Looking for Patterns in Data Obtained from Program-theory Model Based Evaluations

Kathryn Race
Race & Associates, Ltd.

Presented at the
American Evaluation Association Conference
Denver, Colorado
October 2014

Overview of Think Tank

- Provide quantitative data examples from modest-level budget evaluations
- Examples are excerpts from these evaluations
  (not the whole evaluation and not all of the analyses)
- Show how these investigations align with the program model
- Time for you to muse or discuss similar approaches you might take for your data and/or a general discussion

Ask: Is this (more or less) what you expected to see in this Think Tank?

Structure of Think Tank

- Present examples (three examples 15 minutes)
- Facilitate discussion and questions (15 minutes) – after each example or after all
- Other ways to identify patterns or thinking about data patterns and then using this information to guide analysis (15 minutes of what is the take away from this session)

Make sense?

What We Propose to Do

- Examples are from small(er) data sets
- Careful data exploration
- Data sense-making
- Using data to guide our conclusions and next steps
- Thus, not letting data guide us but guiding our decisions with data
- Use these methods in support of traditional statistical tests (when appropriate)

We are suggesting that we can adapt some of the techniques that are applied to complex mathematical models (at least the reasoning behind why an approach is taken) as well as take an approach that may be analogous to determining “common themes” in qualitative data.

Assumptions

- That consistency in the data reflects something meaningful
- Aligning our data explorations with program outcomes and core strategies from our program theory model offers structure
- Attempt to rule out other explanations (e.g., such as use of a single method to collect data) to support that the consistency may be meaningful
- Understanding the difference between the “signal” and the “noise”

Sometimes it is difficult to differentiate the noise (random variance/non systematic variance) and the signal (the intervention or the program). Nate Sliver in his book, *The Signal and the Noise* talks about the issue of over-correction in complex mathematical models, the difficulty in knowing when the model may be over-corrected or under-corrected. I am going to borrow from Nat’s language but apply it to data situations where complex mathematical models are not applicable because of the size of the data set.

What we're not talking about

We are linking or aligning these investigations to a program model – not a general exploration or "fishing trip" where we look long enough to find something of interest. So we are talking about purposeful investigations of the data – sometimes to let the data "speak" for themselves.

What we’re not talking about

What we don't have

- Large data sets with large sample sizes
- Large, complex mathematical models .....

But we do have examples that are:

- Mostly quantitative analyses from small scale evaluations supported by statistical tests
- Could talk about implications for qualitative data if there is time and interest

Any questions so far?

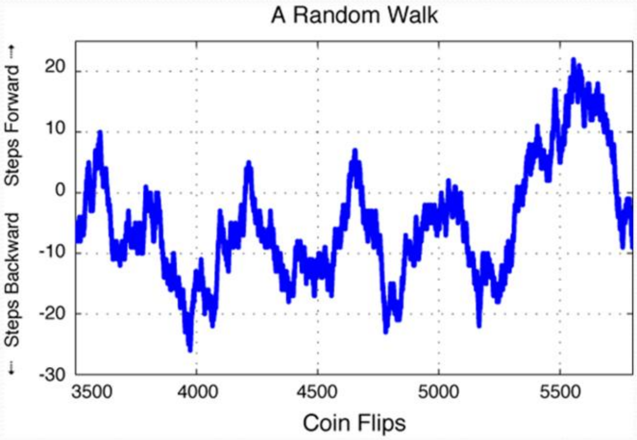
Take-Aways ??

- How does such an approach resonate with the evaluation data that you have?
- Are there ways in which you can incorporate these ideas within the evaluation work you are currently doing?
- Do you see value in data explorations of this type?

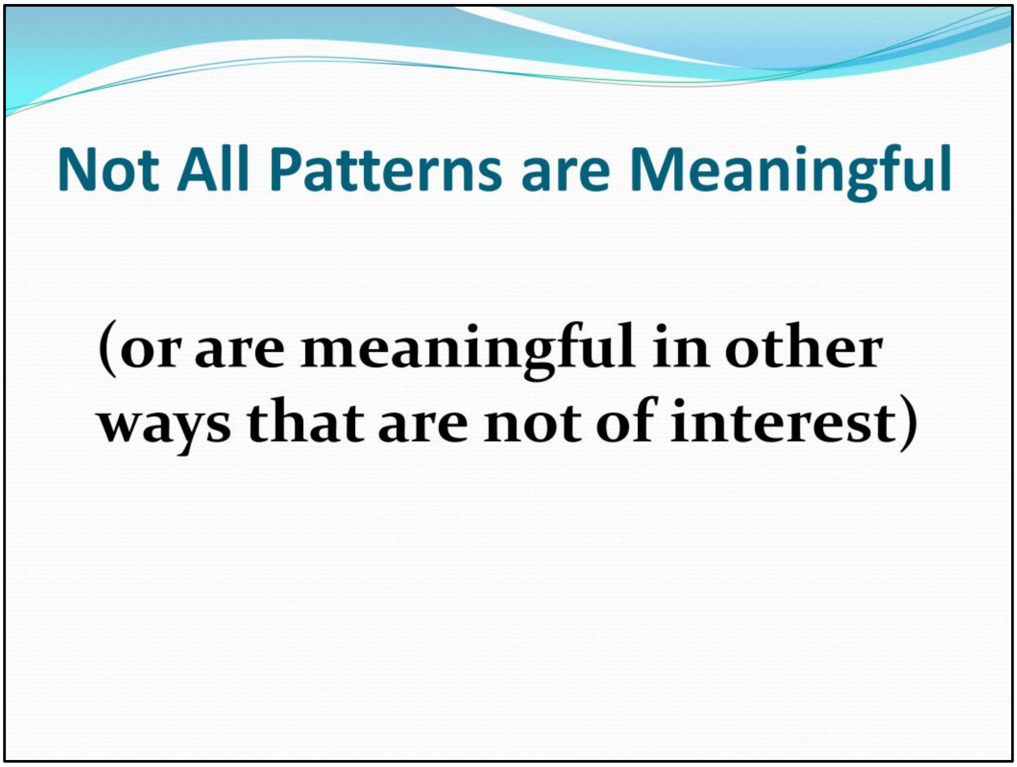
Not All Patterns Are Meaningful

A few caveats …..

Here is an Example of a Random Walk



It looks likes a trend doesn't it? But it is not it is a random walk (taken from Google Images based on "Random Walk Coin Flips"). In a time series analysis a step called differencing would show that this is not a trend; but in the absence of statistical options such as these (e.g., smaller data sets where time series is not appropriate) we could be tempted to see something in these data that are not there.



What I am attempting to highlight here is that consistency in data while encouraging is no guarantee that what we are seeing is meaningful in the way we think it might be.

Common Method Bias

Variance that is attributable to the method used rather than the constructs of interest and that are measured.

Podsakoff, P.M., MacKenzie, S. B., Lee, J. Y. & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology,* 88(5), 879-903.

These slides are intended to convey that I don't think that what we are talking about is a panacea for measurement issues and to recommend that we have to be cognizant of these issues in any approach we take. In the three examples shown, we have used data from other instrument sources to suggest that these patterns are not likely a function of common method bias. These corroborating data come from qualitative data taken from open-ended questions posited in surveys, interviews, and data obtained from case studies.

MBRS-RISE – Minority Biomedical Research Support (MBRS) – Research Initiative for Scientific Enhancement (RISE) program
MARC - Minority Access to Research Careers (MARC)
MHIRT - Minority Health and Health Disparities International Research Training
Each URM program is different but it has a common goal of offering hands-on laboratory research experience to undergraduate students and has this common component; but there are differences too, for example the MARC programs offers academic course work to complement this research experience and offers an off-campus summer research experience; these two programs, MARC and RISE are conducted during the academic year and during the summer; the MHIRT program offers a summer research internship at an oversees laboratory.

First Example (con’t.)

MBRS-RISE – Sophomores

MARC – Juniors

MHIRT – Seniors (some Juniors)

It is important to note that the sample sizes are small here – RISE (about 10 students) MARC (5 students) MHIRT (about 12 students) a year – so we want to cautious in any decisions we make about these data but over time these numbers being to multiple RISE (43 students) MARC (20 students) MHIRT (34 students).

MBRS-RISE Program -- National Institutes of Health (NIH), National Institute of General Medical Science (NIGMS), Grant#R25 GM59218
MARC Program -- National Institutes of Health, National Institute of General Medical Science (NIGMS), award number 5T34GM08388-05; CFDA number 93.859
MHIRT Program – National Institutes of Health, National Institute on Minority Health and Health Disparities Grant (NCMHD) MD001429

First Example (con't.)

Preprogram **Attitudes toward**

Usefulness of Science

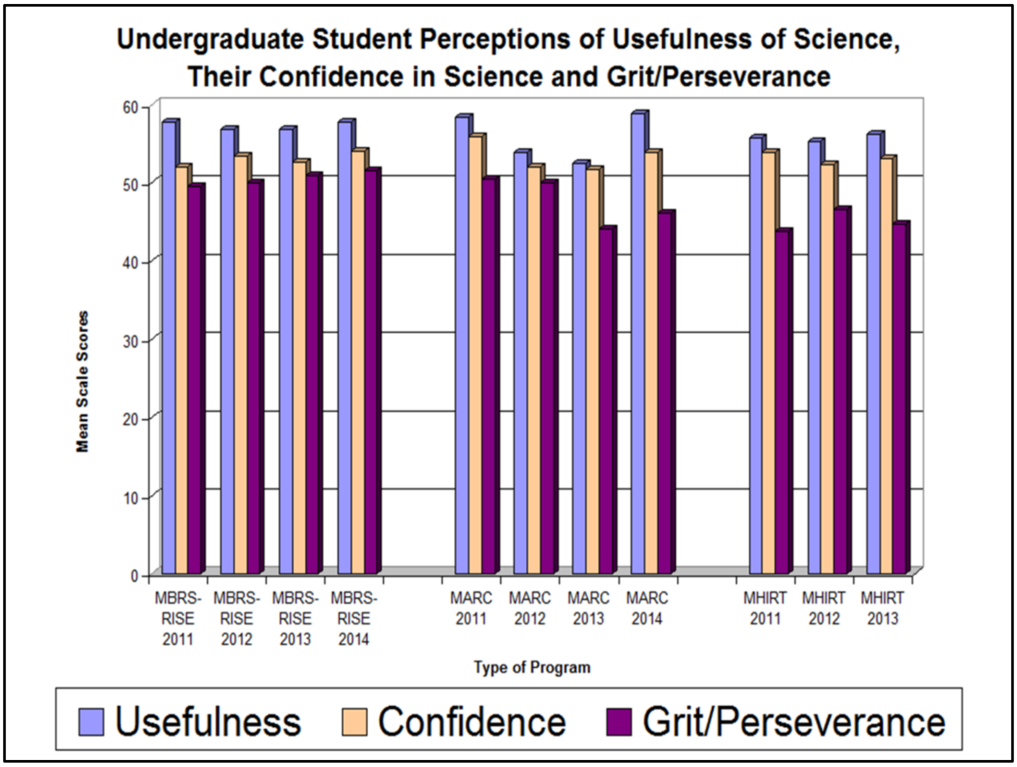
Confidence in Conducting Science

Grit/Perseverance

Each scale based on 12 items using a 5-point Likert-type scale.

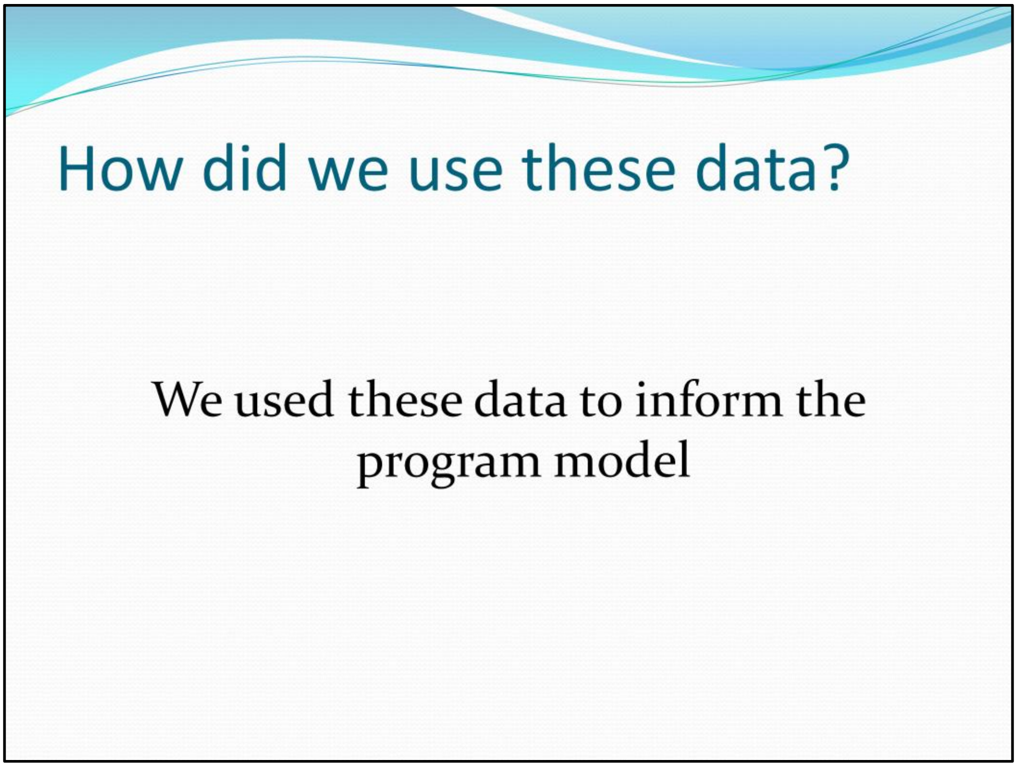
- Adopted from Doepken, D., Lawsky, E., & Padwa, L. Modified Fennema-Sherman Attitude Scales www.woodrow.org
- Duckworth, A.L., Peterson, C., Matthews, M.D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 9*, 1087-1101.

Going into these projects I expected to see a rise in positive attitudes across the three programs as each program begins at a different time in a student's undergraduate experience – RISE program for sophomores; the MARC program beginning during the junior year; and MHIRT at the end of junior/senior year. I expected attitudes to increase as the student progressed through their undergraduate academic/research experiences even those we were looking at different groups of students with mean scores increased for seniors (usefulness, confidence and grit) -- MJIRT compared to less mature students for RISE and MARC students. But this is not what happened ……..



I am most interested in the step-down pattern that is evident in comparing Usefulness, Confidence and Grit scores. This step-down pattern was evident across programs and across program years. And this pattern was supported by statistical analyses. (The higher the score, the more positive the attitudes.)

Based on MANOVA analysis there was a statistically significant program effect for Grit scores (with Grit scores lower for MHIRT students compared to the MBRS-RISE students) [$F_{(8, 74)} = 2.39$, $p < .03$] but no statistically significant program differences based on Usefulness and Confidence scores; nor were there any statistical differences based on program year. When program and program year are collapsed for Usefulness and Confidence scores, these is a statistical difference between usefulness scores (higher than) confidence scores [$F_{(1, 72)} = 43.80$, $p. < .001$]. So, the statistical tests support this step-down pattern.



For a variety of reasons, we did not measures these attitudes in a typical pre-post design. I can explain why this was case in more detail if there is interest.

How did we use these data?

- Explored adding program strategies to help increase students' confidence in conducting science in the laboratory
- Explored program feedback from students on opportunities to gain necessary skills in the lab
- Explored feedback from faculty/mentors on student performance in the lab
- Explored post-program student perceived development through open-ended interview and survey questions.

We explored this pattern *after* the program by asking students (in one of these programs) the following open-ended question, "In your own words, please tell us if or how the MARC program helped you develop your confidence in your ability to conduct science and your ability to go to the next level."

Questions

- Points of Clarity?
- Thoughts about how you might use a similar approach?

Second Example

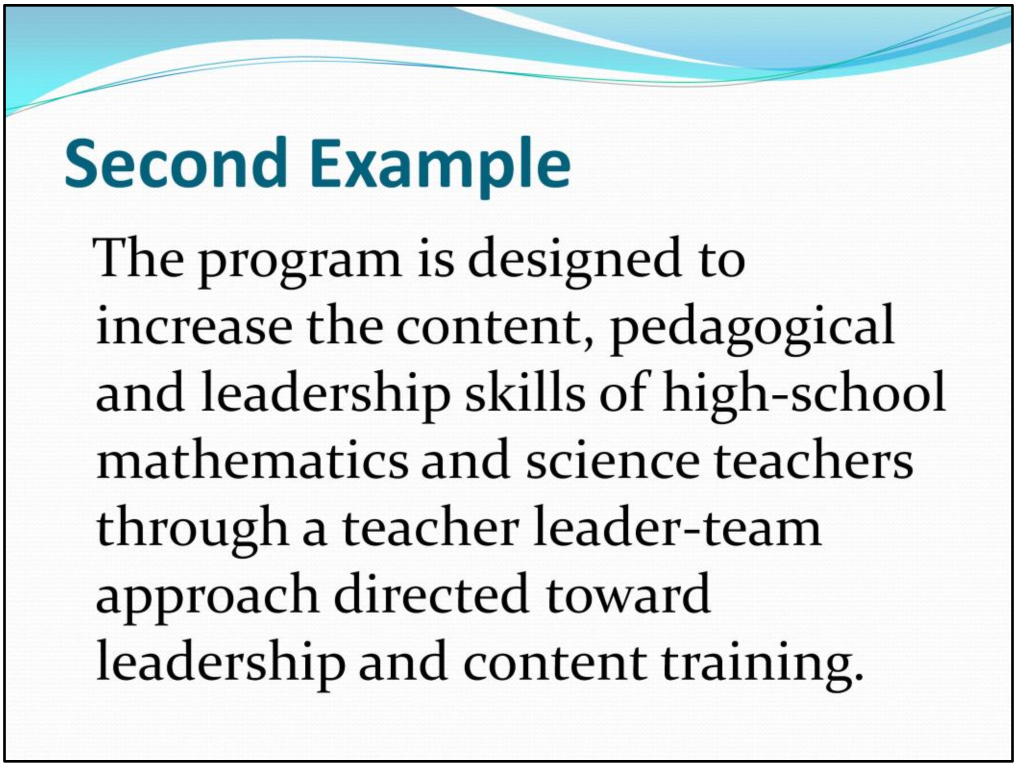
Chicago Transformation Teacher Institutes (CTTI)

CTTI program is funded by a National Science Foundation Mathematics Science Partnership (MSP) grant (NSF-DUE-0928669)

CTTI Steering Committee

- Don Wink, Principal Investigator
- John Baldwin, Co-PI
- Dean Grosshandler, Project Coordinator
- Norm Lederman, Co-PI
- Steven McGee, Co-PI
- Stacy Wenzel, Internal Evaluation Team
- Lynn Narasimhan, Co-PI
- Chandra James, (Co-PI Chicago Public Schools)

External Evaluator, Race & Associates, Ltd.



Chicago Transformation Teacher Institutes (CTTI), a math and science partnership program. CTTI program is funded by a National Science Foundation Mathematics Science Partnership (MSP) grant (NSF-DUE-0928669). This program involved the partnership of five Chicago-based universities: University of Illinois at Chicago, Loyola University Chicago, DePaul University, Illinois Institute of Technology, and Northwestern University plus CPS.

Second Example (con’t.)

- Focuses on how the program model informed the protocol used to observe 12th grade math and science classes ....

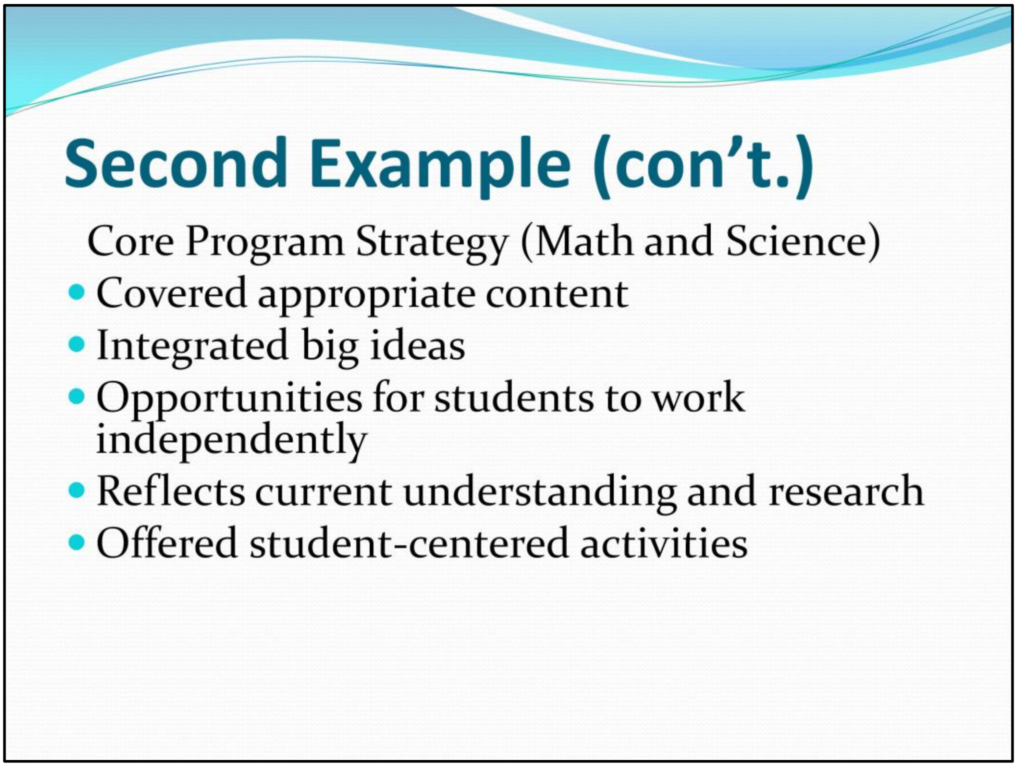
- And how we analyzed these data

This is a very large project with 20 schools and a desired 160 teachers – but the observation component of this project is a much smaller effort within this large program.

Second Example (con't.)

Extracted core program strategies and program outcomes from the program model

For this presentation, I have focused just on the observations which occurred in 12th grade high school *science* courses.



We will focus on just the science classes that were observed. This first set of strategies that were reflective of both math and science.

Full language of strategies:

1. Covered content that is appropriate to the specific discipline in order to prepare students for post-secondary careers and college work in mathematics and/or science.
2. Integrated big ideas in mathematics and/or science.
3. Offered students the opportunity to work individually and collaboratively on meaningful mathematics and/or science.
4. Reflected current understanding and research in mathematics and/or science.
5. Offered student-centered activities, questions, or problems directed by student learning.

Second Example (con't.)

Core Program Strategy/Program Outcome (For Science)

- Content reflects current opportunities and needs within science as a discipline
- Is project-based
- (Capstone) project opportunity is inquiry-based and incorporates inquiry-based activities

These strategies are specific to science. Each of the strategies that extracted from the program model are further defined by a operational definitions and further clarifications.

Second Example (con't.)

Core Program Strategy/Program Outcome (For Science)

- Provides opportunities for teacher to exercise best practice pedagogy
- Emphases nature of science and knowledge about scientific inquiry
- Students actually doing science that demonstrates that research has unknown outcomes, uncertainties, and loose ends

Each of the strategies that extracted from the program model are further defined by a operational definitions and further clarifications.

Second Example (con't.)

Then, we incorporated these in an observation protocol supplement

Second Example (con't.)

Also used the individual items from the RTOP (Reformed Teaching Observation Protocol)

All items rated on a 5-point scale

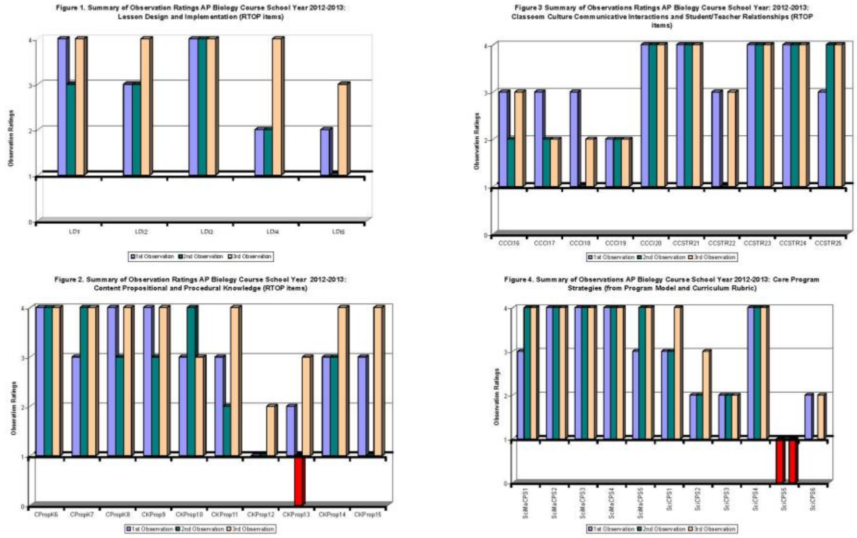
- 0 = Never occurred;
- 4 = Very descriptive of the lesson

And a Cannot Determine/DNA option

We used the RTOP because we wanted to compare these ratings with a more standardized observation protocol or at least one that has been used in field and is well known. We did not, however, use cumulative RTOP scores.

Sawada, D., Piburn, M., Falconer, K., Turley, J., Benford, R., & Bloom, I. (2000). Reformed Teaching Observation Protocol (RTOP). Arizona Collaborative for Excellence in the Preparation of Teachers: Arizona State University.

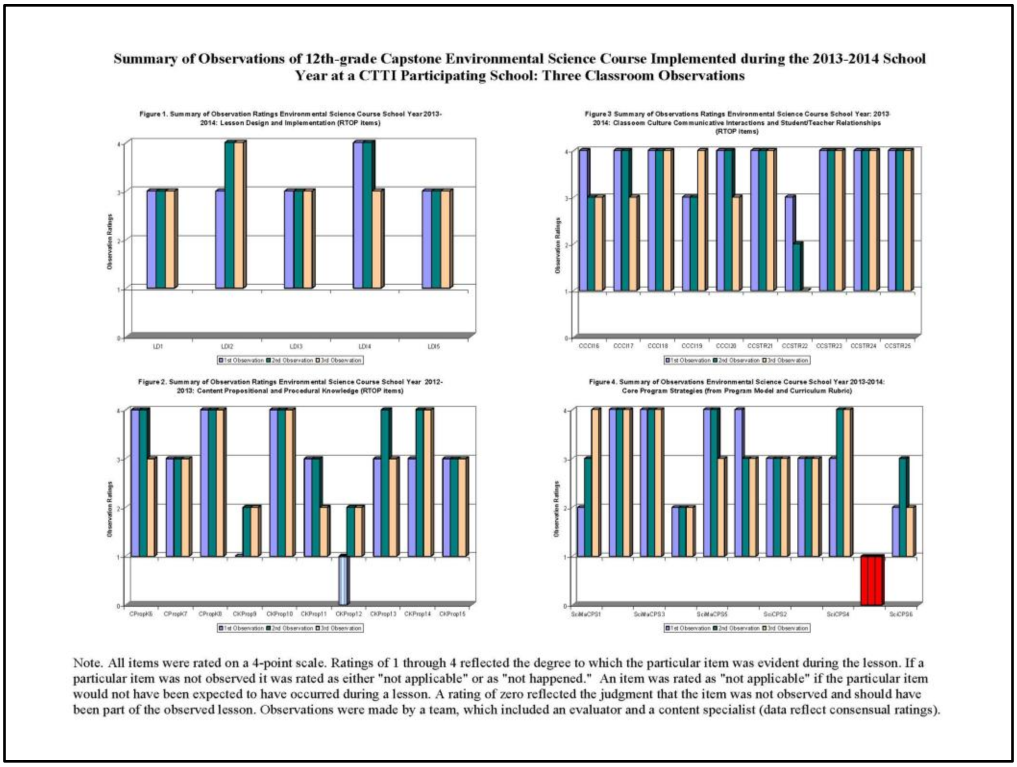
Summary of Observations of 12th-grade Capstone AP Biology Course Implemented during the 2012-2013 School Year at a CTTI Participating School: Three Classroom Observations



Note. All items were rated on a 4-point scale. Ratings of 1 through 4 reflected the degree to which the particular item was evident during the lesson. If a particular item was not observed it was rated as either "not applicable" or as "not happened." An item was rated as "not applicable" if the particular item would not have been expected to have occurred during a lesson. A rating of zero reflected the judgment that the item was not observed and should have been part of the observed lesson. Observations were made by a team, which included an evaluator and a content specialist (data reflect consensual ratings).

The first three graphs show observation ratings (based on consensual ratings of a content specialist and evaluator) reflect RTOP item scores – one graph from each section of the RTOP -- and the graph in the lower right hand corner shows the evidence for core program strategies and program outcomes for these observed lessons. For these graphs we have deliberately raised the xy plane such that any observed strategy (a rating of 1 or above) is above this plane; and a strategy not observed (a rating of 0) is below the plane. Although we are interested in the strategies where there is evidence of their occurrence we are suggesting that the variability of occurrences above the plane may or may not be meaningful. – Might be due to observations based on an individual lesson rather than the full curriculum or rating variations. Those strategies that are below the line were not present in the observed lesson and these strategies are reflected as the red bars in these graphs because we are very interested in these findings. The argument we are making is that it is easier to determine whether or not something occurred than evaluating the level at which that something occurred, when it has occurred. For this 12th grade course, there were three observed lessons.

We have concluded that overall these data suggest a set of very well executed lessons that align with the program strategies emphasized in the CTTI program.



Here is an another example of an observed science lesson. Here we also conclude that these were well executed lessons that integrated program strategies that were emphasized during the CTTI program. Again, for this 12th grade course, there were three observed lessons.

Summary of Observations of 12th-grade Forensic Science Course Implemented during the 2012-2013 School Year at a CTTI Participating School: One Classroom Observation

Figure 1. Summary of Observation Ratings Forenic Science Course School Year 2012-2013: Lesson Design and Implementation (RTOP items)

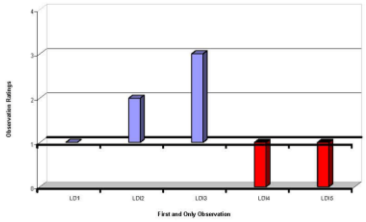


Figure 2. Summary of Observation Ratings Forenic Science Course School Year 2012-2013: Content Propositional and Procedural Knowledge (RTOP items)

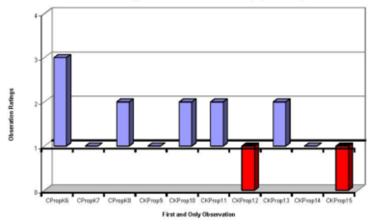


Figure 3 Summary of Observations Ratings Forenic Science Course School Year: 2012-2013: Classoom Culture Communicative Interactions and Student/Teacher Relationships (RTOP items)

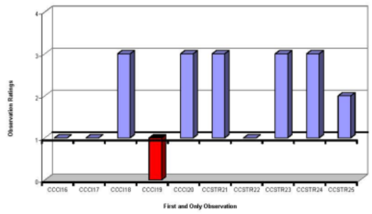
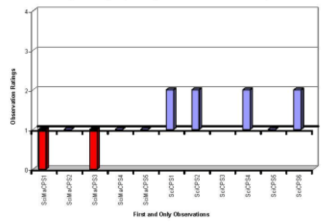


Figure 4. Summary of Observations Forenic Science Course School Year 2012-2013: Core Program Strategies (from Program Model and Curriculum Rubric)



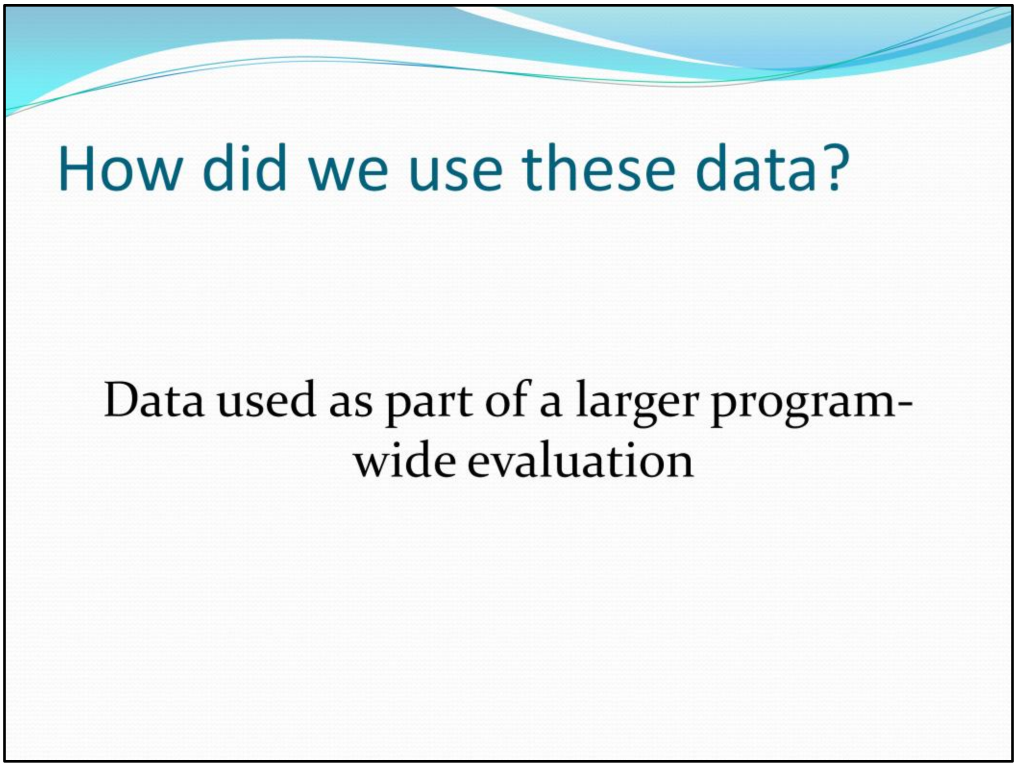
Note. All items were rated on a 4-point scale. Ratings of 1 through 4 reflected the degree to which the particular item was evident during the lesson. If a particular item was not observed it was rated as either "not applicable" or as "not happened." An item was rated as "not applicable" if the particular item would not have been expected to have occurred during a lesson. A rating of zero reflected the judgment that the item was not observed and should have been part of the observed lesson. Observations were made by a team, which included an evaluator and a content specialist (data reflect consensual ratings).

This third graph set is here to show an example of a lesson that were observed from a science class that did not well integrated the program strategies.

Summary of RTOP and CTTI Supplement Scores:
Three 12th-grade Science Courses

Course/Observation	RTOP Scores					CTTI Supplement	
	Sec1	Sec2	Sec3	Sec4	Sec5	Math and Science Strategies	Science Strategies
AP Biology							
Observation 1	15	18	12	15	16	18	13
Observation 2	13	18	16	11	13	19	12
Observation 3	19	19	19	13	14	20	15
Environmental Science							
Observation 1	16	16	12	19	19	16	15
Observation 2	17	17	16	18	18	17	16
Observation 3	16	16	14	17	17	17	15
Forensic Science							
Observation 1	6	9	5	8	12	3	9

If we had just used the section scores from the RTOP and a cumulative scores from our supplement we would have concluded (as we did from the graphs) that the AP Biology and Environmental Science lessons were well executed which was not the case for the Forensic Science lesson which was a poorly implemented lesson. That being said, we would have missed the tendency for both AP Biology and Environmental Science lessons to have missed the opportunity to engage in nature of science discussions despite the fact that these lessons were otherwise well implemented.



Offer the evidence to suggest that the instructional strategies that were emphasized in the CTTI program were reflected in these observed lessons – but still a very small set of observed lessons to draw conclusions; we are collecting more observation data (and we have included observations of math classes we have just elected to focus on science lessons for the same of simplicity of presentation). Going forward, we have been able to determine the level of CTTI-engagement with these observation data based on teachers where we know the actual program attendance, responses to a survey asking teachers about how they have used or adapted either content or pedagogy gained from program participation as well as information gathered from case studies of select participating schools.

Questions

- Points of Clarity?
- Thoughts about how you might use a similar approach?

Third Example

Teaching Evolution through Human Examples (TEtHE)

Smithsonian Institution,

A NSF-funded Discovery Research K-12 Program

Award Number:1119468

Teaching Evolution through Human Examples (TEtHE) program. National Science Foundation Award Number:1119468 Briana Pobiner (Principal Investigator), Richard Potts (Co-Principal Investigator) and William Watson (Co-Principal Investigator) and Race & Associates, Ltd. External Evaluator.

Third Example

- Briana Pobiner, Principal Investigator
- Richard Potts (Co-Principal Investigator)
- William Watson (Co-Principal Investigator)

External Evaluator, Race & Associates, Ltd.

Third Example

Purpose:

Development of four curriculum supplements for AP Biology that uses human examples in the teaching of evolution

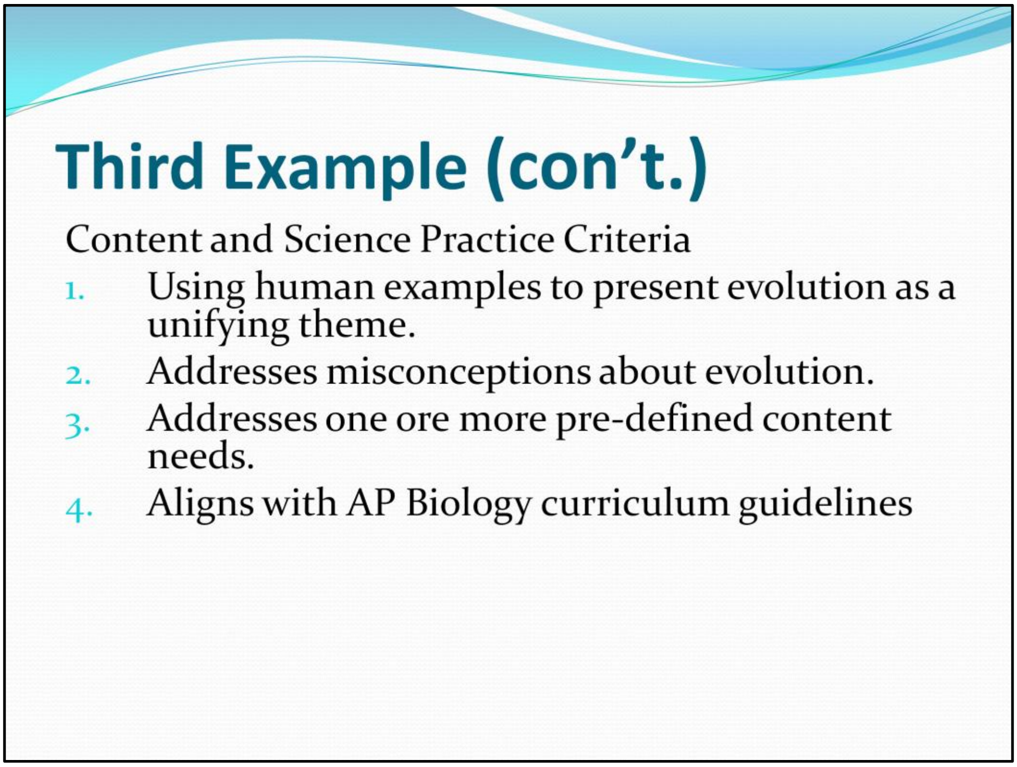
The purpose of the Teaching Evolution through Human Examples (TEtHE) program is to develop curriculum supplements and teaching strategies for use in high school AP Biology focused on human evolution and assess how their use affects the understanding, learning and teaching of evolution.

Third Example (con't.)

Using Core Program Strategies as Criteria to Determine

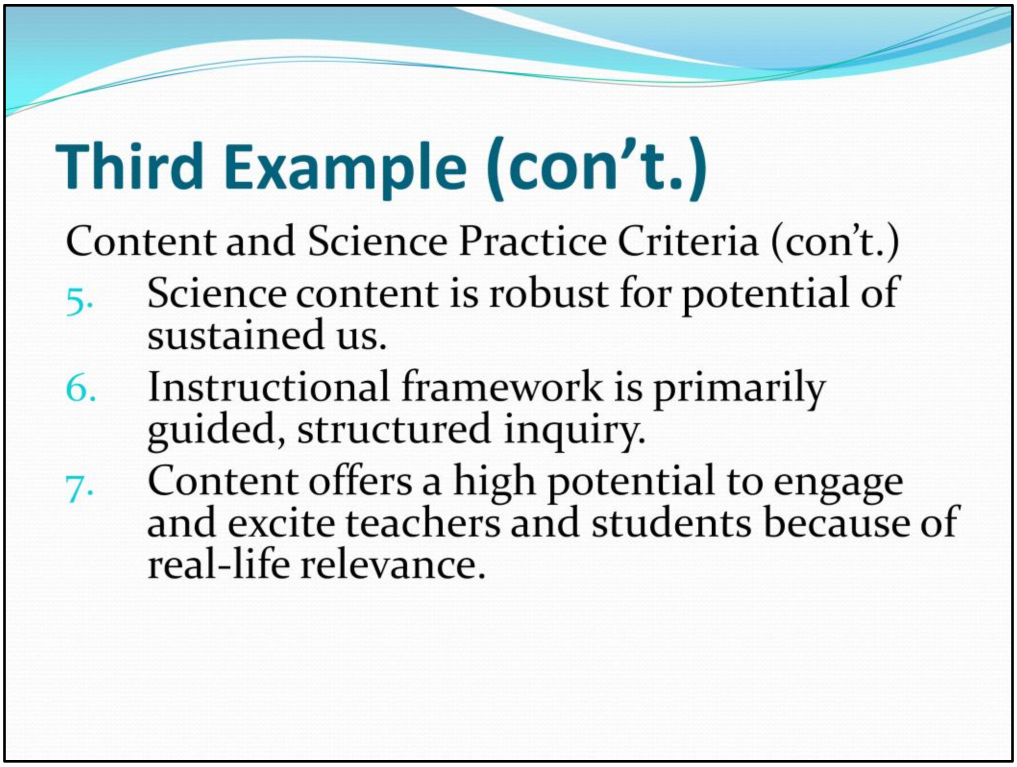
- Program Design Fidelity
- Program Implementation Fidelity

These core program strategies were articulated in the TEtHE program model. Four curriculum supplements were developed and tested: Altitude Adaptation, Malaria, Skin Color, and What Does it Mean to Be Human, the latter of which uses the resources of the National Museum of Natural History of the Smithsonian including virtual use of early humanoid skulls.



Content and Science Practice Criteria (actual wording)

1. Uses human evolution as instructional content and context for presenting the big idea of evolution as a unifying them.
2. Addresses common teacher and/or student misconceptions about evolution when appropriate.
3. Addresses one or more pre-defined content needs (i.e., evolution, mutation, natural selection, extinction, phylogenetics, genetics).
4. Aligns with AP Biology curriculum guidelines (i.e., enduring understandings, science practices, and learning objectives.
5. Incorporates science content that is sufficiently robust of the potential of sustained use (i.e., science content is well-accepted enough not to be speculative and not likely to change substantially in the near future.
6. Instructional framework is primarily guided, structured inquiry that incorporates important components of the nature of science.
7. Presents content that offers a high potential to engage and excite teachers and students because it is relevant to their lives.



Content and Science Practice Criteria (actual wording)

1. Uses human evolution as instructional content and context for presenting the big idea of evolution as a unifying them.
2. Addresses common teacher and/or student misconceptions about evolution when appropriate.
3. Addresses one or more pre-defined content needs (i.e., evolution, mutation, natural selection, extinction, phylogenetics, genetics).
4. Aligns with AP Biology curriculum guidelines (i.e., enduring understandings, science practices, and learning objectives.
5. Incorporates science content that is sufficiently robust of the potential of sustained use (i.e., science content is well-accepted enough not to be speculative and not likely to change substantially in the near future.
6. Instructional framework is primarily guided, structured inquiry that incorporates important components of the nature of science.
7. Presents content that offers a high potential to engage and excite teachers and students because it is relevant to their lives.

First Used to Anchor Expert Panel Review

Expert Panel reviewed curriculum supplements:

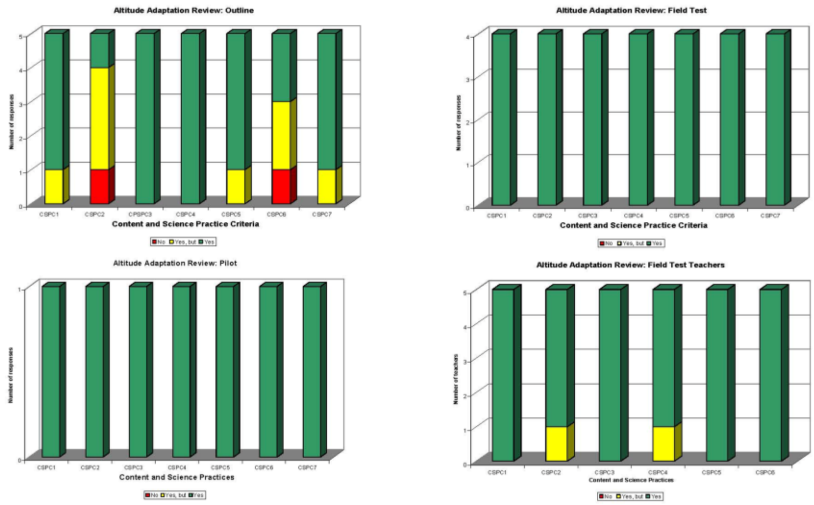
- Outline
- Pilot Test
- Field Test

And reviewed by Field Test Teachers

Each curriculum supplement was also reviewed by an expert panel for each curriculum at the following stages of its development: at Outline, at Pilot, at Field Test. These reviews were based on a different 4-point scale, “Yes,” “Yes, but” “No” and “Unsure.” These observations were conducted by the PI of the project, Briana Pobiner, and the curriculum specialist, Paul Beardsley of the TEtHE project using a protocol specifically designed for this purpose.

Field teachers used the following scale: “Just Right,” “Not Enough,” “Too Much.”

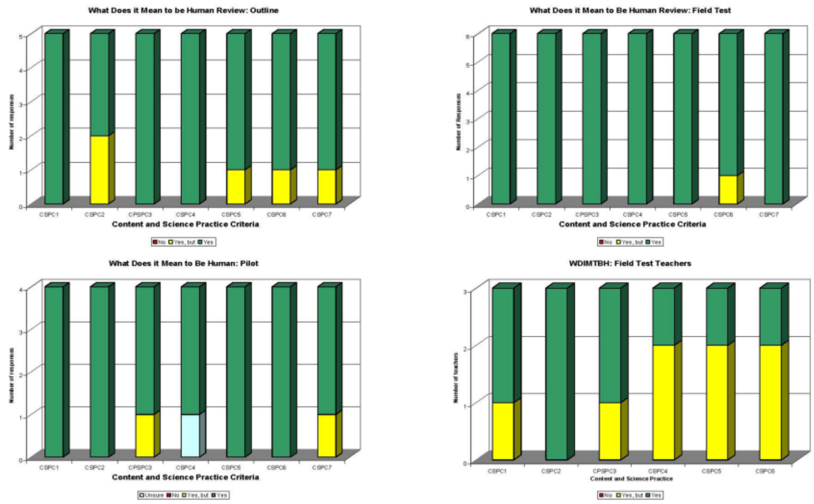
Figure Set Summary of Review of Altitude Adaptation TEtHE Curriculum Supplement by Expert Panel and Field Test Teachers



Note. All items were rated on a 4-option scale, "Yes;" "Yes, but needs a bit of work or fine tuning;" "No" and "Unsure." Content and Science Practice Criteria are: 1. Uses human evolution as instructional content and context for presenting the big idea of evolution as a unifying theme. 2. Addresses common teacher and/or student misconceptions about evolution when appropriate. 3 Addresses one or more pre-defined content needs (evolution, mutation, natural selection, extinction, phylogenetics, genetics. 4. Aligns with AP Biology curriculum guidelines (i.e., enduring understandings, science practices, and learning objectives). 5. Incorporates science content that is sufficiently robust for the potential of sustained use. 6. Instructional framework is primarily guided, structured inquiry that incorporates important components of the nature of science. 7. Present content that offers a high potential to engage and excite teachers and students because it is relevant to their lives. Members of the expert panel reviewed these materials at various points in their development: at Outline, Pilot Test, and Field Test. Teachers reviewed these materials as these were implemented during the Field Test.

Here is an example of one of the curriculum supplements reviewed by expert panel members at the various stages of its development and by the field test teachers who implemented the curriculum. All reviews were based on a 4-point scale, “Yes,” “Yes, but” “No” and “Unsure.” For the first three figures – these data suggest that there was fidelity to these criteria at the program *design* level (based on expert panel review). Figure 4 (lower right hand corner) suggests the fidelity to these criteria as implemented at the field test level (as reviewed by field test teachers). These data were supported by open-ended responses on the review form itself, email discussions, and responses from teacher group-interviews.

Figure Set Summary of Review of WDIMTBH TEtHE Curriculum Supplement by Expert Panel and Field Test Teachers



Note. All items were rated on a 4-option scale, "Yes;" "Yes, but needs a bit of work or fine tuning;" "No" and "Unsure." Content and Science Practice Criteria are: 1. Uses human evolution as instructional content and context for presenting the big idea of evolution as a unifying theme. 2. Addresses common teacher and/or student misconceptions about evolution when appropriate. 3 Addresses one or more pre-defined content needs (evolution, mutation, natural selection, extinction, phylogenetics, genetics. 4. Aligns with AP Biology curriculum guidelines (i.e., enduring understandings, science practices, and learning objectives). 5. Incorporates science content that is sufficiently robust for the potential of sustained use. 6. Instructional framework is primarily guided, structured inquiry that incorporates important components of the nature of science. 7. Present content that offers a high potential to engage and excite teachers and students because it is relevant to their lives. Members of the expert panel reviewed these materials at various points in their development: at Outline, Pilot Test, and Field Test. Teachers reviewed these materials as these were implemented during the Field Test.

Here is a summary of the reviews of the WDIMTBH curriculum supplements at three stages for expert panel review (program design) and for the field test teachers' review (program implementation). There was a great deal of discussion by expert panel members, senior personnel and the PI based on responses on the actual review form, emails and phone conversations. In addition, (most) field-test teachers participated in group interviews where we discussed what and how the curriculum supplements were implemented.

Third Example (con't.)

Also used the individual items from the RTOP (Reformed Teaching Observation Protocol)

All items rated on a 5-point scale

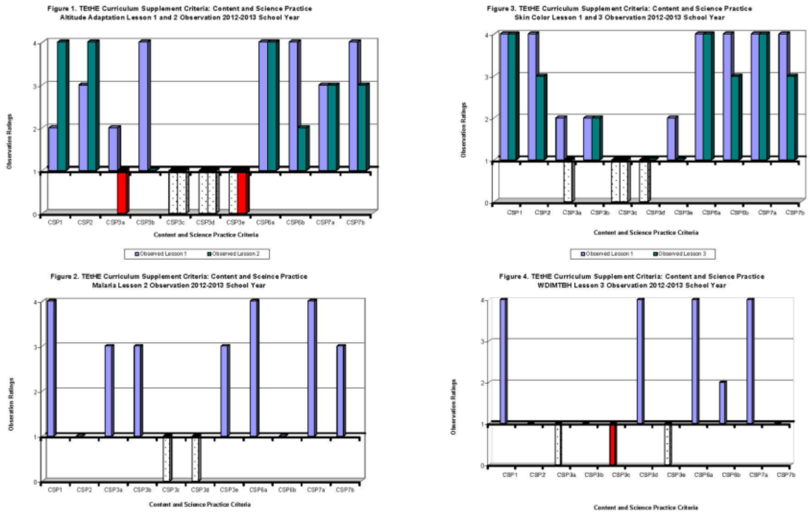
0 = Never occurred;

4 = Very descriptive of the lesson

And a Cannot Determine/DNA

Observations used a different scale (based on the RTOP). These observations were conducted by the PI of the project, Briana Pobiner, and the curriculum specialist, Paul Beardsley of the TEtHE project using a protocol specifically designed for this purpose.

Pilot Test of TEtHE Curriculum Supplement Lesson Observation Ratings during the 2012-2013 School Year



Note. All items were rated on a 4-point scale. Ratings of 1 through 4 reflected the degree to which the particular item was evident during the lesson with a rating of 4 assessing that lesson as "Very Descriptive" for that criterion. If a particular item was not observed it was rated as either "cannot determine" or as "never occurred." An item was rated as "cannot determine" if the particular item would not have been expected to have occurred during a lesson. A rating of zero reflected the judgment that the item was not observed and could (perhaps should) have been part of the observed lesson. Not all content criteria are expected to be covered in a specific lesson; when this is the case a patterned bar is shown reflecting that the lesson did not cover this content and was not expected to cover it. Observations were made by the project's PI.

For the observations we broke out what content needs were covered in the individual lesson that was observed. These areas again our: evolution, mutation, natural selection, extinction, phylogentics, and genetics. If there is a pattern to an observation that falls below the xy plane of these graphs it means that the strategy was not observed and not expected to be observed. Those strategies below the plane where a strategy was not observed and should have been evident – these are shown with red bars. These observations were conducted by the PI of the project, Briana Pobiner, and the curriculum specialist, Paul Beardsley of the TEtHE project using a protocol specifically designed for this purpose.

Field Test of TEtHE Curriculum Supplement Lesson Observations during the 2013-2014 School Year

No altitude lessons were observed

Figure 3. TEtHE Curriculum Supplement Criteria: Content and Science Practice Skin Color Field Test Lessons 2, 4 and 5

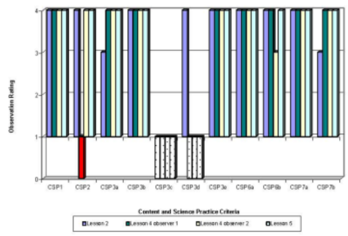


Figure 2. TEtHE Curriculum Supplement Criteria: Content and Science Practice Malaria Field Test Lessons 1 and 3

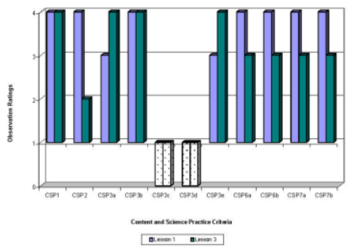
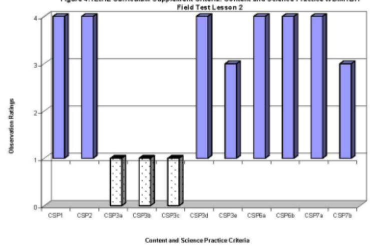
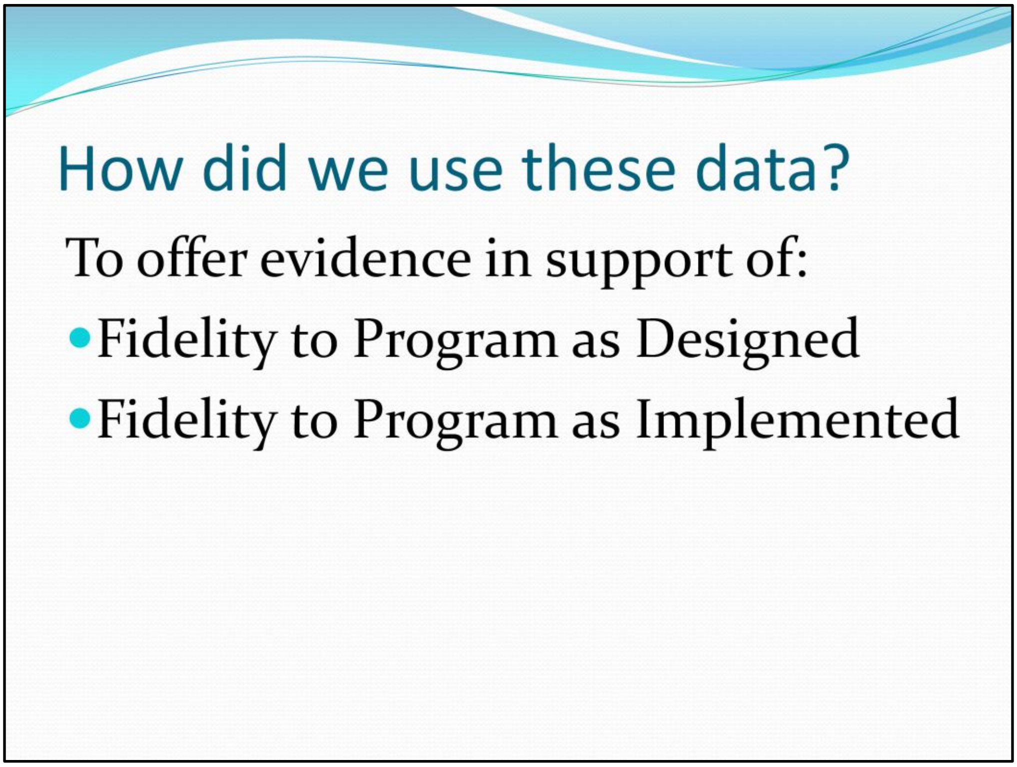


Figure 4. TEtHE Curriculum Supplement Criteria: Content and Science Practice WDIMTBH Field Test Lesson 2



Note. All items were rated on a 4-point scale. Ratings of 1 through 4 reflected the degree to which the particular item was evident during the lesson with a rating of 4 assessing that lesson as "Very Descriptive" for that criterion. If a particular item was not observed it was rated as either "cannot determine" or as "never occurred." An item was rated as "cannot determine" if the particular item would not have been expected to have occurred during a lesson. A rating of zero reflected the judgment that the item was not observed and could (perhaps should) have been part of the observed lesson. Not all content criteria are expected to be covered in a specific lesson; when this is the case a patterned bar is shown reflecting that the lesson did not cover this content and was not expected to cover it. Observations were made by the project's PI and senior curriculum development personnel.

Here are the observation results at field test. These observations were conducted by the PI of the project, Briana Pobiner, and the curriculum specialist, Paul Beardsley of the TEtHE project using a protocol specifically designed for this purpose.



Note. Other data were collected and analyzed as part of this project including feedback from students, student focus groups and measures of attitudes toward evolution by students and student achievement. We have used responses from participating field-teachers to help us interpret these fidelity data as well as to help us better understand differences obtained (by teacher) when student attitudes toward evolution and their content knowledge was assessed.

Questions

- Points of Clarity?
- Thoughts about how you might use a similar approach?

Take-Aways ??

- How does such an approach resonate with the evaluation data that you have?
- Are there ways in which you can incorporate these ideas within the evaluation work you are currently doing?
- Do you see value in data explorations of this type?

Looking for Patterns in Data Obtained from Program-theory Model Based Evaluations

Kathryn Race
Race & Associates, Ltd.
race_associates@msn.com
www.raceassociates.com

Thank you!