SMU | Annette Caldwell Simmons School of Education & Human Development

Research in Mathematics Education

# STEM Teaching Observational Protocol (STEM TOP) Exploratory Factor Analyses

**Technical Report 19-06**

# STEM Teaching Observational Protocol (STEM TOP) Exploratory Factor Analyses

Elizabeth L. Adams • Anthony Sparks • Leanne R. Ketterlin-Geller

Southern Methodist University

**Fall 2019**

Published by
Southern Methodist University
Department of Education Policy & Leadership
Simmons School of Education & Human Development
PO Box 750114
Dallas, TX 75275-0114
Contact information: rme@smu.edu

This research was supported by The Texas Instruments Foundation and the O'Donnell Foundation; GT00317. Opinions expressed herein do not necessarily reflect those of The Texas Instruments Foundation, the O'Donnell Foundation or individuals within.

This document is available in alternative formats upon request.

## Table of Contents

# STEM Teaching Observational Protocol (STEM TOP) Analyses

# Introduction

The STEM Teaching Observational Protocol (STEM TOP) was developed to measure variation in teachers' use of active learning strategies in middle school science classrooms. The measure was developed to guide instructional coaches' feedback to teachers following a classroom observation. The STEM TOP requires ten to fifteen minutes of coaches' time to score. The STEM TOP includes 22 items, each of which is scored on a four-point scale. In order to examine initial validity evidence for the STEM TOP, this study investigates: To what extent does a multi-level exploratory factor analysis (MEFA) support the internal structure of the STEM TOP?

# Sample

During the 2018-19 school year, five coaches, who were trained and calibrated on the STEM TOP, conducted 267 observations with 39 teachers in a large urban school district (mean lessons per teacher = 6.8). We excluded one teacher who only participated in two classroom observations. To investigate the internal structure of the measure and account for the nesting of observations within teachers, we fit a multi-level EFA. In some cases, teachers were observed by a second observer for reliability checks, resulting in two observations of the same lesson per teacher. In those cases, we retained the scores assigned by the teachers' coach, not the second observer. We made this decision because in we were in future work, we are interested in investigating variation or bias in scoring attributable to the teachers' coach. It was hypothesized that the measure captured four dimensions of STEM teaching (i.e., lesson structure, learner-centered instruction, evaluation and feedback, and management and discipline). The purpose of this analysis is to generate empirical support for the theoretical factors, which will be used as scales in future analyses.

# Multi-level Exploratory Factor Analysis

We fit the multi-level exploratory factor analysis MEFA with the full sample of 39 teachers with 267 observations with the preferred categorical estimator, weighted least squares means and variances (WLSMV). Some teachers did not demonstrate change within teacher on certain items, suggesting these items are stable across time. All items demonstrated adequate ***between*** teacher variation as indicated by intra-class correlations (ICCs) above .30, which is above the threshold of .10 suggested by Raudenbush and Bryk (2001). Thus, the goals of this study are to: (a) identify variation and item groupings at the between (teacher) level, and (b) identify item groupings at the within (lesson) level if item groupings at that level are supported empirically.

## Fit One-Level EFA

Prior to testing a two-level model, the model should be supported with a one-level model (Byrne, 2012). We tested the one-level EFA with between one and six factors using the categorical

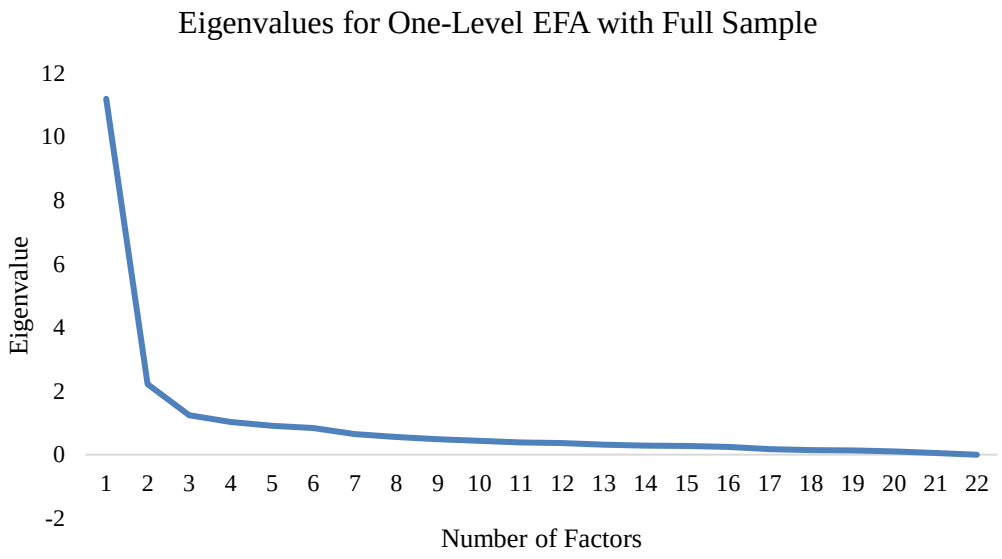
estimator WLSMV. Fit criteria were set such that an acceptable model would exhibit a chi-square p-value above 0.05, an RMSEA below .06, and a CFI/TLI above .90 (Byrne, 2012). In addition, a smaller chi-square and AIC/BIC were desirable.

*Figure 1: Summary of Model Fit Information for the One-Level Model with Full Sample and Categorical Estimator*

| Model | # of Parameters | Chi-square | DF | P-value | AIC/BIC | RMSEA | CFI/TLI | SRMR |
|---|---|---|---|---|---|---|---|---|
| 1 | 22 | 1038.134 | 209 | 0.000 | NA | 0.122 | 0.899<br>0.889 | 0.112 |
| 2 | 43 | 475.869 | 188 | 0.000 | NA | 0.076 | 0.965<br>0.957 | 0.062 |
| 3 | 63 | 324.460 | 168 | 0.000 | NA | 0.059 | 0.981<br>0.974 | 0.048 |
| 4 | 82 | 227.596 | 149 | 0.000 | NA | 0.044 | 0.990<br>0.985 | 0.038 |
| 5 | 100 | 168.487 | 131 | 0.015 | NA | 0.033 | 0.995<br>0.992 | 0.031 |
| 6 | 117 | 137.729 | 114 | 0.065 | NA | 0.028 | 0.997<br>0.994 | 0.026 |

*Note: AIC/BIC is not available when using a categorical estimator (Muthen & Muthen, 2017). Models demonstrating relatively better fit indicated in gray.*

## Plot the Eigenvalues



Both the model fit statistics and eigenvalue plots suggest between two to four factors without accounting for the nesting of lessons within teachers.

## Fit two-level MEFA

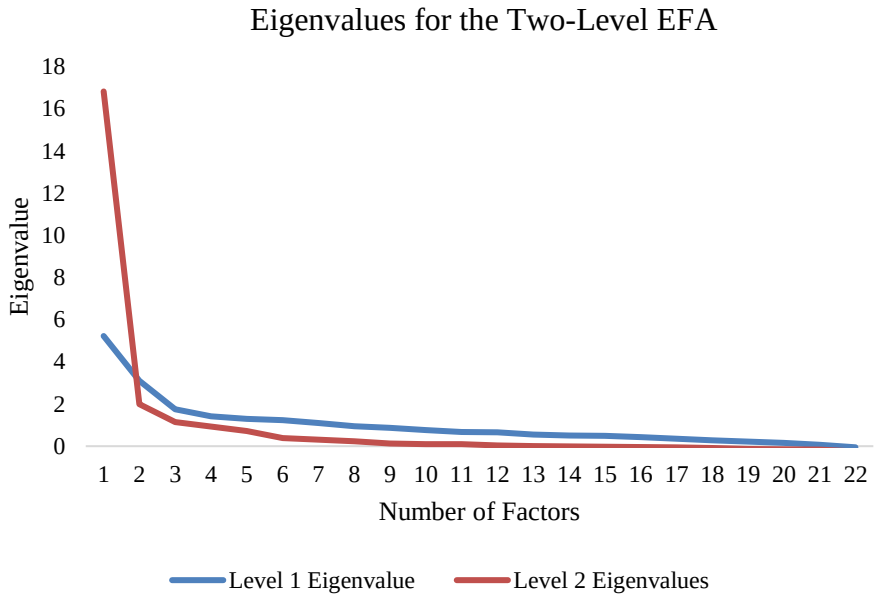
We fit the MEFA with the full sample of 39 teachers and 267 lessons treating the items as categorical (WLSMV estimator). Based on the one-level MEFA fit statistics, we examined the results of the two-level MEFA for between two and four factor solutions at the between and within levels. We also examined the results specifying the within-level as unrestricted to test if the model fit best at the between (teacher) level only. Again, fit criteria were set such that an acceptable model would exhibit a chi-square p-value above 0.05, an RMSEA below .06, and a CFI/TLI above .90 (Byrne, 2012). In addition, a smaller chi-square and AIC/BIC were desirable.

Summary of Model Fit Information for the Two-Level MEFA with *Full Sample and Categorical Estimator*

| # of Factors Between | # of Factors Within | # of Parameters | Chi-square | DF | P-value | AIC/BIC | RMSEA | CFI/TLI | SRMR Within Between |
|---|---|---|---|---|---|---|---|---|---|
| **2** | **2** | **174** | **456.307** | **376** | **0.0028** | **NA** | **0.028** | **0.920 0.902** | **0.091 0.070** |
| **2** | **3** | **194** | **391.079** | **356** | **0.0970** | **NA** | **0.019** | **0.965 0.955** | **0.072 0.070** |
| 2 | 4 | -- | -- | -- | -- | -- | -- | -- | -- |
| 3 | 2 | 194 | 434.056 | 356 | 0.0029 | NA | 0.029 | 0.922 0.899 | 0.091 0.055 |
| 3 | 3 | 214 | 368.886 | 336 | 0.1047 | NA | 0.019 | 0.967 0.955 | 0.072 0.055 |
| 3 | 4 | 233 | 339.570 | 317 | 0.1835 | NA | 0.016 | 0.978 0.967 | 0.064 0.055 |
| 4 | 2 | 213 | 413.701 | 337 | 0.0027 | NA | 0.029 | 0.924 0.895 | 0.091 0.040 |
| 4 | 3 | 233 | 348.882 | 317 | 0.1052 | NA | 0.019 | 0.968 0.954 | 0.072 0.040 |
| 4 | 4 | -- | -- | -- | -- | -- | -- | -- | -- |
| **2** | **Unrestricted** | **131** | **191.346** | **188** | **0.4184** | **NA** | **0.008** | **0.997 0.992** | **0.000 0.070** |
| 3 | Unrestricted | 151 | 158.688 | 168 | 0.6848 | NA | 0.000 | 1.000 1.025 | 0.000 0.055 |
| 4 | Unrestricted | 170 | 130.617 | 149 | 0.8584 | NA | 0.00 | 1.000 1.057 | 0.000 0.040 |

Note: Dashes indicate that the model did not converge.

## Plot the Eigenvalues




We identified the best-fitting, most parsimonious models, which are highlighted in gray, based on the fit criteria and eigenvalues. Both the model fit statistics and eigenvalue plots suggest two factors at the between (teacher level) and two, three, or an unrestricted factor solution at the within (lesson level). See Appendix A for item loadings and Appendix B for M*plus* results.

# Item Assignment

For three best-fitting models, we examined the item groupings to examine the correspondence between the theoretical or hypothesized groupings and the empirical groupings. We assigned the item to the grouping with the highest, statistically significant loading ($p<.05$).

- At the between level: We observed strong evidence for the two factor solution at the between (teacher) level. Empirical evidence suggests that the items for the first three hypothesized dimensions of instruction (i.e., lesson structure, learner-centered instruction, evaluation and feedback) measure one dimension of instruction (i.e., active learning). The items hypothesized to measure the dimension "management and discipline" are empirically supported.
- At the within level: We observed moderate evidence for the two factor at the within (lesson) level. Because we might be interested in using data at the lesson level, it was important to specify the model at the within level if the factors were empirically supported. We moved forward with the two factor solution because the three factor solution only included two items on the third factor, which is too few items to measure a

separate dimension of instruction (e.g., prone to bias, may not replicate in other work). For the two factor solution, the model was the same as the between level, with one exception. One item "LEA9: i. Teacher involves all students (e.g., calling on non-volunteers, facilitating student-student interaction, checking in with hesitant learners, etc.)" demonstrated the highest loading with the second factor "management and discipline." Given that the item loading was statistically significant on both the first and second factors and that this item loaded on the first factor at the between level, we assigned this item to the first factor to retain the most parsimonious model.

## Possible Cross-Loading Items

The following items exhibited statistically significant loadings on both factors. In all cases except for one at the within level, items were assigned to the factor with the highest loading. The exception was made based on theoretical support and to maintain parsimony across the between and within levels.

*Within level only:*

- "MAN4: d. Teacher efficiently manages time (e.g., transitions, wait time, pacing)" was assigned to "management and discipline" because the item loading was almost twice as strong as the other loading (.47 compared to .26).
- "FEE3: c. Students evaluate their own or other's work" was assigned to "active learning" based on theoretical evidence and because the item loading was stronger than the other loading (.34 compared to .25).
- "LEA9: i. Teacher involves all students (e.g., calling on non-volunteers, facilitating student-student interaction, checking in with hesitant learners, etc.)" was assigned to "active learning" based on theoretical evidence and to maintain similar item structure at the between and within levels, even though the item loading was higher for the other factor (.15 compared with .31).
- "STR4: d. The lesson is clearly connected to students' prior knowledge and experiences" was assigned to "active learning" based on theoretical evidence and because the item loading was higher than the other loading (.35 compared to .29).

*Between level only:*

- "MAN1: a. Students are on task throughout the class" was assigned to "management and discipline" because the item loading was more than twice as strong as the other loading (.71 vs. .36).
- "LEA6: f. Teacher poses cognitively demanding, open-ended questions" was assigned to "active learning" based on theoretical evidence and because the item loading was higher than the other loading (.63 compared to .42).
- "LEA1: a. Students explain and justify their thinking" was assigned to "active learning" based on theoretical evidence and because the item loading was higher than the other loading (.64 compared to .42).

*Both within and between levels:*

- "FEE2: b. Teacher provides feedback focused on expanding learning and understanding (formal and informal formative), not correctness or the end product (summative)" was assigned to "active learning" because the item loadings were almost twice as strong (.48 compared to .29 for within; .63 compared to .35 for between).
- "LEA10: j. Teacher is attentive to students' academic and social/emotional needs (e.g., use of cooperative learning, language-appropriate strategies and materials, awareness of student comfort)" was assigned to "active learning" based on theoretical evidence and because the item loadings were higher than the other loadings (.31 compared to .28 for within; .67 compared to .38 for between).

Based predominately on theoretical evidence and to maintain the most parsimonious, least complex model, we moved forward with the two factor solution at the within and between levels without cross loadings. The final model suggests two factors at the within (lesson) and between (teacher) levels: active learning (17 items) and classroom management and discipline (5 items).

# Conclusion

The evidence of internal structure presented in this report indicates that the STEM TOP is a viable tool for measuring two dimensions of STEM instruction including: (a) active learning, and (b) management and discipline, both of which are important aspects of effective STEM instruction. The reduction in time required to score this tool is practically meaningful for busy school-based personnel.

## References

Byrne, B. M. (2012). *Structural equation modeling with M*plus*: Basic concepts, applications, and programming.* New York, NY: Routledge.

Raudenbush, S. W. & Bryk, S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: SAGE Publications.

Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus User's Guide. Seventh Edition.* Los Angeles, CA: Muthén & Muthén.

# Appendix A – MEFA Factor Loadings

| Results Two-level EFA with Full Sample | | | Two Factor Between, Two Factor Within | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Within | | | | Between | | | |
| | | | Factor 1 | | Factor 2 | | Factor 1 | | Factor 2 | |
| Number | Item | ICC | Loading | Descriptio | Loading | Descriptio | Loading | Descriptio | Loading | Descriptio |
| **STR1** | **a. The lesson objectives are clear to students** | 0.53 | 0.374* | active lear | 0.264* | | 0.933* | active lear | -0.017 | |
| **STR2** | **b. The lesson is structured to build understanding and maintain a sense of purpose** | 0.56 | 0.451* | active lear | 0.131 | | 0.839* | active lear | 0.175 | |
| **STR3** | **c. The lesson includes an investigative or problem-based approach (e.g., students investigate or discover scientific ideas)** | 0.37 | 0.769* | active lear | -0.198* | | 0.897* | active lear | -0.002 | |
| **STR4** | **d. The lesson is clearly connected to students' prior knowledge and experiences** | 0.46 | 0.351* | active lear | 0.285* | possible c | 0.902* | active lear | 0.131 | |
| LEA1 | a. Students explain and justify their thinking | 0.50 | 0.533* | active lear | 0.053 | | 0.636* | active lear | 0.418* | possible c |
| LEA2 | b. Students engage in behaviors reflective of the process standards | 0.32 | 0.782* | active lear | -0.208* | | 0.854* | active lear | 0.092 | |
| LEA3 | c. Students direct their own learning (e.g., are provided with flexibility or choices during the lesson) | 0.37 | 0.579* | active lear | -0.064 | | 0.731* | active lear | 0.161 | |
| LEA4 | d. Teacher engages students in appropriately challenging content (e.g., critical thinking, problem-solving strategies) | 0.48 | 0.658* | active lear | 0.000 | | 0.693* | active lear | 0.168 | |
| LEA5 | e. Teacher openly welcomes discussion about mistakes or misconceptions | 0.72 | 0.462* | active lear | 0.093 | | 0.988* | active lear | -0.132 | |
| LEA6 | f. Teacher poses cognitively demanding, open-ended questions | 0.43 | 0.456* | active lear | -0.004 | | 0.627* | active lear | 0.419* | possible c |
| LEA7 | g. Teacher explicitly connects learning to the real world (e.g., careers, current events) | 0.34 | 0.182* | active lear | 0.055 | | 0.773* | active lear | 0.150 | |
| LEA8 | h. Teacher explicitly connects learning to other disciplines (e.g., social studies, mathematics) | 0.44 | 0.261* | active lear | 0.195 | | 0.799* | active lear | -0.142 | |
| LEA9 | i. Teacher involves all students (e.g., calling on non-volunteers, facilitating student-student interaction, checking in with hesitant learners, etc.) | 0.59 | 0.147* | active lear | 0.313* | possible c | 0.996* | active lear | -0.111 | |
| LEA10 | j. Teacher is attentive to students' academic and social/emotional needs (e.g., use of cooperative learning, language-appropriate strategies and materials, awareness of student comfort) | 0.66 | 0.315* | active lear | 0.276* | possible c | 0.667* | active lear | 0.382* | possible c |
| **FEE1** | **a. Teacher uses a variety of assessment strategies (e.g., large group questions, one on one discussion, small group feedback, exit tickets, quiz or test, informal progress check)** | 0.53 | 0.365* | active lear | 0.142 | | 0.866* | active lear | -0.022 | |
| **FEE2** | **b. Teacher provides feedback focused on expanding learning and understanding (formal and informal formative), not correctness or the end product (summative)** | 0.50 | 0.477* | active lear | 0.288* | possible c | 0.626* | active lear | 0.354* | possible c |
| **FEE3** | **c. Students evaluate their own or other's work** | 0.52 | 0.341* | active lear | 0.248* | possible c | 0.879* | active lear | 0.003 | |
| MAN1 | a. Students are on task throughout the class | 0.66 | -0.005 | | 0.790* | managem | 0.335* | possible c | 0.713* | managem |
| MAN2 | b. Students demonstrate an understanding of expectations for behavior | 0.70 | -0.051 | | 0.980* | managem | 0.048 | | 0.953* | managem |
| MAN3 | c. Students demonstrate an understanding of classroom procedures/routines | 0.62 | 0.142 | | 0.962* | managem | -0.006 | | 1.004* | managem |
| MAN4 | d. Teacher efficiently manages time (e.g., transitions, wait time, pacing) | 0.70 | 0.259* | possible c | 0.469* | managem | 0.070 | | 0.894* | managem |
| MAN5 | e. Teacher redirects off task or disruptive behavior (NA if no disruptive behavior) | 0.61 | -0.101 | | 0.492* | managem | -0.049 | | 0.892* | managem |

See Excel document saved: https://smu.box.com/s/vng4ay370x7lgvfyq6ztxjwrl8jn9bhp

M*plus* analysis file saved: https://smu.box.com/s/uh978rxy7gvastavb40r7wbwjunoc9rc

# Appendix B – MEFA Results for the Best Fitting Models

Two Factor Within, Two Factor Between Solution

WITHIN LEVEL RESULTS

GEOMIN ROTATED LOADINGS (* significant at 5% level)

| | 1 | 2 |
|---|---|---|
| STR1 | 0.374* | 0.264* |
| STR2 | 0.451* | 0.131 |
| STR3 | 0.769* | -0.198* |
| STR4 | 0.351* | 0.285* |
| LEA1 | 0.533* | 0.053 |
| LEA2 | 0.782* | -0.208* |
| LEA3 | 0.579* | -0.064 |
| LEA4 | 0.658* | 0.000 |
| LEA5 | 0.462* | 0.093 |
| LEA6 | 0.456* | -0.004 |
| LEA7 | 0.182* | 0.055 |
| LEA8 | 0.261* | 0.195 |
| LEA9 | 0.147* | 0.313* |
| LEA10 | 0.315* | 0.276* |
| FEE1 | 0.365* | 0.142 |
| FEE2 | 0.477* | 0.288* |
| FEE3 | 0.341* | 0.248* |
| MAN1 | -0.005 | 0.790* |
| MAN2 | -0.051 | 0.980* |
| MAN3 | 0.142 | 0.962* |
| MAN4 | 0.259* | 0.469* |
| MAN5 | -0.101 | 0.492* |

GEOMIN FACTOR CORRELATIONS (* significant at 5% level)

| | 1 | 2 |
|---|---|---|
| 1 | 1.000 | |
| 2 | 0.091 | 1.000 |

BETWEEN LEVEL RESULTS

GEOMIN ROTATED LOADINGS (* significant at 5% level)

| | 1 | 2 |
|---|---|---|
| STR1 | 0.933* | -0.017 |
| STR2 | 0.839* | 0.175 |
| STR3 | 0.897* | -0.002 |
| STR4 | 0.902* | 0.131 |
| LEA1 | 0.636* | 0.418* |
| LEA2 | 0.854* | 0.092 |
| LEA3 | 0.731* | 0.161 |
| LEA4 | 0.693* | 0.168 |
| LEA5 | 0.988* | -0.132 |
| LEA6 | 0.627* | 0.419* |
| LEA7 | 0.773* | 0.150 |
| LEA8 | 0.799* | -0.142 |
| LEA9 | 0.996* | -0.111 |
| LEA10 | 0.667* | 0.382* |
| FEE1 | 0.866* | -0.022 |
| FEE2 | 0.626* | 0.354* |
| FEE3 | 0.879* | 0.003 |
| MAN1 | 0.335* | 0.713* |
| MAN2 | 0.048 | 0.953* |
| MAN3 | -0.006 | 1.004* |
| MAN4 | 0.070 | 0.894* |
| MAN5 | -0.049 | 0.892* |

GEOMIN FACTOR CORRELATIONS (* significant at 5% level)

| | 1 | 2 |
|---|---|---|
| 1 | 1.000 | |
| 2 | 0.653* | 1.000 |

GEOMIN FACTOR CORRELATIONS (* significant at 5% level)

| | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 1.000 | | |
| 2 | 0.657* | 1.000 | |
| 3 | 0.324 | 0.208 | 1.000 |