Governors State University

1 University Parkway
University Park, IL 60484
708.534.5000
www.govst.edu

Continuous Improvement Communities of Practice
Inter-Rater Reliability Process

The faculty and staff of Governors State University's (GSU) Educator Preparation Provider (EPP) continuously work to improve teacher preparation. By examining its current practices, collecting and analyzing data, and having rich discussions with stakeholders, the EPP can systematically implement changes. Through this process, the EPP identified the need to improve the implementation of the Charlotte Danielson Framework rubric used to observe and assess teacher candidates during clinical and student teaching experiences. Two Danielson Framework rubric inter-rater reliability exercises were conducted to assess the effectiveness of the implementation of the rubric. The results from the first exercise, conducted within the EPP, did not meet the desired outcome of 75% or greater agreement. A more extensive, collaborative exercise that included EPP and district teachers led to improved results. These results, when compared to an expert scorer, revealed the need to both continue and expand the efforts to improve the accuracy in pre-service teacher evaluations.

Inter-Rater Reliability Exercise #1:

On May 9, 2018, thirty-five GSU's EPP faculty and staff participated in an inter-rater reliability exercise to ensure the accuracy, consistency, and precision of the implementation of the Student Teaching Danielson Rubric.

In advance of the meeting, participants were asked to review a summary of Charlotte Danielson's "*Step-by-Step- Approach"* for using the Danielson Framework. In addition, participants were also asked to review and assess a 13-minute video of a middle school math lesson using the EPP Student Teaching Danielson Rubric - Domains 2 and 3. It should be noted, that the EPP Student Teaching Danielson Rubric is a slight modification of the Danielson Framework to better align the expected performance levels to those of a student teacher rather than an experienced teacher.

During the meeting, participants watched a math lesson video together and tallied their respective scores. Using the inter-rater reliability formula, the outcome resulted in a 58% agreement. After the initial scoring, the team had an in-depth group discussion regarding the individual interpretation of the tool and lesson. The participants completed a second scoring resulting from the discussion which led to an overall inter-rater reliability agreement of 64%.

The overall inter-rater reliability agreement was below the desired 75% goal; however, the agreement increased by six percentage points after the group discussion. As a result of the findings (see separate report), the following strategies were recommended:

- Conduct a separate activity to revise the assessment tool
- Repeat the exercise using a full-length lesson
- Repeat the exercise using more participants, including supervising teachers
- Take additional time for discussion
- Rate the second round one domain at a time together through threaded discussion
- Allow more time for a similar exercise

Collaborative Transformation Background

In Spring 2018, a Memorandum of Understanding (MOU) was developed between GSU's EPP as an institution of higher education and Crete-Monee High School (CMHS), a local high-needs high school. The MOU was developed to promote simultaneous renewal with benefits to both GSU's EPP and CMHS through a clinical immersion experience for secondary English candidates. This partnership would allow the candidates to complete an entire year of clinical experience at the same school with content mentor teachers.

An initiative was developed and sponsored by the Illinois State Board of Education (ISBE) and Branch Alliance established Continuous Improvement Communities of Practice (CICP). Through this initiative, GSU's EPP and CMHS were able to catalyze collaborative learning and leadership around key issues of teacher recruitment, preparation, and retention in high-need subjects and schools. Specifically, the GSU – CMHS community partnership focuses on the clinical immersion experience as a preparation pathway to student teaching and placement.

Through a CICP strategic process, GSU's EPP/CMHS discovered inconsistencies in the implementation of the Danielson Rubric, which is an integral component of preparing candidates during the clinical immersion experience. Three major discoveries were made: (1) The EPP uses a modified version of the Danielson Rubric to closer align with student teaching expectations; CMHS examines the candidate with teacher expectations; (2) the EPP focuses on Danielson Domain 1 for clinical experience; CMHS focuses more on Domains 2 and 3 during teaching observations, and (3) the mentor teachers responsible for providing feedback to candidates had not been trained to use the Danielson rubric.

Governors State UNIVERSITY

1 University Parkway
University Park, IL 60484
708.534.5000
www.govst.edu

Inter-Rater Reliability Collaboration Exercise #2

On September 6, 2018, the GSU EPP faculty and staff along with partnering school CMHS participated in an inter-rater reliability collaborative exercise. This exercise was conducted to ensure the accuracy, consistency, and precision of implementing the Danielson Rubric Domains 2 and 3, which is used to assess clinical, student teaching, and teacher observation experiences.

The meeting began with the introductions of 34 individuals, consisting of 20 English and social science mentor teachers from CMHS, 2 administrators from CMHS, 9 GSU's EPP faculty/supervisors, 2 GSU's EPP administrators, and 1 GSU's EPP assessment coordinator. Of the 34 individuals who attended the gathering, only 29 participated in the actual scoring. In preparation for the exercise, the following were discussed: (1) the significance of using the same rubric; (2) the significance of using Domains 2 and 3 during both the clinical and student teaching experiences, and (3) instructions on the use of the original Danielson Framework observation tool (with no modifications). The video of a full-length 5th-grade math lesson was provided using the Talent Ed Calibration Module. This module is ISBE-approved for K-12 evaluator recertification.

Participants were given an opportunity to review the Danielson Rubric and acclimate to its design. The participants then watched the video of the lesson together and were subsequently separated into six different discussion groups comprised of combined GSU's EPP and CMHS faculty/staff. The use of a common assessment tool was instrumental in communicating how we define teaching excellence. During the small group discussions about the different elements within Domains 2 and 3, the participants shared their observation perspectives and interpretations, which influenced their individual scores. After the discussion, the participants submitted their individual scores. As the scores were being tallied, participants took this opportunity to share their perceptions about the exercise.

The overall inter-rater reliability agreement was 82% which was above the anticipated 75% goal.

Governors State UNIVERSITY

1 University Parkway
University Park, IL 60484
708.534.5000
www.govst.edu

Scores and Interpretation

GSU's EPP/CMHS Level of Agreement

Table 1:

2a	86%
2b	86%
2c	52%
2d	62%
2e	93%
Overall	**76%**

3a	77%
3b	79%
3c	90%
3d	97%
3e	97%
Overall	**88%**

Overall		**82%**

The individual scores of the participants were averaged and compared to the Expert Scored Video. The overall results of the expert scorer were 25 compared to the average GSU/CM collaborative group assessment of 30.24, showing a 5.25 difference.

Table 2

Scores Compared to Expert Scorer Agreement

Calibration Matrix			
Overall Adjacent Calibration Score (percent within 1 point - all scores combined)	100.00%		
Valence Distance Scores (positive # = points > master score, negative # = points < master score)	**-1**	**0**	**1**
Percent Average Valence Distance	0.38%	41.00%	58.62%
Percent Domain 2 Average Valence Distance	0.00%	33.62%	66.38%
Percent Domain 3 Average Valence Distance	0.69%	46.90%	52.41%
Absolute Distance from Master Scores (higher #, farther away from master score)		**0**	**1**
Percent Average Absolute Distance (all scores combined)		41.00%	59.00%
Percent Domain 2 Average Absolute Distance		33.62%	66.38%
Percent Domain 3 Average Absolute Distance		46.90%	53.10%
Percent Exact Agreement			
Percent Exact Agreement Scores (all scores combined)	41.00%		
Percent Exact Agreement Domain 2	33.62%		
Percent Exact Agreement Domain 3	46.90%		

Governors State UNIVERSITY

1 University Parkway
University Park, IL 60484
708.534.5000
www.govst.edu

The overall adjacent calibration score shows that all (100%) scores by raters were within one point of the master score, so all raters were in close agreement to the master scores. The valence distance scores show that approximately 59% of all scores were higher than the master scores, 41% were in agreement, and 0.38% (one score) was lower than the master scores. Because there was only one score lower than the master score and all scores were within one point of the master score, the absolute distance from master scores and percent exact agreement showed the same basic results as those of the valence distance scores. Overall, the raters performed better in rating Domain 3 than Domain 2. There was a 47% exact agreement with Domain 3, compared to a 34% exact agreement with Domain 2.

The overall trend is that the majority of raters assessed the teacher's performance higher than those of the master raters. Moving forward, we should discuss the tendency for teachers to submit higher ratings than master scorers. Additionally, in examining the raw data, components 2a, 2b, 3d, and 3e of Danielson represent scores for which there were the fewest exact agreements. We should pay particular attention to those components.

Expert Scores and Recommendations

In an effort to continuously and systematically improve practices to increase the quality of teacher preparation, the following considerations are offered:

1. Provide additional training for observers using Talent Ed/Teaching Channel Videos
2. Review and compare the expert's collection of evidence to the individual reviewer's collection of evidence.
3. Consider the EPP's use of Domains 2 and 3 for clinical experiences prior to student teaching.
4. Consider the EPP's use of the original, unmodified version of the Danielson rubric.
5. Repeat similar exercise on an annual basis.
6. Use training video of a lesson closer aligned to grade level and content.
7. Establish best practices of an effective observer.

Concluding Remarks

Both the university and the P-12 school benefitted from this exercise. The EPP, who is ultimately responsible for the assessment of clinical and student teaching experience, gained insight on P-12 needs to have teachers ready day one. Teachers as mentors not only experienced the observation training from the perspective of an observer, but also as a teacher.

Governors State UNIVERSITY

1 University Parkway
University Park, IL 60484
708.534.5000
www.govst.edu

During conversations about practices, particularly when such conversations are organized around a common framework, participants were able to learn from one another and to thereby enrich individual and collective practices. It is through this collaborative exercise and opportunities like this that the conversation becomes rich and valued. It is through collaborative, professional conversations about frameworks components that these components are validated for the use of clinical and student teaching experiences.

This inter-rater reliability collaborative exercise was a critical step to enriching the professional lives of educators and to ensure that the components used in a given setting actually do apply there. As an agreed-upon framework for excellence between the university and the P-12 partner, the Danielson observation instrument for teaching serves to structure conversations among educators about exemplary practice.

Future Steps

GSU's DOE will continue to pursue inter-rater reliability collaborative exercises both inside the institution with professors and staff and outside the institution with partner LEA's on an annual basis. It is obvious that this type of shared governance and mutual understanding of assessments utilized in teacher licensure by both the university and its cooperating partners are not only beneficial by vital to continuous improvement.

Participants

GSU's EPP

Amy Vujaklija, Glenna Howell, Pam Guimond, Lisa Pennington, Linda Ruhe-Marsh, Audrey Manley, Steven Sharp, David Conrad, Megan McCaffrey, Melinda Elliott, Tim Harrington, Joi Patterson.

CM – High School

Ms. Marcia Mikal, Mr. Mark O'Connor, Mr. Brian O'Donnell, Mr. Richard Posey ,Mr. Taurus Scurlock ,Mr. Dave Surlak ,Mr. Brook Swanson , Ms. Carly Benard, Ms. Jamie Booth, Ms. Michelle Cresto, Ms. Melanie Duffy,Mr. William Fitzgerald,Ms. Michelle Rivero,Ms. Denise Graney,Ms. Rebecca Johnson,Mr. Kevin Kenealy,Ms. Allison Lewandowski, Ms. Jennifer Santor,Mr. Robert Welch,Ms. Cynthia Hysell

Governors State UNIVERSITY

1 University Parkway
University Park, IL 60484
708.534.5000
www.govst.edu

Resources:

Charlotte Danielson – Enhancing Professional Practice – A Framework for Teaching; 2007; 2nd edition.

Talent Ed Calibration Module

Thanks to all the many individuals who have input in preparing our candidates and for participating in this collaborative exercise. Special Thanks to:

- CM faculty for mentoring pre-service teachers and participating in the inter-rater reliability training exercise
- GSU EPP faculty for preparing and supporting pre-service teachers and participating in the inter-rater reliability training exercise
- Amy Vujaklija for implementing a clinical immersion program at a high-need school and fully participating in the development of this collaboration
- John Konecki for facilitating the clinical immersion program and fully participating in the development of this collaboration
- Laura Hirsch for consenting to the partnership, fully participating in the development of this collaboration, and providing the resources necessary to conduct the clinical immersion program and inter-rate reliability training exercise
- Melinda Elliott for providing technical and data support for the CICP initiative and fully participating in the development of this collaboration