Formative Assessment in Elementary Science: A Multi-level Path Analysis

Yunyun Dai, Joan Herman, Ellen Osmundson

University of California, Los Angeles

AERA 2013 Annual Meeting

Division H - Research, Evaluation and Assessment in Schools
San Francisco, CA

April 28th, 2013

National Center for Research
on Evaluation, Standards, & Student Testing

CRESST

UCLA | Graduate School of Education & Information Studies | cresst.org

Outline

- **Study context**

- **Research questions**

- **Study samples**

- **Study methods, analyses and findings**

- **Conclusions**

Study Context

1. Large-scale, multi-state, randomized controlled study of the effects of adding formative assessment tools and strategies to an elementary science curriculum

2. Schools (and teachers within them) randomly assigned to treatment and control groups

3. All teachers had prior experience with curriculum

   ✓ *All received professional development (2 days)*

   ✓ *All taught 2 curriculum modules*

4. Treatment re-conceptualized as one element of formative assessment: availability of quality tools

National Center for Research on Evaluation, Standards, & Student Testing

CRESST

Treatment vs. Control

- **Treatment Teachers**: Full Implementation of ASK/FOSS Embedded Assessment System

  Instruction, Embedded Assessments, Benchmark Assessments, Next-Step Strategies, Study Group, weekly logs, Professional development

- **Control Teachers:** Full Implementation of FOSS System

  Instruction, Assessment, weekly logs, Professional development.

National Center for Research on Evaluation, Standards, & Student Testing

Research Questions

1. How does formative assessment influence teacher content-assessment knowledge and student learning?

2. How does formative assessment influence student learning?

3. What are the inter-relationships among and between teacher knowledge, availability of formative tools, use of assessment, and student learning?

CRESST

Study Variables & Instrumentation

Students
- Knowledge of magnetism and electricity: pre-post measure
- Demographics and prior achievement state, grade level: archival data

Teachers
- Quality assessment tools: treatment condition
- Teacher content knowledge: pre-post measure
- Intensity of curriculum implementation: weekly teacher log
- Frequency of on-going assessment: weekly teacher log

National Center for Research on Evaluation, Standards, & Student Testing

CRESST

Teacher Content Assessment Knowledge

1.22 Anne is investigating objects and magnets. She made this observation in her science journal.
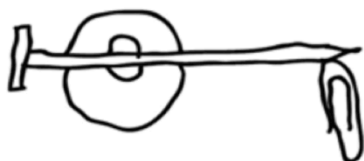
"I was surprised! A nail was stuck to the magnet. When I accidentally touched the nail to a paper clip, the paper clip stuck to the nail. I wonder why that happened?"

a. Explain to Anne why the paper clip stuck to the nail. Use diagrams or pictures if necessary.

Anne and her friend were asked by her teacher why they thought the paper clip stuck to the nail. Here are their responses to the question:

Anne's response: The paper clip turned into a magnet too.

Anne's friend's response: The nail gets stuck on the magnet, and the nail turns into a magnet, so the paper clip can stick on the nail.

b. What inferences can you draw about the students' understanding of magnetism and electricity? What do these students know? What do these students not know/need to learn?

c. If these students were in your class, what would you do next in your instruction to help the students learning progress?

CRESST

Teacher Weekly Log

Teacher Log for the Week of 09/07/2009

☐ I did not teach FOSS/ASK this week

General

1a) Which Investigation/s did you work on this week? (check the appropriate boxes)

Water:

☐ Survey (pretest)	☐ 3.2
☐ 1.1	☐ 3.3
☐ 1.2	☐ 3.4
☐ 1.3	☐ 4.1
☐ 2.1	☐ 4.2
☐ 2.2	☐ 4.3
☐ 2.3	☐ 4.4
☐ 3.1	☐ Posttest

(This information will be made available to support staff and contractors to allow them to better support you.)

1b) On which days did you teach FOSS/ASK this week? ☐ Mon ☐ Tue ☐ Wed ☐ Thu ☐ Fri

1c) On the days that you taught science, approximately how many minutes did you spend teaching FOSS/ASK? (record number of minutes in each box)

Mon	Tue	Wed	Thu	Fri

1d) This week, approximately how many minutes each day did you spend looking at student work after teaching FOSS/ASK? (record number of minutes in each box)

Mon	Tue	Wed	Thu	Fri	Wkend

This week, during FOSS/ASK instruction, how many days did you engage in the following activities?

Resources — DAYS USED THIS WEEK

2a) Used the "At a Glance" to review focus (investigation) questions, science content, and assessment opportunities for your teaching — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

2b) Used the "ASK Teacher Guide Insert Pages" to guide instruction and formative assessment practices — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Assessments — DAYS USED THIS WEEK

3a) Planned and used an **embedded assessment** for the lesson (e.g., notebook sheet, notebook entry, response sheet, observation) — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3b) Analyzed student work in **science notebooks** — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3c) Analyzed student work on the **response sheets** — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3d) Analyzed **observations** of students — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3e) Analyzed student work for **patterns and trends** — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3f) Recorded and used assessment information on an **informal data chart** — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3g) Provided **feedback to individual students** based on analysis of student work — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3h) Used sticky notes, conferences, etc. to provide **individual feedback** to students based on analysis of student work — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3i) Provided **feedback to the entire class** based on analysis of student work — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3j) Selected and used a **next-step strategy** — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3k) **Retaught** content based on analysis and interpretation of student work — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

3l) What did you learn about students' understanding of science concepts from your analysis of student work? Please provide examples and specific details.

Benchmark Assessments for Investigations — DAYS USED THIS WEEK

4a) Checked on student understandings at the end of an Investigation — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4b) Engaged students in self-assessment of science learning — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4c) Administered an I-Check Benchmark Assessment — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4d) Used **coding guides** in the Benchmark Folio to code I-Check items — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4e) Recorded I-Check codes on the "Summary Coding Sheets" — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4f) Conducted **student self-assessment sessions** based on I-Check analysis — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4g) Checked students' reflections after self-assessment — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4h) Used a **next-step strategy** based on self-assessment sessions — ○ N/A ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

4i) Describe the self-assessment activities you provided for the class.

4j) What did you learn about students' understanding about science concepts based on information from the I-Checks? Please provide examples and specific details.

Comments

5a) What percentage of your students do you think understand the core concepts of the Investigation(s) you taught this week?

5b) Do you have any questions or feedback about your experience with the project this week? This question is not confidential and responses will be made available to support staff and contractors to allow them to better support you.
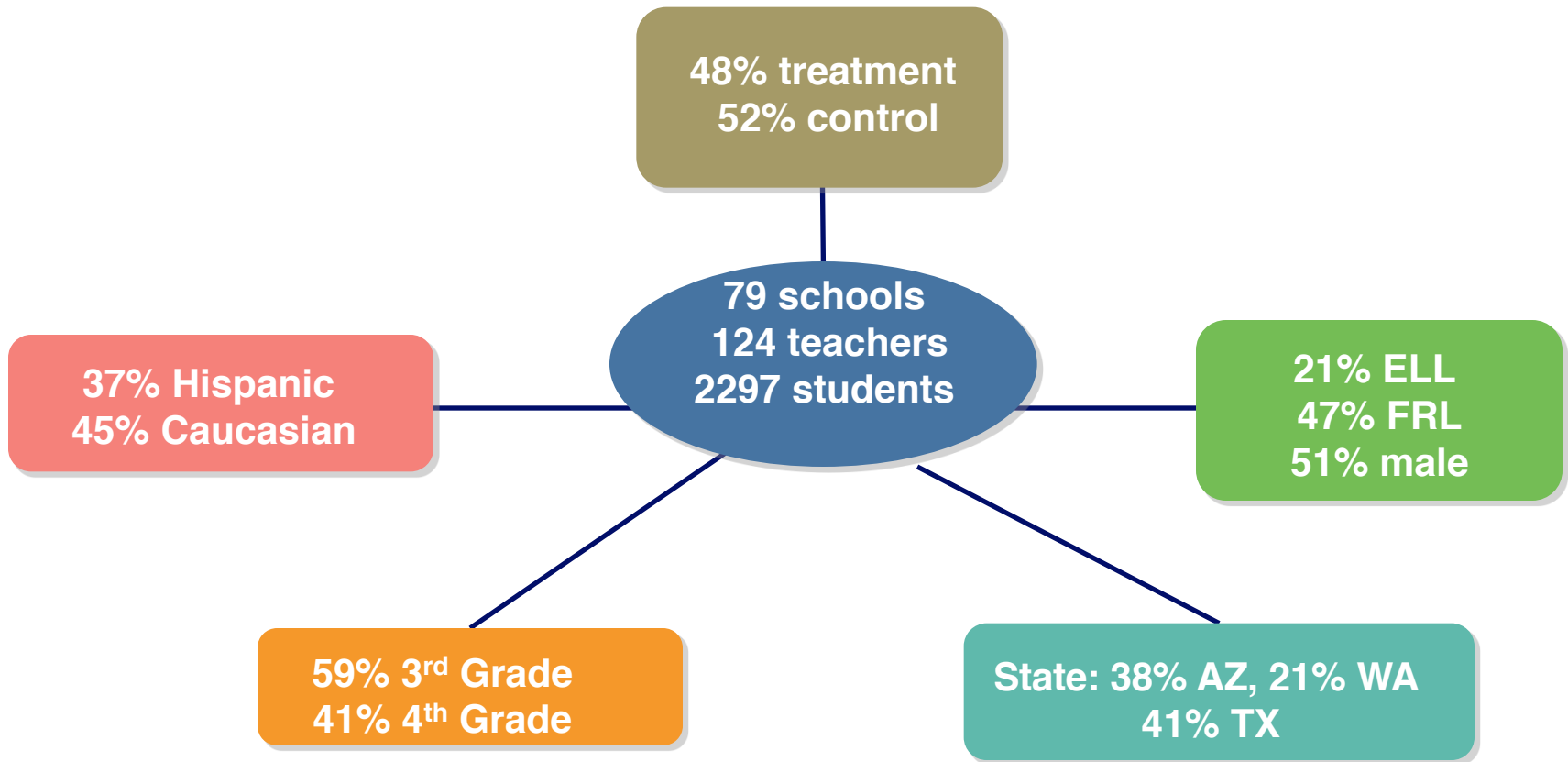
FOSS/ASK Study Group

6) Did you meet with your FOSS/ASK Study Group this week? ○ YES ○ NO

Once you submit this log you will not be able to go back and edit it. Please make sure all your answers are entered correctly!

[Submit Teacher Log]

...ndards, & Student Testing

CRESST

Description of Study Sample

48% treatment
52% control

37% Hispanic
45% Caucasian

79 schools
124 teachers
2297 students

21% ELL
47% FRL
51% male

59% 3rd Grade
41% 4th Grade

State: 38% AZ, 21% WA
41% TX

National Center for Research on Evaluation, Standards, & Student Testing

CRESST

Descriptive Results

	N	Mean	Std. Deviation
Teacher content survey			
Teacher pretest score	117	20.78	7.43
Teacher posttest score	117	30.35	6.99
Teacher weekly log			
Total number of teacher logs	117	7.62	2.47
Total number of days teaching AF per week	117	3.13	0.79
Average minutes of teaching AF	117	49.5	12.45
Average minutes of reviewing student work	117	9.54	5.72
Student M&E test scores			
Student pretest score (M&E)	2297	18.72	5.15
Student posttest score (M&E)	2297	30.34	5.33

National Center for Research on Evaluation, Standards, & Student Testing

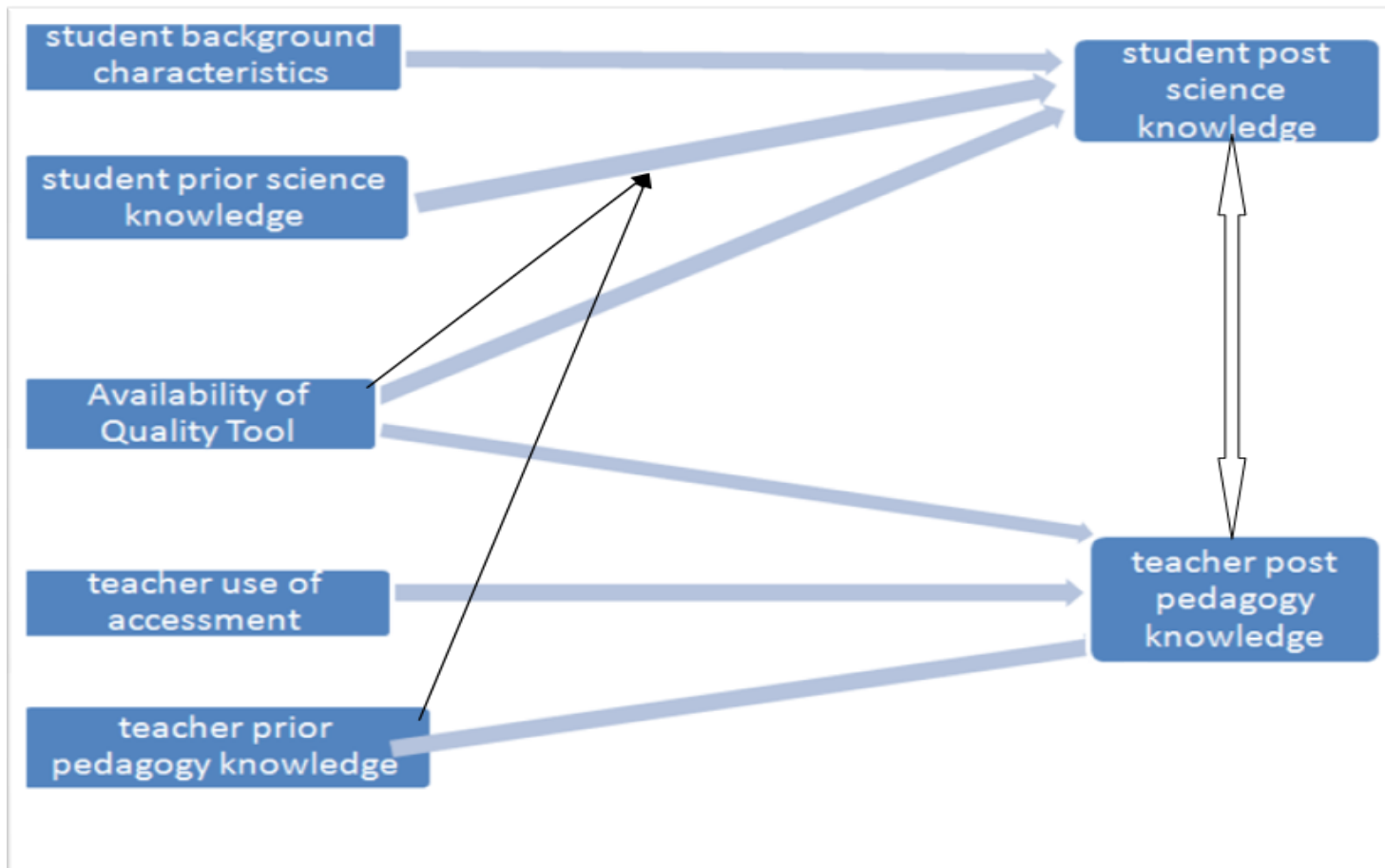# Study Analysis Methods

Prior analysis: Multilevel regression analyses (HLM)

- separate analyses of each student and teacher outcomes

Current analysis: Multilevel path analysis

- combined analyses of both student and teacher outcomes.

- explored the inter-relationships among student science learning with multiple teacher measures

CRESST

Conceptual Model of Pedagogy, Content, and Learning

National Center for Research on Evaluation, Standards, & Student Testing

CRESST

Analyses and Findings

Parameter	Estimate	S.E.	Est./S.E.	P-Value
Log-likelihood				-2691.81
Number of Parameter				26
AIC				5435.62
BIC				5584.85
Parameter	**Estimate**	**S.E.**	**Est./S.E.**	**P-Value**
Within Level				
student post M&E score	ON			
Hispanic	0.03	0.03	1.18	0.24
Caucasian	0.11	0.02	4.42	0.00
English-language learner	-0.06	0.03	-2.04	0.04
Free/reduced lunch	-0.07	0.02	-3.10	0.00
male	0.02	0.02	1.24	0.22
Residual Variances				
student post M&E score	0.53	0.02	21.67	0.00
Between Level				
level 1 random slope	**ON**			
treatment	-0.03	0.03	-1.05	0.29
teacher pedagogy pretest	0.02	0.02	0.77	0.44
teacher pedagogy posttest	**ON**			
treatment	0.31	0.08	3.93	0.00
teacher pedagogy pretest	0.42	0.08	5.11	0.00
number of teacher log	-0.05	0.09	-0.63	0.53
total days of teacher AF per week	-0.18	0.09	-2.05	0.04
minutes of teaching AF per day	-0.18	0.06	-2.84	0.01
minutes of reviewing student work per day	0.10	0.08	1.17	0.24
4th grade	0.25	0.11	2.40	0.02
Arizona	0.09	0.11	0.81	0.42

student post M&E score	**ON**			
treatment	0.16	0.05	3.33	0.00
4th grade	-0.08	0.07	-1.10	0.27
Arizona	0.11	0.07	1.56	0.12
student post M&E score	**WITH**			
teacher pedagogy posttest	0.05	0.05	1.18	0.24
Intercepts				
teacher pedagogy posttest	0.06	0.08	0.68	0.49
student post M&E score	0.01	0.05	0.17	0.87
level 1 random slope	0.30	0.03	11.81	0.00
Residual	**Variances**			
teacher pedagogy posttest	0.65	0.09	7.47	0.00
student post M&E score	0.25	0.04	6.87	0.00
level 1 random slope	0.02	0.01	2.00	0.05

National Center for Research on Evaluation, Standards, & Student Testing
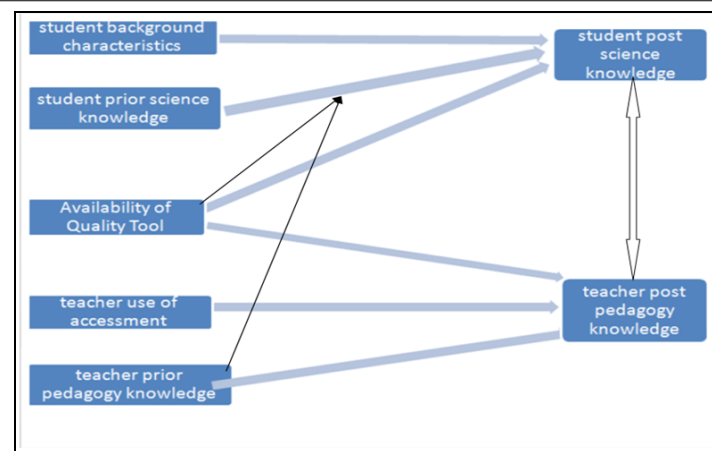
# Estimated Effects of Student and Teacher Variables on Student Post Test

- Caucasian: 0.11 (0.00)
- English-language learner: -0.06 (0.04)
- Free/reduced lunch: -0.07 (0.00)

- Access to Quality Assessment Tools:0.16 (0.00)

National Center for Research on Evaluation, Standards, & Student Testing
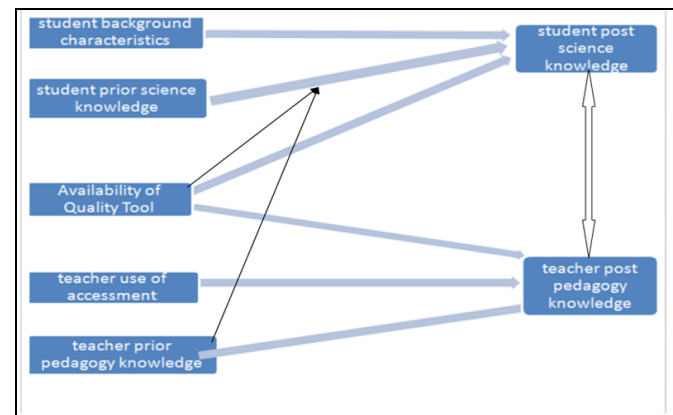
Estimated Effects of Teacher Variables on Teacher Content Assessment Posttest

- Access to Quality Assessment Tools: 0.31, (0.00)
- Teacher content assessment pretest: 0.42 (0.00)
- Total days of teaching AF per week: -0.18 (0.04)
- Ave. minutes of teaching AF per day: -0.18 (0.01)

- 4th grade: 0.25 (0.02)

CRESST
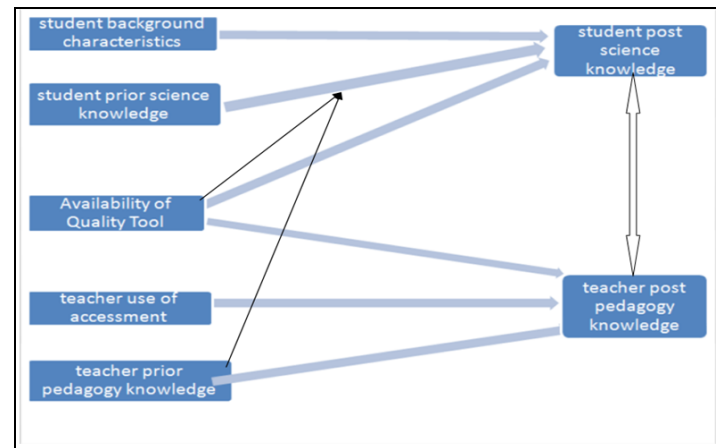
Conclusions

- Availability of quality assessment tools
    - leads to significant improvement in teachers' content knowledge
    - significantly impacts students' science learning.

- Study findings underscore the value of quality, curriculum embedded, formative assessment tools in supporting both student and teacher learning.

- Importance of formative assessment components in curriculum development

- Implication of curriculum selection

CRESST

For More Information

• Dai, Y., Herman, J., Osmundson, E., & Chai, Y., (2013) *The Role of Formative Assessment in Student Learning: An Upper Elementary School Example.* Presentation at Annual meeting of the American Educational Research Association, San Francisco, CA, May, 2013

• Herman, J., Osmundson, E., Dai, Y., Ringstaff, C., Timms, M. (2011) Relationships between teacher knowledge, assessment practice and student learning: Chicken, egg or omelet? *CRESST Technical Report #809.* Los Angeles, CA: CRESST.

• Herman, J., Osmundson, E. & Silver, D (2010). Capturing quality in formative assessment practice: Measurement challenges. *CRESST Report #770.* Los Angeles, CA: CRESST.

• Muthén, B., Asparouhov, T (2009) Beyond Multilevel Regression Modeling: Multilevel Analysis in a General Latent Variable Framework. In Handbook of Advanced Multilevel Analysis. J. Hox & J.K Roberts (eds), Taylor and Francis.

• Osmundson, E., Herman, J., Ringstaff, C., Dai, Y., & Timms, M. (2012). Measuring fidelity of implementation – Methodological and conceptual issues and challenges. (Technical Report No. 811). Los Angeles: UCLA / CRESST.

• Osmundson, E., Dai, Y., & Herman, J. (2011). Year 3 ASK/FOSS efficacy study. (Technical Report No. 782). Los Angeles: UCLA / CRESST.

National Center for Research on Evaluation, Standards, & Student Testing

National Center for Research
on Evaluation, Standards, & Student Testing

UCLA | Graduate School of Education & Information Studies

VISIT US ON THE WEB

cresst.org

dai@cse.ucla.edu