

CAEP Annual Report: UTeach Observation Protocol (UTOP) Summary and Data (Instrument used by BSED Secondary Education, Sciences and Mathematics and MAT-S Teaching Science with Certification)

1. Case for Meeting Standard Component

- The UTeach Observation Protocol (UTOP) evaluation form is a proprietary instrument developed by the University of Texas at Austin. The development of the UTOP evaluation instrument was based in part on the InTASC Standards, and it was intentionally selected by NAU Professional Education Programs, the Department of STEM Education, and the Department of Mathematics and Statistics due to the explicit alignment between the criteria and InTASC Standards (**CAEP Standard Components R1.1, R1.2, R1.3, R1.4**). It is a validated instrument that requires inter-rater reliability training (i.e., calibration sessions) to support the consistent implementation and reliability of the data collected. The UTOP evaluation form is used by University Supervisors to evaluate a student teacher's design and delivery of a lesson through a formal mid-term and final observation, thus providing evidence for **CAEP Standard Component R3.3**, Competency at Completion. We have used the UTOP instrument since the 2015-2016 academic year.
- The UTeach Observation Protocol (UTOP) for Mathematics and Science was selected for the following reasons:
 - Correlation between the instrument best practice indicators for science and mathematics instruction and positive impact on student learning (**CAEP Standard Component R1.1** "The Learner and Learning"; **CAEP Standard Component R.3** "Instructional Practice").
 - Opportunity to use an instrument being used by other UTeach programs, thus bringing a level of standardization between NAUTeach and other UTeach programs.
 - Positive recommendations from other UTeach programs using the instrument.
 - Flexibility to make minor adaptations to the instrument to ensure appropriate National Science Teachers Association or NSTA (SPA) standards were addressed and assessed.
 - Opportunity to use an instrument designed specifically to assess science and mathematics instruction indicators congruent with the best practices and principles promoted in the UTeach model. (**CAEP Standard Component R1.2** "Content").
 - Opportunity to use an instrument whose validity and reliability had already been established and documented (**CAEP Standard Component R5.2**).
- The UTOP evaluation form indicators are categorized into four domains which are aligned to the InTASC Categories and relevant SPA standards as follows:

InTASC Category I: The Learner and Learning (InTASC Standards 1-3):

- 1.1 Classroom Engagement
- 1.2 Classroom Interactions
- 1.3 Classroom On-Task
- 1.4 Classroom Management
- 1.5 Classroom Organization
- 1.6 Classroom Equity
- 3.2 Implementation Involvement

InTASC Category II: Content (InTASC Standards 4-5):

- 4.1 Content Significance
- 4.2 Content Fluency

- 4.3 Content Accuracy
- 4.5 Content Abstraction
- 4.6 Content Relevance
- 4.7 Content Interconnections
- 4.8 Content Societal Impact

InTASC Category III: Instructional Practice (InTASC Standards 6-8):

- 2.1 Lesson Sequence
- 2.2 Lesson Importance
- 2.3 Lesson Assessments
- 2.4 Lesson Investigation
- 2.5 Lesson Resources
- 3.1 Implementation Questioning
- 3.3 Implementation Modification
- 3.4 Implementation Timing
- 3.5 Implementation Connections
- 3.6 Implementation Safety
- 4.4 Content Assessments

InTASC Category IV: Professional Responsibility (InTASC Standard 9):

- 2.6 Lesson Reflection

- The following NSTA SPA standards were also added to the UTOP instrument; only student teachers completing the BSED Secondary Education, Science (Biology, Chemistry, Earth Science, General Science, Physics) and the MAT-S Teaching Science With Certification programs are evaluated on the following additional criteria:
 - 6a: Practices safe and proper techniques for storage, handling, and disposal of instructional materials. (NSTA 12: 4a)
 - 6b: Follows emergency procedures and maintains safety equipment and ensures safety procedures that are appropriate for students’ activities and abilities. (NSTA 12: 4b)
 - 6c: Demonstrates legal and ethical responsibility for the welfare of students and other living organisms. (NSTA12: 4c)
- *Summary:* The AY 2020-2021 and AY 2021-2022 data results were shared internally with the Initial Teacher Preparation (ITP) Coordinating Council and with Cooperating Teachers (key stakeholders) through new feedback items that were added to the instrument cooperating teachers complete. These additional items regarding these evaluation results were piloted in Fall 2021 and fully implemented in Spring 2022. The Cooperating Teacher feedback is publicly available in the [EPP’s CAEP Annual Report webpage](#), in the accordion file labeled “Stakeholder involvement.” The raw data results and data analysis from these additional items as well as open-ended comments provided by Cooperating Teachers were provided to ITP Coordinating Council members and program coordinators at the March 2022 ITP CC meeting for further review and to determine if any instructional or program improvements are needed and if so, what curriculum changes should be made. PEP leadership and staff followed up with programs at the April 2022 ITP CC meeting and in Fall 2022 to request documentation of any curriculum changes based on the data results. In addition, at the conclusion of each fall and spring semester, program level assessment files (i.e., Assessment 4) for all undergraduate and graduate Secondary Mathematics Education and Science Education initial teacher preparation programs are updated as appropriate with UTOP results. Use of data results are noted in the

interpretation section of this assessment file (i.e., item d.). These examples illustrate use of data in Section 4 Findings, Implications, and Use of Data.

2. Data Sources and Methods

- **Data Source:** The UTOP evaluation form is used for two formal evaluations during student teaching in all undergraduate and graduate Secondary Mathematics Education and Science Education initial teacher preparation programs. A copy of the UTOP is provided as Appendix A at the end of this evidence file. This evidence file contains UTOP data results from Fall 2020 to Spring 2023 (i.e., three cycles of data). Data results are updated in this evidence file at the conclusion of each academic year.
- **Methods:** The UTOP is used for two formal evaluations during student teaching for the secondary education mathematics and science education programs. These programs are affiliated with the UTeach program at the University of Texas at Austin and use the UTOP, an instrument developed by the UTeach program. The primary purpose of the UTOP is to provide targeted feedback to a candidate regarding their performance related to discipline-related best practices and the InTASC Standards. The UTOP is provided to student teachers through the university's learning management system. The scoring rubric was created in the learning management system to collect evaluation data.

At the initial student-teaching team meeting, the University Supervisor introduces the UTOP and scoring guidelines to the Cooperating Teacher and Student Teacher. The Student Teacher also completes a self-assessment using the UTOP, which is discussed at the initial meeting. A copy of the UTOP, instructions, and scoring guidelines are included as Appendix A of this evidence file. As recommended by the University of Texas at Austin, a passing score is considered an overall **“human,”** or synthetic, mean score of 3 or higher on the final evaluation for each UTOP category (or rating section). The instrument and training guides indicate that the evaluator must not calculate a mathematical mean score when providing a holistic rating of the candidate's performance on a particular UTOP rating section. Instead, the rater should provide an overall impression of the lesson, using a holistic view of the domain and providing a **“human average”** of the entire lesson. Evidence supporting the synthesis ratings (feedback and comments) should be provided in the appropriate spaces on the UTOP rubric. Furthermore, an evaluation must only be given after the observation of the lesson has taken place, and the rater has had an opportunity to review their field notes as needed to provide evidence for each rating assigned. A score of **“NA”** (Not Applicable/Not Observed) may only be provided for the five indicators that specifically mention an NA option (i.e., indicators 1.2, 2.6, 3.6, 4.4, and 4.5). To maintain consistency among UTOP ratings, no mandatory rating section or indicator should be left blank, even if the rater feels the indicator is not applicable to the observed lesson.

Candidates are scored on a 0-5 scale on the UTOP Observation Protocol and on the Cooperating Teacher Evaluation Instrument. On both instruments, a score of 1 or 2 is considered approaching, a score of 3 or 4 is considered meeting, and a score of 5 is considered exceeding expectations. In general, the numerical values for the Likert scale on the UTOP can be interpreted as follows:

1 = Not observed at all / Not demonstrated at all

2 = Observed rarely / Demonstrated poorly

3 = Observed an adequate amount / Demonstrated adequately [Target performance]

4 = Observed often / Demonstrated well

5 = Observed to a great extent / Demonstrated to a great extent

According to the training guide, each value on the rating scale corresponds to two attributes of the candidate's performance—the frequency of the occurrence of the indicator, and the quality of the implementation of that indicator—though only one of these attributes may be appropriate for a given lesson.

The Professional Education Programs (PEP) Director serves as the primary instructor for all student teaching courses and sections. Before posting individual grades, PEP staff verify all student teaching requirements have been successfully completed by each candidate, including a passing score on the UTOP (CAEP Standard R1 and Standard Component R3.3).

- **Data Reporting:** At the completion of each semester, the UTOP data are extracted from the university's LMS (Learning Management System) and archived in Business Objects, a university supported reporting tool. The data are disaggregated by program and results are added to Assessment 4 of the current version of each program review report for all undergraduate and graduate secondary mathematics education and science education initial teacher preparation programs. For analysis at the EPP level, the instrument items are presented by program and InTASC Standard, and mean scores are computed. Both data sets are stored on a SharePoint website that faculty and college leaders can review and analyze for program evaluation and continuous improvement. Through this process, the EPP provides relevant internal stakeholders with evaluation data that can be used in program design, evaluation, and continuous improvement (**CAEP Standard Component R5.3**).

The UTOP, including a description of the instrument and reasons NAU chose this instrument, is shared with stakeholders through the Teacher Preparation Advisory Council (see **Evidence File R5.3_EF342_Clinical Partners and Stakeholders Meetings and Feedback.docx**). As noted above, initial UTOP results were also shared with cooperating teachers through additional survey items added to the evaluation instrument that is completed by these teachers regarding their candidates. Cooperating teachers are asked to rate if they observed similar strengths and areas for improvement as identified through the aggregated UTOP results.

2.a. Evidence of Data Credibility (CAEP Standard Component RA5.2)

- The inter-rater reliability and validity of the UTOP instrument has been widely examined and documented by the UTeach Institute, using the data from the UTeach program at University of Texas, Austin. This information can be found at <https://utop.uteach.utexas.edu/>. Online training regarding the UTOP and use of the UTOP is available for university supervisors.

2.b. Participants

- Across the Secondary Education Mathematics and Science Education programs utilizing the UTOP, there were the following numbers of program graduates: in Fall 2020, there were 10 program graduates, and in Spring 2021, there were 8 program graduates. In Fall 2021, there were 6 program graduates, and in Spring 2022, there were 19 program graduates. In Fall 2022, there were 7 program graduates; in Spring 2023 there were 18 program graduates.

The number of program graduates and response rates related to the data reported are listed in the left column of the tables in Section 5 Data at the end of this evidence file.

The number of programs graduates by program and semester are listed in the table below. The data presented below only represent candidates in the following programs which adopted the UTOP instrument: BSEd Secondary Education, Mathematics, BSEd Secondary Education, Sciences (Biology, Chemistry, Earth Science, General Science, Physics), and MAT-S Teaching Science with Certification. All other NAU initial teacher preparation programs use the NIET Aspiring Teacher Rubric for the student teaching evaluation (see **Evidence File R1.1_EF02_Student_Teaching_Evaluation_Aspiring Teacher Rubric**).

Teacher Preparation Program	Number of Program Graduates					
	Fall 2020	Spring 2021	Fall 2021	Spring 2022	Fall 2022	Spring 2023
BSEd Secondary Education - Mathematics	6	3	2	9	3	7
BSEd Secondary Education, Sciences (Biology, Chemistry, Earth Science, General Science, Physics)	4	5	4	10	4	7
MAT-S Teaching Science with Certification	0	0	0	0	0	4
Totals	10	8	6	19	7	18

3. Data Analysis

- PEP staff, the Department of STEM Education chair, and a Math Education faculty member reviewed the aggregate data from the UTOP results for the past two academic years (i.e., AY 2019-2020 and AY 2020-2021). For the AY 2021-2022, PEP staff reviewed the aggregate data from the UTOP results and shared the findings with leadership and faculty for the Secondary Education Mathematics and Science Education programs. Due to the small number of program graduates for the programs that utilize the UTOP, data charts are not included in this file but are reported and reviewed internally. Analysis of the UTOP data results focused on identifying overall strengths and areas for improvement based on the UTOP mid-term student teaching assessment results. We were primarily looking for items where there were patterns of average scores around 4.5 or higher at mid-term. Areas for improvement were based on overall results and across secondary education math and science students. We were primarily looking for items where there were patterns of average scores below 3.0 at mid-term. We chose mid-term results intentionally since these results more directly reflect a candidate’s preparation through coursework and program requirements prior to student teaching. Once we identified strengths and areas for improvement based on the criteria below, we referenced the UTOP website (see <https://pd.utexas.edu/utop-iii-rating-scales>) for the detailed language for the indicators. This information was used to tag each strength and area for improvement.

Based on faculty feedback in relation to findings based on the AY 2019-2020 and AY 2020-2021 UTOP data results, PEP staff re-reviewed the UTOP aggregate data and confirmed that 1.4 Classroom Management and 1.6 Classroom Equity did not fall below a 3.0 at mid-term, which is why they were not included as areas for improvement. There was one exception, but it was data

for a single student, so that data point was not considered because there was not a pattern across multiple semesters.

Results Patterns Across AYs 2020-2021, 2021-2022, and 2022-2023

The following strengths and areas for improvement were identified:

Strengths (patterns of average scores around 4.5 or higher at mid-term)

- Organizing classroom appropriately such that students can work in groups easily and get to lab materials as needed, and the teacher can move to each student or student group (Classroom Organization, InTASC Standard 3)
- Accurate written and verbal content information (Content Accuracy, InTASC Standard 4)
- Using appropriate resources (e.g., presentation tools, visual organizers, calculators, lab equipment, manipulatives, worksheets, etc.) to implement the lesson (Lesson Resources; InTASC Standard 7)
- Implementing safe, ethical, and environmentally appropriate lab procedures and/or classroom activities (Implementation Safety; InTASC Standard 8)
- Reflecting critically about their practice after the lesson including recognizing strengths and weaknesses related to planning, structure of the lesson, and instructional decision-making during the lesson (Lesson Reflection; InTASC Standard 9)
- Content Knowledge was a strength (Content Knowledge: InTASC Standard 4)

Areas for Improvement (patterns of average scores below 3.0 at mid-term)

- Communicating to students how the content fits into the big picture of the discipline and making it clear why the concepts are significant and important to learn (Content Relevance; InTASC Standard 5)
- Connecting math and science concepts across the disciplines to help generalize the content and make it more coherent (e.g., math lesson on graphing quadratic equations connects to related physics principles) (Content Interconnections; InTASC Standard 5)
- Discussing the content topic in relation to history, current events, or relevant “real-world problems” (Content Societal Impact; InTASC Standard 5)

Results Disaggregated for Race/Ethnicity and Gender

The following patterns were revealed when results for Fall 2021, Spring 2022, Fall 2022, and Spring 2023 were analyzed.

- Of the female candidate ($n = 32$) and male candidates ($n = 18$), female candidates had higher overall mean scores for two semesters and male candidates had higher overall scores for two semesters. This indicates that women as well as men are succeeding in the math and sciences and that the programs have successfully minimized gender bias.
- When analyzed for race/ethnicity, there were no clear patterns of racial/ethnic inequities. There was a strong presence of traditionally underrepresented racial/ethnic groups, with 30% of candidates identifying as American Indian/Alaska Native, Latine, or 2 or more races ($n = 15$); 68% of candidates identified as White ($n = 34$). One candidate did not identify a race ($n = 1, 2\%$).

4. Findings, Implications, and Use of Data

- *Findings/Implications:* Across all of the data cycles, the results indicated candidate strengths in relation to InTASC Standards 7, 8, and 9. The results patterns across AYs 2020-2021, 2021-2022, and 2022-2023 also indicated InTASC Standards 3 and 4 as strengths of candidates in the

secondary education mathematics and sciences programs. The results, across all of the data cycles, also suggested that the UTOP indicators and corresponding InTASC Standard where secondary education mathematics and science candidates are struggling the most at the mid-term evaluation are related to InTASC Standard 5, Application of Content.

- *Use of Data:* These data results were shared internally with the Initial Teacher Preparation (ITP) Coordinating Council in Fall 2021 and Fall 2022, respectively, and with Cooperating Teachers (key stakeholders) through new items added to the instrument cooperating teachers complete. The additional items regarding these evaluation results were piloted in Fall 2021 and fully implemented in Spring 2022. The raw data results and data analysis from these additional items as well as open ended comments provided by Cooperating Teachers have been provided to ITP Coordinating Council members and program coordinators to review further and determine if any instructional or program improvements are needed and if so, what curriculum changes should be made. For a full discussion of these Cooperating Teacher feedback items, please see evidence file **Evidence File R5.3_EF342_Clinical Partners and Stakeholders Meetings and Feedback.**

In addition, at the conclusion of each fall and spring semester, program level assessment files (i.e., Assessment 4) for all secondary education mathematics and science programs are updated as appropriate with UTOP results. Use of data results are noted in the interpretation section of this assessment file (i.e., item d.) and the Assessment 4 file is submitted as part of the formal program review process as a supplemental document for the Arizona Department of Education program review process.

- While student teachers performed well at mid-term during the past two academic years for the areas of Content Significance and Content Accuracy (InTASC Standard 4, Content Knowledge), data indicated several areas for improvement related to InTASC Standard 5, Application of Content, including:
 - Communicating to students how the content fits into the big picture of the discipline and making it clear why the concepts are significant and important to learn (Content Relevance; InTASC Standard 5)
 - Connecting math and science concepts across the disciplines to help generalize the content and make it more coherent (e.g., math lesson on graphing quadratic equations connects to related physics principles) (Content Interconnections; InTASC Standard 5)
 - Discussing the content topic in relation to history, current events, or relevant “real-world problems (Content Societal Impact; InTASC Standard 5)

5. Data

**UTeach Observation Protocol (UTOP) Midterm and Final Evaluations
Summary Data
by InTASC Standards
Fall 2020 to Spring 2023**

- The three teacher preparation programs utilizing the UTOP are the BSEd Secondary Education, Mathematics, BSEd Secondary Education, Science (Biology, Chemistry, Earth Science, General Science, and Physics), and MAT-S Science Education with Certification (Secondary).

The UTOP has five performance levels (or scales), with a “human average” score of “3” indicating the target performance. Mean scores in the tables below should be interpreted based on this benchmark and indicate the average scores candidates achieved on the rubric rows that comprise each InTASC category. The overall mean score for each program shown in the last column indicates the average score of each rubric row. The UTOP Observation Protocol evaluation indicators and results are presented in relation to InTASC Standards in the data tables below.

In all cases below, the response rate is 100%.

