Intern Evaluation ς Mid-Experience: Pilot Data: Brief Report (Draft)

The following report briefly summarizes basic item-level, reliability

Reliability Evidence (Internal Consistency)

To provide a preliminary evaluation of reliability evidence for the intern evaluation, internal consistency reliability was examined. Internal consistency reliability is commonly used to evaluate the reliability of a set of test or questionnaire items. It provides an indication of an instrument's reliability by estimating the extent to which items on an instrument consistently measure the same construct (e.g., intern performance).[1] Values exceeding 0.70 are considered 'adequate', while values exceeding 0.80 are preferred for pilot work. For comparative purposes and because the response option 'No opportunity to observe' can feasibly be regarded as missing data, internal consistency reliability was examined both with the 'No opportunity to observe' response option included as well as with the 'No opportunity to observe' response option recoded as missing data.

Intern Evaluation ς Final Experience: Pilot Data: Brief Report (Draft)

The following report briefly summarizes basic item-level, reliability, and descriptive information for the revised intern evaluation. The obtained data summarized in this section are based on the second final-experience evaluation of 33 candidates, rated by both mentors and supervisors and completed during the fall, 2017 semester.

Item-Level Information

Item-level information is summarized in the following tables, first for mentors and then for supervisors.

	How Often Each Response Was Selected (Frequencies)		
Item	No opportunity	Unsatisfactory	Developing

( )De0(scre1mg7(ps)10(r) Tm[Item)]TJETQ132.504 531ET 314.45 13.44 reW

Item	How Often Each Response Was Selected (Frequencies)				
	No opportunity	Unsatisfactory	Developing	Proficient	Exemplary
Learner Development[a]	0	0	1	7	25

of how strongly mentors' and supervisors' responses to each evaluation item were related. Higher values indicate a stronger relationship between scores, and are indicative of stronger inter-rater reliability. Correlation values ranging from 0.35 to 0.59 are generally described as being moderate in strength, while values ranging from 0.60 to 0.79 are considered strong; values exceeding 0.80 would be described as very strong. In the following table, item-level correlation values delineating the relationships between mentors' and supervisors' ratings are presented.

Item	Correlation value	Interpretation
1. Learner Development[a]	0.49	Moderate
2. Learner Differences[a]	0.41	Moderate
3. Learning Environments[a]	0.42	Moderate
4. Managing Classroom Procedures[a]	0.31	Weak
5. Content Knowledge[b]	0.56	Moderate
6. Content Application[b]	0.55	Moderate
7. Pedagogical Procedures[b]	0.60	Moderate
8. Flexibility and Responsiveness[b]	0.44	Moderate
9. Learner Assessment[c]	0.45	Moderate
10. Learner Feedback[c]	0.51	Moderate
11. Impact on Learning[c]	0.38	Moderate

The findings indicate that scores based on mentors' and supervisors' ratings were significantly and moderately to strongly related. In particular, these correlations provide some degree of convergent