Independent Student Practice Opportunities: Classroom Observations of Student-Teacher Interactions

Keith Smolkowski & Barbara Gunn
Oregon Research Institute

Chris Doabler
University of Oregon
Center on Teaching and Learning

Conference on Learning Disabilities, October 3, 2014

Classroom Observations of Student-Teacher Interactions (COSTI)

- Student-teacher interactions
  - Critical for development of basic skills
  - Malleable
    (Greenwood et al., 2002; Kirschner, Sweller, & Clark, 2006)
- Observable, measurable behaviors that impact on student learning
  - Teacher Demonstrations
  - Student Independent Practice
  - Student Errors
  - Teacher Correction of Errors
- Supported by substantial research and theory

COSTI History

- Supervise and coach student teachers—mid 1990s
  (Barb Gunn & Kathleen Jungjohann)
- Open Court Study—late 1990s
  (Gunn, Kameenui, & Smolkowski, 2000)
  - Observe second grade reading instruction
  - Strong correlations between practice and student outcomes
- SHIP Project—1995 to 2000
  (Gunn, Smolkowski, Biglan, Black & Blair, 2005)
  - Supervise and coach instructional assistants
  - Pilot in Grade 1 classrooms in two small communities
- Read Well K Effectiveness Trial—2005 to 2008
  (Gunn, Smolkowski, & Vadasy, 2011; Smolkowski & Gunn, 2012)

COSTI and Treatment Intensity

- Rate of demonstrations, practice, errors, and corrections directly affects student acquisition and retention of basic skills
  (Archer & Hughes, 2011, Engelmann & Carnine, 1991)
- Treatment intensity allows researchers to examine conditions which may enhance or detract from an intervention’s effect
  (McGinty et al., 2011; Warren, Fey, & Yoder, 2007)

Treatment Intensity

- Total Intervention Duration
  - Time period over which the intervention takes place
  - Example: 12 weeks, entire school year
- Dose Frequency
  - Frequency of intervention sessions provided per period
  - Example: 2 sessions per day, 4 times per week
- Dose or Dose Intensity
  - Number or rate of teaching episodes in a single session
  - Length of intervention session
  - Examples
    - 60 episodes in an hour
    - 5 episodes/min. for 30 minutes

Cumulative Intervention Intensity

- Product of dose intensity × dose frequency × total duration
- Example
  - (4 episodes/min. × 20 minute) × 4 sessions/week × 10 weeks = 3,200 episodes
- Dose Form
  - Typical task or activity within which teaching episode is delivered—the active ingredient
  - Examples
    - Imitative prompts ➔ structured practice
    - Positives reinforcers for appropriate behavior
Why Treatment Intensity?

- More Valuable Research Hypotheses
  - Do interventions differ on cumulative intensity?
  - Does cumulative intensity determine changes?
- Examine Value of Components
  - For a given instructional intervention . . .
    - is there an optimal dose (rate of teaching episodes)?
    - are dose frequency and session duration commutative? (e.g. are four 20-minute sessions equivalent to two 40-minute sessions?)
    - How do duration, frequency, and intensity dimensions interact?

Limited Reports of Intensity

- Only One or Two Dimensions
  - Total sessions, total hours or days, or days per week
    - “12 sessions of 2-3 hours each for three months”
    - “Groups of 5 to 14 parents met weekly for 2.25 hr” and “Groups [of students] met twice a week for 2 hr” (Smolkowski et al., 2005)
  - Easy to report duration
  - Duration does not inform on intensity
- Limits on cumulative dose intensity reports
  - Inadequate conceptualization
    - What is a teaching episode?
  - Measures unavailable

Measuring Intensity with the COSTI

- Explicit Instruction (e.g., Archer & Hughes, 2011)
  - Teacher demonstrations
  - Student independent practice
  - (Student errors)
  - Teacher corrections
- Interactions or Behaviors Frame Student Practice
- Metrics
  - Frequency of behaviors per session—duration dependent
  - Rate of behaviors per minute
  - Contingent proportions of behaviors

Teacher Demonstrations

- Definition
  - Teacher provides new information
  - Teacher shows how to use a new skill
    - Example: “the letter m says /mmm/.”
- Evidence
  (Archer & Hughes, 2011; Meltzoff et al., 2009; Rosenshine, 1997)
  - Modeling the appropriate response allows students to imitate
  - Faster than trial-and-error learning or individual discovery
  - Shown to be more effective than alternatives (e.g., inquiry-based learning)

Independent Practice

- Definition
  - A combination of classroom behaviors: responding to teacher questions, participating in tasks
  - Similar to opportunities to respond.
  - COSTI: student verbally practices or applies new skill without help from the teacher.
- Example
  - Teacher: What sound does m make?
  - Student: /mmm/
  - Teacher: What is this word? [pointing to the word “cat”]
  - Student: “cat”

Independent Practice

- Evidence
  (Brophy & Good, 1986; Ericsson et al., 2007; Melzoff et al., 2009; Sutherland et al., 2003)
  - Higher practice rates improve academic outcomes and behavior
  - Deliberate practice is a defining feature of expert proficiency in music, academics, sports, and games (e.g., chess)
  - High rates of independent practice improves attention
- Biological Evidence—White Matter (Fields, 2008)
  - Axons carry signals faster when myelin-insulated
  - Oligodendrocyte cells manufacture myelin and wrap axons
  - Astrocyte cells . . .
    - Listen in on axon traffic
    - Signal oligodendrocytes to produce more myelin
### Student Errors

- **Definition**
  - Incorrect or no response during independent practice
- **Implications of too many errors**
  - Student may not understand what to do (possibly due to insufficient demonstration)
  - Student has not (completely) learned the skill (possibly due to insufficient opportunities to practice)
- **Implications of very few errors**
  - Overly slow pace of instruction
  - Inadequate teacher attention to errors
  - Not a malleable instructional behavior in isolation

### Teacher Corrections

- **Definition**
  - Teacher directs students’ attention to incorrect response
  - Teacher provides relevant information about the task or skill
  - Example: “The letter *m* says /mmm/.”
  - Negative example: repeating the question
- **Evidence**
  - (Barbetta et al., 1994; Gardner, 1998; Hattie & Timperley, 2007; McCoy & Pany, 1986)
  - Direct corrective feedback enhances word recognition accuracy and, in some cases, reading comprehension
  - Young children require more corrective feedback

---

### The COSTI – Coaching

<table>
<thead>
<tr>
<th>Teacher Name</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lisa McFarland</td>
<td>05/10/11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School Name</th>
<th>Coach/Observer</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roosevelt Elementary</td>
<td>T.O.</td>
<td></td>
</tr>
</tbody>
</table>

**Instructional Focus**: [ ]

**Start Time**: 9:00  **End Time**: 9:55  **Minutes**:      

**COSTI Code**: [ ]  **Total**: [ ]

**Invalid Notes**: [ ]

**Instruction**: [ ]

---

### COSTI – Research

**Activity 1**

<table>
<thead>
<tr>
<th>Start Time</th>
<th>End Time</th>
<th>Student Engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00</td>
<td>9:55</td>
<td>High</td>
</tr>
</tbody>
</table>

**Content Codes**

- 01 High, 02 Medium, 03 Low

**Activity 2**

<table>
<thead>
<tr>
<th>Start Time</th>
<th>End Time</th>
<th>Student Engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00</td>
<td>9:55</td>
<td>High</td>
</tr>
</tbody>
</table>

**Evidence**

- (Barbetta et al., 1994; Gardner, 1998; Hattie & Timperley, 2007; McCoy & Pany, 1986)

---

### The COSTI – Close Up

**Activity 2**

**Start Time**: 9:00  **End Time**: 9:55

**Student Engagement**

- O High  O Medium  O Low

**Content Codes**

- 01 High, 02 Medium, 03 Low

**Evaluating the Effectiveness of Read Well Kindergarten (RWK)**

Barbara Gunn, Keith Smolkowski
Oregon Research Institute
Patricia Vadasy
Washington Research Institute

US Department of Education, Institute of Education Sciences
Grant # R305F060003
Read Well Kindergarten Trial

- Evaluation of Read Well Kindergarten (RWK)
  - Kindergarten reading curriculum
  - First year of a K-3 curriculum
- Research study
  - Design: School-Randomized Trial
  - Intervention schools taught RWK, usually small groups
  - Control schools’ reading programs
    - Also: Zoo Phonics, Scott Foresman, Spalding, Animated Literacy, Celebrate Reading, Reading Milestones, Explode the Code, and Focusing on Language and Academic Instructional Renewal
  - Details: Gunn, Smolkowski, & Vadasy (2010)

Secondary Goal: Technical Adequacy of the COSTI

- Research aims
  - Reliability
    - Observation level: interobserver reliability & stability across observations
    - Classroom level: reliability of classroom means
  - Predictive validity
    - Primary focus: independent practice
- Sample
  - 24 schools in Oregon and New Mexico
  - 79 classrooms; 54 teachers, 25 participated two years
  - 1,519 students in fall; 1,427 in spring

The COSTI Collection

- Observer training
  - Overview & explanation of codes and procedures
  - Practice with video
  - Live practice in kindergarten classrooms
  - Agreement established at 80% or higher
- Procedures
  - Observe entire literacy period – usually small groups
    - Average group size: 7 students
    - Range: 1 to 24 students
  - Three observations per year
  - Reliabilities: two observations collected on 24 occasions

The COSTI Metrics

- Frequencies – raw counts
- Rates
  - Duration-independent
  - Example: 100 independent practice opportunities in 20 minutes → 5 practice opportunities per minute
- Proportions
  - Reduces dependence of practice, errors, corrections
  - Key: Proportion of practice followed by an error
  - Example: errors after 5 of 100 independent practice opportunities → .05 practice opportunities followed by an error

Student Measures

- Measures
  - Letter names and sounds
  - Sight words and decodable words
  - Curriculum based
  - Woodcock Reading Mastery Test (Word ID & Word Attack)
  - Oral reading fluency
  - Phonological processing – CTOPP
  - Vocabulary – PPVT
- Procedures
  - Three-hour training for assessors
  - Standardized administration

The COSTI Means, SDs, & Correlations

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Demonstration Rate</td>
<td>0.8</td>
<td>0.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2. Practice Rate</td>
<td>5.9</td>
<td>4.2</td>
<td>.23</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3. Error Rate</td>
<td>0.4</td>
<td>0.1</td>
<td>.14</td>
<td>.38</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4. Correction Rate</td>
<td>0.2</td>
<td>0.2</td>
<td>.18</td>
<td>.26</td>
<td>.75</td>
<td>-</td>
</tr>
<tr>
<td>5. Prop’n of Practice</td>
<td>.07</td>
<td>.04</td>
<td>.01</td>
<td>-.15</td>
<td>.59</td>
<td>.44</td>
</tr>
<tr>
<td>Followed by Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N = 235 observations. All correlations of .14 or above are statistically significant.
Interobserver Reliability

- Observations nested within occasions
  - Two reliability observations collected on 24 occasions
  - ICC – intraclass correlation coefficient: \( r^2/(r^2 + \sigma^2) \)
  - Occasion-level variance (between occasions): \( r^2 \)
  - Observation-level variance (within occasions): \( \sigma^2 \)

- Interpretation
  - Teacher-level variance as a proportion of total variance
  - Reliability levels (Landis & Koch, 1977)
    - .00 – .20 Slight reliability
    - .21 – .40 Fair
    - .41 – .60 Moderate
    - .61 – .80 Substantial
    - .81 – 1.00 Nearly perfect

Reliability of the Observed Mean

- Observations occasions within classrooms
  - Reliability, \( \lambda \), slightly different than ICC
    - \( \lambda = \frac{\tau^2/(\tau^2 + \sigma^2/n_j)}{\tau^2/(\tau^2 + \sigma^2/n_j) + n_j/(n_j-1) \cdot \rho} \)
    - Classroom-level variance: \( \tau^2 \)
    - Occasion-level variance: \( \sigma^2 \)
    - Number of observation occasions per classroom: \( n_j \)
  - Reliability of mean improves as number of observations increases

- Interpretation
  - Suggested criteria (Smith, McCarthy, & Anderson, 2000)
    - < 0.70 Inadequate
    - 0.70–0.80 Moderately adequate
    - 0.81–1.00 Acceptable
  - Criteria depends on situation (Rosenthal & Rosnow, 2008)

Classroom Stability

- Observations occasions within classrooms
  - 255 observation occasions collected in 79 classrooms
  - ICC – intraclass correlation coefficient: \( r^2/(r^2 + \sigma^2) \)
    - Classroom-level variance (between classrooms): \( \tau^2 \)
    - Observation-level variance (within classrooms): \( \sigma^2 \)

- Interpretation
  - Observers coding different events across time
  - Average correlation between any two observations
    - Analogous to test-retest correlations (Shrout & Fleiss, 1979)
  - Replicates required to estimate behaviors (Shoukri et al., 2004)
    - \( > .50 \) No more than 3 observations required (e.g., per year)
    - 0.20-0.50 Three to six observations required

Predictive Validity

- Students nested within classrooms
  - ~1405 students in 79 classrooms
- Students in predicted by average COSTI measures
  - \( Y_i = \pi_0 + \pi_1 X_{i1} + \epsilon_{i1} \)
    - student score, \( Y_i \)
    - predicted by intercept, \( \pi_0 \)
    - and pretest covariate, \( \pi_1 X_{i1} \)
  - \( \pi_0 = \beta_0, O + \pi_0 \) classroom mean, \( \pi_0 \)
    - predicted by intercept, \( \beta_0, O \)
    - and observation effect, \( \beta_0, O \)

- Effect sizes used the COSTI effects (\( \beta_0, O \) term)
  - Partial correlation, like \( r \) with effects of covariates removed
  - Pseudo-R\(^2\), proportion of classroom-level variance explained by predictors (COSTI measures)

Predictive Validity Results: Partial \( r \)

<table>
<thead>
<tr>
<th>Measures</th>
<th>Letter Names</th>
<th>Letter Sounds</th>
<th>CTOPP</th>
<th>Sight Words</th>
<th>Decodable Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demonstrations</td>
<td>.20</td>
<td>.16</td>
<td>.17</td>
<td>.17</td>
<td>.15</td>
</tr>
<tr>
<td>Practice Opportunities</td>
<td>.32**</td>
<td>.43***</td>
<td>.36**</td>
<td>.55***</td>
<td>.59***</td>
</tr>
<tr>
<td>Student Errors</td>
<td>.12</td>
<td>.24</td>
<td>.26</td>
<td>.38**</td>
<td>.27**</td>
</tr>
<tr>
<td>Teacher Corrections</td>
<td>.06</td>
<td>.16</td>
<td>.21</td>
<td>.31**</td>
<td>.24**</td>
</tr>
<tr>
<td>Prop’n of Practice Followed by Error</td>
<td>-.13</td>
<td>-.19</td>
<td>-.04</td>
<td>-.23**</td>
<td>-.36**</td>
</tr>
</tbody>
</table>

Table reports partial-\( r \) values for the classroom-level effects of the observation measure on each outcome controlling for student-level pretest scores. CTOPP measures phonological processing.

*p < .05, **p < .01, ***p < .001, ****p < .0001
Predictive Validity Results: Pseudo-$R^2$

<table>
<thead>
<tr>
<th>Measures</th>
<th>Letter Names</th>
<th>Letter Sounds</th>
<th>CTOPP</th>
<th>Sight Words</th>
<th>Decodable Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demonstrations</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
</tr>
<tr>
<td>Practice Opportunities</td>
<td>.22***</td>
<td>.21***</td>
<td>.25***</td>
<td>.36***</td>
<td>.39***</td>
</tr>
<tr>
<td>Student Errors</td>
<td>.00</td>
<td>.05*</td>
<td>.13**</td>
<td>.15***</td>
<td>.07*</td>
</tr>
<tr>
<td>Teacher Corrections</td>
<td>-.02</td>
<td>.02</td>
<td>.08</td>
<td>.11**</td>
<td>.06*</td>
</tr>
<tr>
<td>Prop’n of Practice</td>
<td>.03</td>
<td>.04</td>
<td>-.03</td>
<td>.05*</td>
<td>.16***</td>
</tr>
<tr>
<td>Followed by Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table reports Pseudo-$R^2$ values, which represent the reduction in classroom-level variance in student covariate-adjusted scores due to the addition of the observation measure as a predictor.

*p < .05, **p < .01, ***p < .001, ****p < .0001

Limitations & Future Directions

- **Limitations**
  - Just 79 classrooms, mostly small-group instruction
  - The COSTI does not address quality of demonstrations, practice, and corrections

- **Future Directions and Replication**
  - Consider corrective and reinforcing feedback separately
  - Code guided practice (separately from independent practice)
  - Explore role of additional information on quality
  - Examine interplay between COSTI and general observations of classroom environment (e.g., CLASS)

Conclusions

- **The COSTI as a Measure of Dose Intensity**
  - Good construct validity
  - Measures well-defined, explicit teaching episodes
  - May offer a more comprehensive portrait of the behaviors that occur during instructional interactions, possibly in conjunction with other observations

- **Technically Adequacy of the COSTI**
  - Reliable data collection
  - Stable rates of student practice; reliable classroom means
  - Rate of practice predicts important literacy outcomes
  - Except vocabulary (PPVT)
  - See Smolkowski & Gunn (2012; ECRQ) for details

Early Learning in Mathematics Trial

- **Evaluation of ELM**
  - Core kindergarten mathematics curriculum
  - 120 45-min lessons; 15 min-calendar lessons

- **Research study**
  - Design: classroom randomized trial
  - Intervention classrooms taught the ELM curriculum
  - Control classrooms (business as usual) used a host of commercially-available math curricula, including Everyday Math, Houghton Mifflin

- **Details:** Clarke, Smolkowski, Baker, Fien, & Chard (2011)
Secondary Goal:
Technical Adequacy of the COSTI

- Research aims
  - Reliability and stability
  - Predictive validity

- Sample
  - 129 Kindergarten classrooms: 69 ELM; 60 control
  - 112 full-day K-programs; 17 half-day programs
  - Average class size: 20.5 students
  - 130 teachers (one class taught by two teachers)
  - Approximately 2,220 students
    - 53% were boys; 17% received SpEd; 32% considered LEP

Student Measures

- Test of Early Mathematics Ability
  (TEMA-3; Pro Ed, 2003)
  - Counting, calculations, and other mathematical concepts

- Early Numeracy Assessments
  (EN-CBM; Clarke & Shinn, 2004)
  - Oral Counting
  - Number Identification
  - Quantity Discrimination
  - Missing Number
  - Standardized, individually administered assessment procedures at pretest and posttest

The COSTI Collection

- Modified version of the COSTI
  - Combined both guided and independent practice
  - New Interaction Codes:
    - Group response – a verbal response from 2 or more students
    - Individual response – A verbal response from a single student, when elicited or reinforced by the teacher
    - Academic feedback: included both affirmations and corrections

- Observer training
  - Overview and explanation of codes
  - Positive and negative examples
  - Practice with video and in live classrooms
  - Agreement established at 80% or higher
  - Collected three observations per year

Reliability & Stability Results

<table>
<thead>
<tr>
<th>Measures</th>
<th>Interobserver Reliability</th>
<th>Classroom Stability</th>
<th>Reliability of the Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher model rate</td>
<td>.67</td>
<td>.13</td>
<td>.27</td>
</tr>
<tr>
<td>Group practice rate</td>
<td>.92</td>
<td>.15</td>
<td>.31</td>
</tr>
<tr>
<td>Individual practice rate</td>
<td>.95</td>
<td>.34</td>
<td>.56</td>
</tr>
<tr>
<td>Student mistake rate</td>
<td>.84</td>
<td>.19</td>
<td>.37</td>
</tr>
<tr>
<td>Feedback rate</td>
<td>.90</td>
<td>.35</td>
<td>.57</td>
</tr>
</tbody>
</table>

Intraclass correlation (ICC) estimates used to describe interobserver reliability and teacher stability. Interobserver reliability data collected for 74 occasions.

Predictive Validity Results: Partial $r$

<table>
<thead>
<tr>
<th>Measures</th>
<th>TEMA</th>
<th>EN-CBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher model rate</td>
<td>-.03</td>
<td>-.08</td>
</tr>
<tr>
<td>Group practice rate</td>
<td>.03</td>
<td>.02</td>
</tr>
<tr>
<td>Individual practice rate</td>
<td>.25*</td>
<td>.21*</td>
</tr>
<tr>
<td>Proportion of practice with errors</td>
<td>-.17</td>
<td>.00</td>
</tr>
<tr>
<td>Proportion of practice with feedback</td>
<td>.09</td>
<td>.13</td>
</tr>
</tbody>
</table>

- Multilevel Model nested ~2200 students within 129 classrooms
  - Students in classrooms predicted by average COSTI measures
  - Partial $r$ values reported at classroom level.
  - EN-CBM: composite scores
  - $^* p < .05$

Predictive Validity Results: Pseudo-$R^2$

<table>
<thead>
<tr>
<th>Measures</th>
<th>TEMA</th>
<th>EN-CBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher model rate</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Group practice rate</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Individual practice rate</td>
<td>.04*</td>
<td>.03*</td>
</tr>
<tr>
<td>Proportion of practice with errors</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>Proportion of practice with feedback</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

- Multilevel Model
  - Pseudo-$R^2$ values: proportion of classroom-level variance in student outcome explained
  - EN-CBM: composite score
  - $^* p < .05$
Discussion

- Rate of individual practice opportunities (i.e., verbal responses) are important for improving student performances on proximal (EN-CBM) and distal (TEMA) measures of mathematics.
- Aligns with recent meta-analysis on mathematics interventions: student verbalizations of mathematical thinking has a relatively large effect ($Hedges\ g = 1.04, 95\% \text{ CI} [0.42, 1.66]$) on math outcomes for students with MD (Gersten et al., 2009).

Overall Implications of COSTI for Research and Practice

- Provides quantitative data on the rate (intensity) and sequencing of student-teacher behaviors during instruction.
- Allows researchers to examine how treatment intensity may enhance or detract from an intervention’s effect.
- Complements observations of the general classroom environment to provide a more comprehensive picture of the dynamics of instruction.
- Provide teachers with specific instructional feedback, if interactions predict student outcomes.