Detailed description of analyses from ESERA presentation:

Comparison of two alternative approaches to quality STEM teacher preparation: Fast track licensure and imbedded residency programs

August 22, 2017
Dublin, Ireland

Gale A. Mentzer, Ph.D.
Acumen Research and Evaluation, LLC
1811 N. Reynolds Road, Toledo, OH 43615
gmentzer@acumenresearcheval.com
419-265-1811

Charlene M. Czerniak, Ph.D.
The University of Toledo
Charlene.Czerniak@utoledo.edu
Toledo, OH

T. Ryan Duckett, M.A.
The University of Toledo
Timothy.Duckett@utoledo.edu

Janet Struble, M.S.
The University of Toledo
Janet.Struble2@utoledo.edu

Teacher self-efficacy

A teacher's confidence in his or her ability to provide quality instruction that will result in student learning has long been accepted as a precursor to the actual implementation of high quality instruction ([Author] et al, 2005; Evans, Luft, Czerniak., 2014; Milner, Sondergeld, Demir, Johnson, & Czerniak, 2012). Using RMM, conversion to an interval scale resulted in the following for the *STEBI*:

OAEL Personal Beliefs $\bar{x}$ = 37.48 Outcome Expectancy $\bar{x}$ = 30.98

LAMP Personal Beliefs $\bar{x}$ = 50.20 Outcome Expectancy $\bar{x}$ = 28.40

An F test indicated unequal variances ($p < 0.001$) and the resulting t test for unequal variances showed the LAMP group scored statistically significantly higher ($p = 0.0008$) than the OAEL on the Personal Beliefs scale. The corresponding effect sizes were small (0.06 for Personal Beliefs favoring LAMP; 0.05 for Outcome Expectancy favoring OAEL). These results are due in part to the differences in variances with the OAEL have standard deviations nearly twice as large as the LAMP suggesting that the LAMP scored much more consistently within their group and the OAEL had some scoring quite high and others rather low. The statistically significant finding on the Personal Beliefs scale is noteworthy considering the small samples. It indicates that the LAMP scholars had a higher degree of confidence in their ability to provide high quality science or math instruction than the OAEL scholars.

The *TSE* was also implemented as a measure of self-efficacy. Following best practices in measurement theory, the response options were reduced from the original nine options to five before the survey was administered (Alwin, 1992). Four independent t-tests (by overall and by subscale) were conducted. First the raw ordinal data were converted to interval logit based scores for the overall TSE scale and individually for each of the subscales using RMM. Table 2 provides the results. There was no statistically significant difference. Based upon the results, we can assume, with caution (due to large degree of measurement error), that group membership did not influence self-efficacy regarding instructional practices, classroom management, and student engagement.

Table 2: *TSE comparison by Licensure Type*

	OAEL (9)		LAMP (7)			
	Mean	SD	Mean	SD	t (14)	p

Overall TSE	42.48	1.90	44.13	3.06	-1.33	0.205
Instruction	21.63	1.20	22.64	1.81	-1.351	.0198
Management	21.05	2.32	22.12	3.23	-0.772	0.453
Engagement	20.23	1.74	21.37	2.05	-1.201	0.250

Teaching style preferences

Fourteen Noyce scholars completed the STIPS (OAEL = 7; LAMP = 7). RMM was used to determine the extent to which participant responses on the STIPS correlated between preference for inquiry and non-inquiry teaching practices. Figure 1 maps item difficulty and shows that the items functioned as expected overall. Item measures had a very high reliability of 0.86 which indicates the accuracy with which the questions can discern between inquiry and non-inquiry teaching practices. Inquiry based questions (I on Figure 1) comprised four out of the five most frequently endorsed items (listed at the bottom of the scale) and non-inquiry based questions (NI) were the five least frequently endorsed items (at top of scale).

Measure	More Person	Rare Item
2		NI 1 NI 9
	x	
	x	
		NI 3
1	x	NI 7
		NI 5
	xxxxx	
	x	I1
0	xx	I 8 NI10 NI 8
		I 4 I 7 NI 2
	x	I 9 NI 4
		I 6
-1		I 2
		NI 6
		I 3
		I 5
-2		
	Less	Frequent

Figure 1: STIPS item difficulty

These non-inquiry items that were the most difficult for respondents to agree with included (in the order of less likely to endorse to more likely):

- Science/math should be taught as a discipline of conclusive and authoritative information that has been verified beyond dispute.
- The primary objectives of lab experiments should be the development of manipulative skills and the ability to follow directions, which lead to planned results.
- During instruction, the student should be explicitly told the important concepts contained within the content dealing with the topic being studied.
- Lab experiments should be designed so that the correct results or answers will emerge for only those who follow the directions and procedures.

- During lab exercises, students should follow specific directions on what to observe, measure, and report to find the right answers to the problem.

The items respondents were the most likely to agree with in the order of easiest first included (inquiry preference unless noted):

- The learning of scientific concepts should include the alternative views, weaknesses of current explanations, and doubts about the validity of the conclusions.
- In the lab, the student should be free to identify on his or her own the relevant questions and means of investigation for pursuing possible results.
- The true nature of science/math should be illustrated to the student through the study of its technological applications and achievements. (non-inquiry)
- To learn science/math, the student should be provided situations that exemplify concepts but which require him or her to figure them out himself or herself from the examples encountered.
- Instructional materials must encourage students to formulate alternative ideas to concepts encountered.

Figure 2 plots the group relationship between inquiry and non-inquiry teaching practices. Respondents from the OAEL as a group had an almost balanced preference for both types of teaching, with three OAEL respondents slightly preferring non-inquiry based strategies (ratio of inquiry to non-inquiry was 1.24:1). By comparison, the LAMP group more uniformly preferred inquiry based instruction (ratio 1.36:1). The N-1 chi square test verified this finding. OAEL had a total inquiry preference score of 247 out of a total 458 points and the LAMP had 273 for inquiry out of 478. The analysis results indicated a practical significance between the groups

with the LAMP group having an 83.62% more likely chance of scoring higher on the inquiry preference scale (p = 0.33) and a 67.25% chance that the two groups are different.

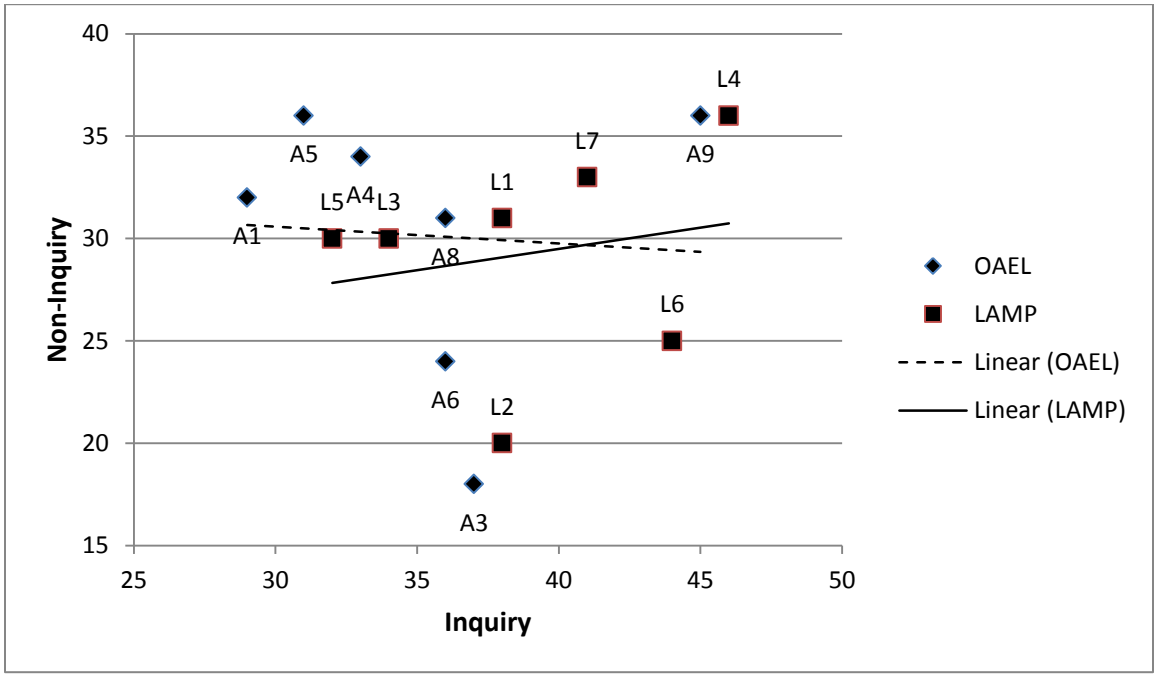

Figure 2: Mapping of STIPS scores

The *TBI* is a measure of theoretical versus practical approaches to teaching (nominal data). To calculate a score, we tallied frequencies of theoretical and practical responses for each respondent then added those frequencies to obtain a group total frequency. The OAEL had a total of 97 theoretical responses and 128 practical responses (total response = 225); the LAMP had 82 theoretical and 93 practical responses (total = 175). Because of the small sample size and the fact that the group sizes were not equivalent, the N-1 chi square test was again utilized. This analysis showed that the OAEL group had a 77.23% chance of scoring higher on the practical scale (two tailed p = 0.455) and there was a 54.47% chance the proportions of responses were different. Contributing factors to these differences may include that the OAEL scholars have been in the field longer, the OAEL were provided less guidance and may have learned "by the seat of their

pants," and the LAMP may have learned more theory as melding theory with practice is a focus for the LAMP program.

Responses to the *TEB Interview* were coded using the four- category scale described earlier. Using RMM to convert the ordinal rankings to an interval scale allowed for the calculation of a mean score and standard deviation for each group: OAEL *mean* = 45.34, SD = 7.04; LAMP *mean* = 51.30, SD = 4.35 (with expected mean of 50). Both groups scored close to the expected mean and appeared to endorse the middle category (transitional). A one-tail independent t test assuming equal variances ($F = 2.62$, $p = 0.12$), conducted under the assumption that the LAMP scholars would score higher, resulted in a statistically significant difference ($t = 1.96$, $df = 14$, $p = 0.035$) between groups suggesting that the LAMP group held more student-focused views of teaching and learning. Because the sample was small and therefore making it more difficult to identify a difference when one exists, the finding is most likely indicative of a real difference.

Closer examination of responses using qualitative analysis saw little difference between groups when asked to provide essential elements of inquiry instruction. Both groups emphasized a well-planned lesson that provided students the opportunity to develop their own solutions to a question. Similarly, both groups felt that student learning was evident not only through formal assessments but by observing student-directed discussions, student responses to oral questioning, and the quality of student questions. They were also similar in how they made decisions to move forward with curriculum (when the majority of students have exhibited mastery or when time dictated to stay on schedule). To maximize student learning, both groups felt differentiating instruction to suit a variety of learning styles was important and within each group about 30% of the respondents provided explicit strategies (electronic notebooks, small group work, modeling).

The *POSTT* explored teaching orientations. Participants indicated how they would implement lessons in a series of classroom scenarios. To assess the spectrum of participants' instructional approaches, responses were numerically coded so that open inquiry responses became a 4, guided inquiry became a 3, etc. RMM transformed the raw, ordinal scores into an interval scale. The results revealed very low person separation (0.56) suggesting the instrument was not sensitive enough to identify differences between teacher preferences. Combined with an additional low person reliability of 0.24, the data suggested that only one level of person ability (i.e. extent to which the participant advocates for open inquiry based science instruction) could be distinguished.

A review of the descriptive statistics provided further information. The range of scores for the instrument is 25 to 100. All of the scores occurred within the 3rd quartile (50 – 75) with observed range of 61 to 76. OAEL group mean was 67 and LAMP group mean was 71.3. An independent t-test revealed that the difference in the mean scores was not significantly statistic (t = -1.521, p = 0.154, df = 12) meaning that the groups did not differ on their teaching orientation preferences. The mean scores do indicate, however, that both groups prefer, in general, a combination of active direct and guided inquiry with more emphasis on the guided inquiry.

Results correspond with each group most commonly endorsing guided inquiry, a condition corroborated by the descriptive statistics (OAEL endorsed guided inquiry 36.6% of the time, LAMP 42.3% of the time). Further analysis using a chi square goodness of fit test indicated that in spite of similar overall teaching orientations, OAEL participants were more likely to endorse the didactic direct, the least desirable instructional approach ($\chi 2$ = 10.62, df = 3, p = 0.001 with residual for didactic category = -2.79) (Figure 3).

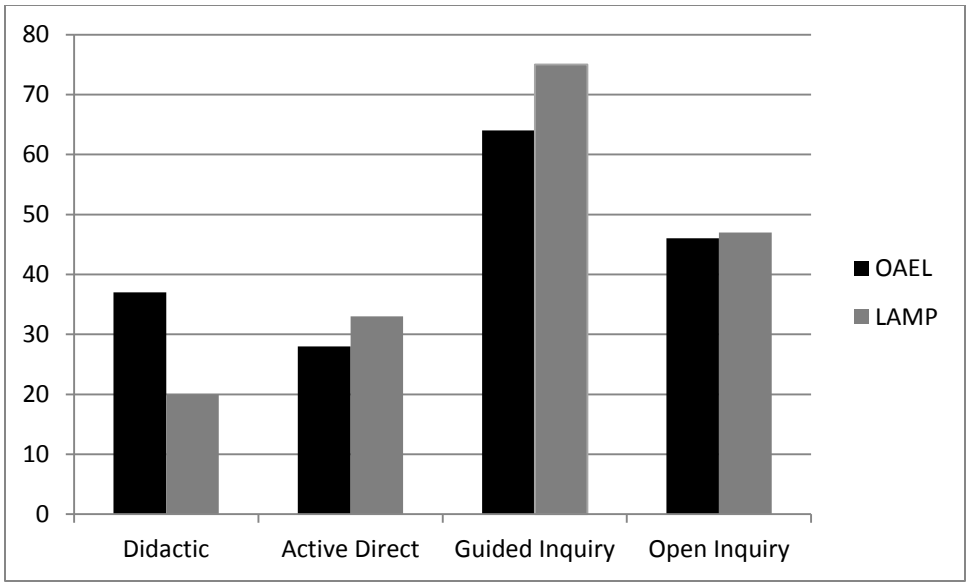

Figure 3: Frequencies of endorsements of instructional approach type

Ability to function successfully in a high needs school

The *Star Screener* quartile rankings provide a general assessment of a teacher's likelihood to succeed in a high needs district/school. The first quartile indicates a highly prepared teacher and the fourth quartile suggests the teacher is not prepared or predisposed to succeed in the high needs environment.

None of the respondents scored in the first or fourth quartile. An N-1 chi square test explored differences between the groups on the percent who scored in the second quartile (5 of 9 in OAEL, 6 of 7 in LAMP). The resulting probabilities were that there is a 78.87% chance the proportions are different and an 89.44% chance the LAMP students have a higher proportion of second quartile rankings than the OAEL students ($p = 0.211$ for a two tailed test). This suggests that there is indeed a practical significance that the proportion of LAMP students placed in the higher quartile ranking is higher than those who went through the OAEL thereby suggesting that LAMP may be better prepared for the high needs classroom. A closer examination of the specific

factors showed that scholars continued to score low on the same factors regardless of their teacher preparation program. Those areas where scholars continued to struggle included:

- Organization and Planning (how and why star teachers plan as well as their ability to manage complex classroom organizations).
- At-Risk Students (the likelihood that the respondent will be able to connect with and teach students of all backgrounds and levels).
- Survive in Bureaucracy (the likelihood that the respondent will be able to function as a teacher in large, depersonalized organization).
- Fallibility (how the teacher plans to deal with mistakes in the classroom).

One construct showed a slight difference between groups with the OAEL obtaining a median and mode score of 2 (average) and LAMP a 1 (below average). This was *Explains Teacher Success* (the criteria the respondent uses to determine teaching success and whether these are relevant to teachers in poverty schools). This difference suggests that LAMP may not use criteria to gauge their success as teachers that are relevant to the high needs classroom. On the other hand, LAMP appeared to be slightly more successful in the Application of Theory to Practice factor with a mode and median of 3 while the OAEL had a median of 2 and a multi-mode of 2 and 3 suggesting that the LAMP graduates may be slightly more successful in implementing practical applications of educational theory.

Data from the *Promoting Classroom Management Survey* identified disruptive behaviors in the classroom and the strategies teachers use to address those behaviors. Responses examined perceived severity and frequency of classroom disruptions. After converting the ordinal data to interval (RMM) an independent t-test was conducted to compare overall assessment of severity of disruptive behaviors. There was not a statistically significant difference in the scores for the

OAEL (M = 69.22, SD = 9.35) and LAMP (M = 64.29, SD= 8.94) (t = 1.067, df = 14, p = .304) suggesting that type of teacher licensure program did not affect overall level of severity assigned to common classroom disruptions.