



## Enhanced Learning Maps (ELM) Project Cohorts 1 and 2 Learning Mathematics For Teaching

### Pre and Posttest Summary

May 2019

---

#### Learning Mathematics For Teacher (LMT) Overview and Methods

Cohorts 1 and 2 teachers who participated in the mathematics component of the Enhanced Learning Maps (ELM) project were asked to complete the Learning Mathematics for Teaching (LMT) assessment following nearly two years involvement in the project. The teachers completed the pretest prior to attending the summer trainings in 2016 and 2017 for Cohorts 1 and 2, respectively. The posttest was completed in spring 2018 for Cohort 1 and spring 2019 for Cohort 2.

The LMT was developed by researchers from the University of Michigan through the support of federal funding from the National Science Foundation and the U.S. Department of Education (<http://www.umich.edu/~lmtweb/>). The purpose of the LMT is to assess mathematical content knowledge needed for teaching at the elementary and middle school levels. The LMT includes items that reflect the real mathematics tasks teachers face in classrooms such as assessing student work, representing numbers and operations, and explaining common mathematical rules or procedures. These teaching tasks require the same types of instructional decision making as are addressed by the ELM project goals.

The LMT includes multiple measures and forms. In consultation with the ELM project staff, McREL evaluators selected two modules for Cohort 1 and 2 participants to complete. These modules were (1) Number Concepts and Operations (NCOP) and (2) Patterns, Functions, and Algebra (PFA). A sample of released items from the LMT is available at <http://lmt.mspnet.org/index.cfm/17924>. Both modules were for teachers of middle level grades and were the 2007 versions of the assessment. McREL evaluators sent invites to the teachers requesting their completion of the LMT pre and posttests. A link to a consent form was included in the invitation. If a teacher consented to participate, he/she then received directions on accessing the LMT which is housed on an University of Michigan server. The assessment required about 30-40 minutes for each teacher to complete. Teachers were administered either Form A or B of the two modules. The LMT system randomly selected which form a teacher would receive. Twenty-one of the 36 Cohort 1 and 2 teachers completed both the pre and post NCOP assessments for a response

rate of 58%. Nineteen of the 36 teachers completed both PFA assessments for a response rate of 53%.<sup>1</sup>

## Analysis

The LMT results are analyzed using item response theory (IRT) scores. The University of Michigan staff used equating methods to determine the relative difficulty of each test form (i.e. Forms A and B). The two equating methods used were common item and common person. Test equating is a method that precisely estimates the relative difficulty of two or more tests and allows for accurate adjustment of scores. Based on the item difficulties, the “raw” scores (i.e., number correct) were assigned a scale or IRT score. The scale scores were adjusted for the relative difficulty of the tests and thus any change in scale score can be interpreted as a real gain in knowledge.

The University of Michigan researchers used equating procedures to generate conversion tables. Conversion tables provide an IRT equated scale score for each possible raw score answer on the test. The IRT scale scores are given in standard deviation units or “standardized” scores. A standardized distribution is one that has a standard deviation of 1 and a mean of 0.<sup>2</sup>

## Findings

The average (mean) change in the teachers’ IRT score on the NCOP was 0.07 ( $SD = 0.05$ ). The average (mean) change in the teachers’ IRT score on the PFA was 0.32 ( $SD = 0.04$ ) (Table 1). A paired samples t-test was used to examine the differences between the pre and posttests. The differences were not statistically significant for the NCOP ( $p=0.70$ ) or the PFA ( $p=0.31$ ).

Effect sizes for these analyses were calculated using Hedge’s  $g$ , which is the adjusted mean difference divided by the unadjusted pooled within-group standard deviation (U.S. Department of Education, 2014). It is important to consider the magnitude of an effect when placing findings into a broader context. Statistical significance will measure whether a program effect is due to chance, whereas effect sizes measure the strength or magnitude of the program’s effect and are not sensitive to the sample sizes. McREL researchers consider an effect size of 0.25 or greater to be educationally meaningful, and an effect size between 0.13 and 0.20 to be substantively important. These benchmarks are based on the What Works Clearinghouse’s methodological guidelines (U.S. Department of Education, 2014) and on a Lipsey et al. (2012) article, which reported an average effect size of 0.13 for 227 randomized controlled trials that examined the effect of curricula or broad instructional programs. Effect sizes (e.g., Hedge’s  $g$ ) were found to be educationally significant for the PFA ( $g = 0.26$ ) and negligible for the NCOP ( $g = -0.07$ ) (Table 1). Figures 1 shows the pre and post change in mean scores for each of the two assessments.

---

<sup>1</sup> Given the very small sample sizes, caution should be exercised when making any inferences from the findings.

<sup>2</sup> In a standardized distribution roughly 95% of individuals will fall within 2 standard deviations of the mean, or between -2 and 2 standard deviation units. Roughly 67% of participants will fall within one standard deviation of the mean. A scale score at 1 standard deviation above the mean would be higher than roughly 85% of expected scores in a standardized distribution. A scale score at 2 standard deviations above the mean would be higher than roughly 97% of the expected scores in a standardized distribution.

Table 1. Descriptive Statistics NCOP and PFA Pre and Posttests

Assessment	n	Pretest		Posttest		Significance	Effect Size
		Mean	SD	Mean	SD		
NCOP	21	0.30	1.02	0.23	0.97	0.704	-0.07
PFA	19	-0.44	1.18	-0.12	1.14	0.308	0.26

Note: Significance is reported using p-value. Effect size is reported using Hedge's *g*.



Figure 1. Pre and Posttest Change Scores for NCOP and PFA

**References:**

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

U.S. Department of Education. (2014). *What Works Clearinghouse Procedures and Standards Handbook, Ver. 3.0*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse. Retrieved from [https://ies.ed.gov/ncee/wwc/docs/referenceresources/wwc\\_procedures\\_v3\\_0\\_standards\\_handbook.pdf](https://ies.ed.gov/ncee/wwc/docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf)