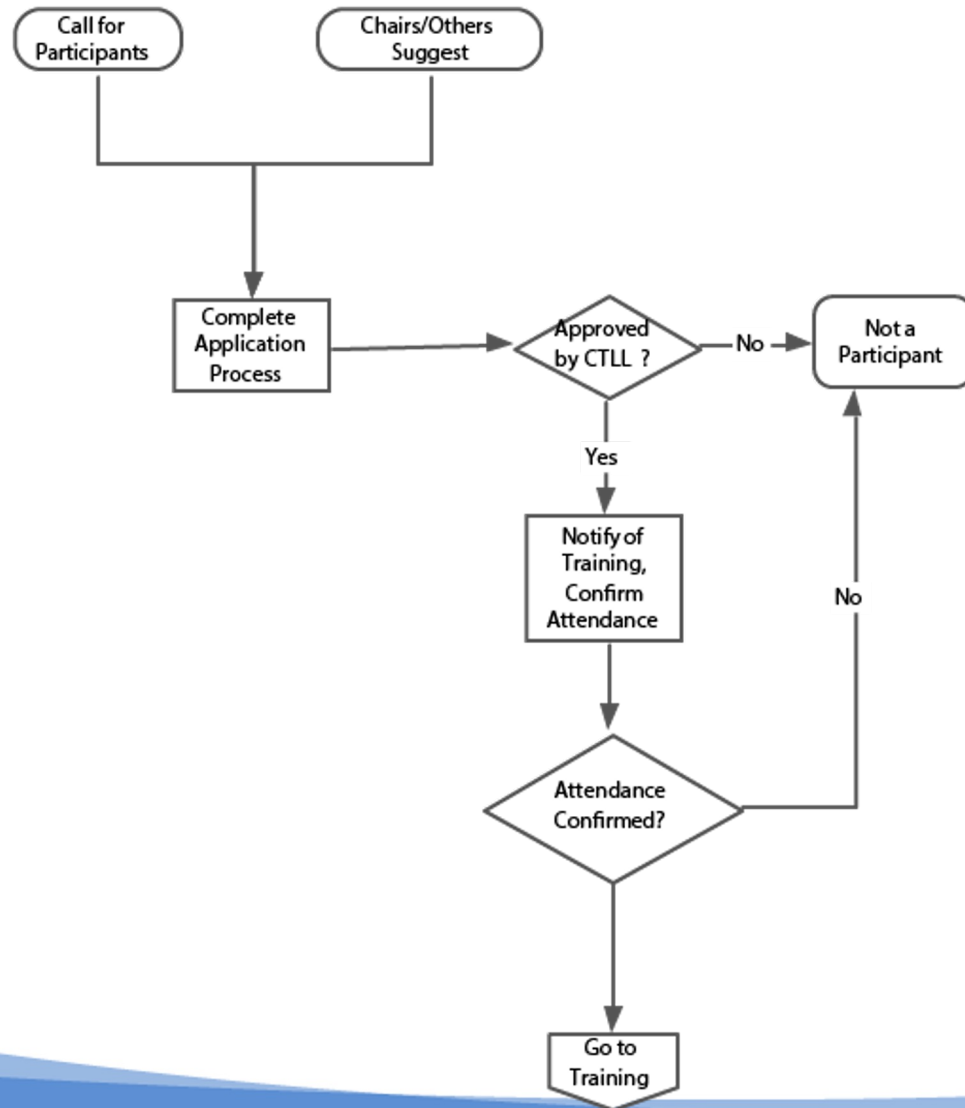


Our Process

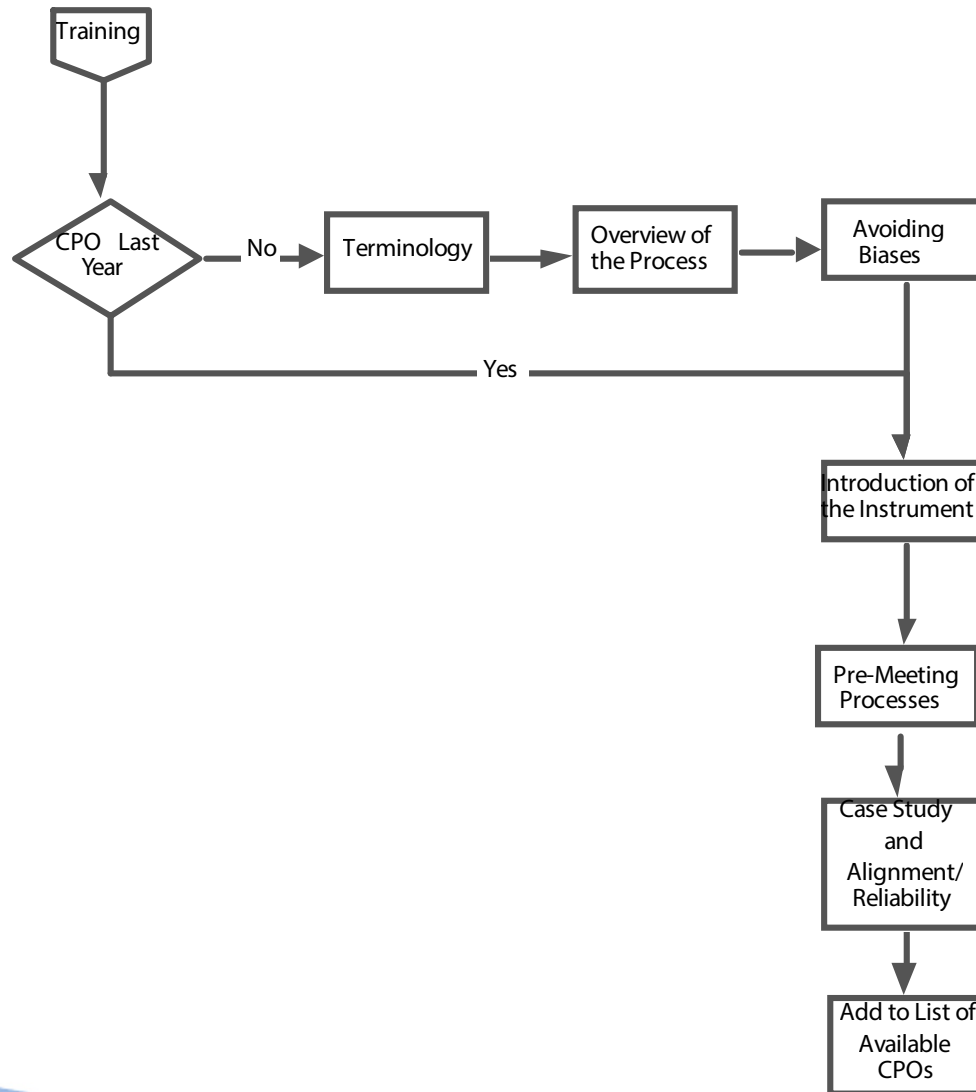
Factors for our program:

- Faculty development opportunity
 - Enhanced by self-awareness
- Flexibility to use the review is up to the faculty member
 - Remember, it is formative

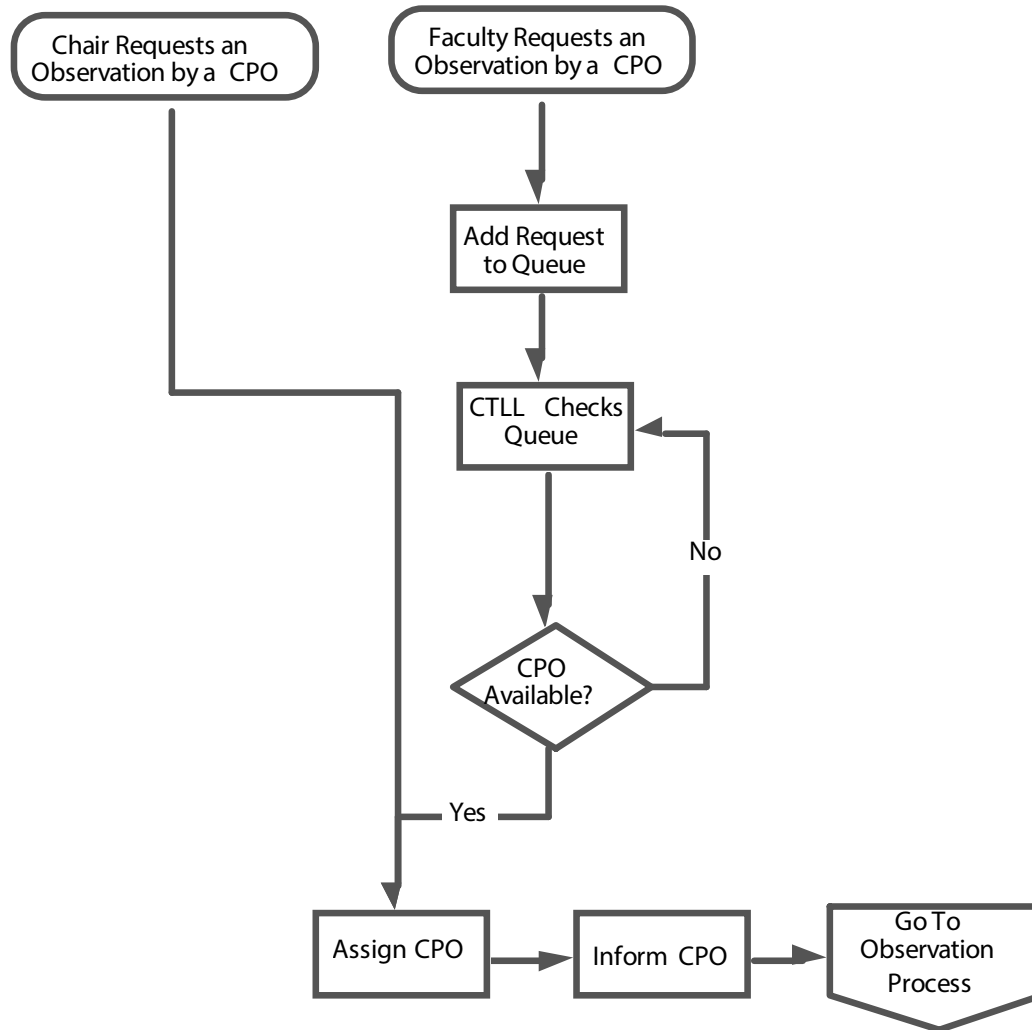
CPO Recruitment



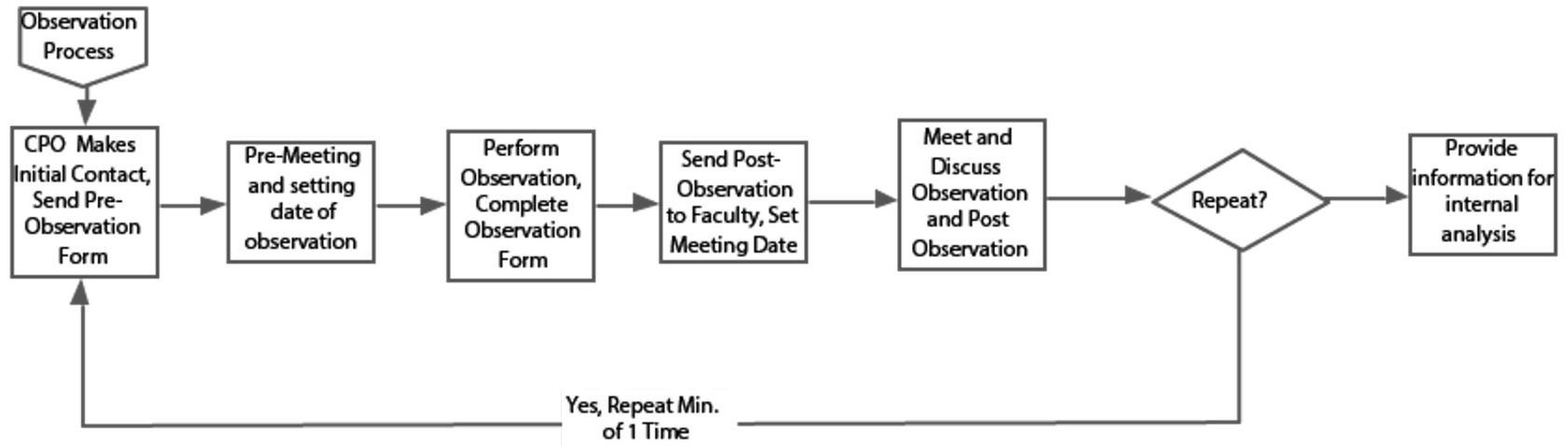
CPO Training



CPO Requesting



Observation



Activity:

[https://padlet.com/cllinsky/4e9rm
t4azm43bluq](https://padlet.com/cllinsky/4e9rm
t4azm43bluq)

Other than SETs what experiences have helped
the most to improve your teaching?

Discussion

- Do you currently have a peer observation program at your institution? If so, please share!
- If you do not have a CPO program, which entities at your institution could help to develop one?
- Could you develop a CPO program within your department? What might be some barriers and/or solutions?

References

- Arreola, R.A. (2007). *Developing a comprehensive faculty evaluation system: a guide to designing, building, and operating large-scale faculty evaluation systems*. Anker Publishing, Boston, MA.
- Berk, R.A. (2013). *Top 10 flashpoints in students ratings and the evaluation of teaching: What faculty administrators must know to protect themselves in employment decisions*. Stylus, Sterling, VA.
- Boring, A., Ottoboni, K., & Stark, P.B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*.
- Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, 41, 71-88.
- Braskamp, L.A., & Ory, J.C. (1994). *Assessing faculty work: Enhancing individual and institutional performance*. Jossey-Bass, San Francisco, CA.
- Centra, J.A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, 44, 495-518.

References

Centra, J.A. (2009). Differences in response to the student instructional report: Is it bias? Educational Testing Service, Princeton, NJ.

Hamermesh, D., & Parker, A. (2005). Beauty in the classroom: Professors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, 24(4), 369-376.

Hativa, N. (2013). *Student ratings of instruction: a practical approach to designing, operating, and reporting*. Oron Publications.

Hornstein, H.A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1), Article 1304016 2017

Hoyt, D.P., & Lee, E. (2002a). *Basic data for the revised IDEA system* (IDEA Paper No. 12). Manhattan, KS: The IDEA Center at Kansas State University, Center for Faculty Education and Development.

Hoyt, D.P., & Lee, E. (2002b). *Disciplinary differences in student ratings* (IDEA Paper No. 13). Manhattan, KS: The IDEA Center at Kansas State University, Center for Faculty Education and Development.

References

Kelly, M. (2012). Student evaluation of teaching effectiveness: Considerations for Ontario University. *COU Academic Discussion Paper #866*, pp.1-15.

Linse, A.R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, 94-106.

Macnell, L., Driscoll, A., & Hunt, A. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40, 291-303.

Marsh, H.W. *Students' evaluations of university teaching: A multidimensional perspective*. In R.P. Perry, J.C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective*, Springer, New York NY (2007), pp. 319-384.

Marsh, H.W., & Duncan, M.J. *Students' evaluations of university teaching: A multidimensional perspective*. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research* (V8). Agathon Press, New York NY (1992).

References

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598-642.

Storage, D., Horne, Z., Cimpian, A., & Leslie, S-J. (2016). The frequency of “brilliant” and “genius” in teaching evaluations predicts the representation of women and African Americans across fields. *PLoS ONE*, 11(3), Article e0150194.

Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, 42, 1-19.

Svinicki, M., & McKeachie, W.J. (2011). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers*. Wadsworth, Belmont, CA.

Trujilo, J.M., Divall, M. V., Barr, J., Gonyeau, M., Van Amburgh, J. A., Matthews, J., Qualters, D. Development of a peer assessment program and a peer observation and evaluation tool. (2009). *American Journal of Pharmaceutical education*, 72(6), Article 147