

Instructional Practice and Student Achievement: Validating a Contemporary Classroom Walkthrough Tool

Dale L. Cusumano

National Implementation Research Network

Matthew A. Lenard¹

Wake County Public School System

Angela Preston

National Implementation Research Network

Caryn S. Ward

National Implementation Research Network

Structured Abstract

Background & Context

How do the impacts of an intervention vary by levels of implementation? Stakeholders often ask this question of researchers and researchers often struggle to provide a response due to the lack of fidelity measures available to them (Dane & Schneider, 1998; Durlak & DuPre, 2008, Cross et al., 2011, Dobson & Cook, 1980, Naleppa & Cagle, 2010). In education research, results are often presented as intent-to-treat (ITT) estimates, suggesting that specific findings may be generalizable to an entire setting. But implementation varies widely by unit, whether they are students, teachers or schools, and validated, on-the-ground measures of instructional implementation for contemporary interventions are few and far between. This paper explores the relationship between one such tool, the Observation Tool for Instructional Supports and Systems (OTISS), and student achievement, and it suggests that for reading interventions, in particular, it may serve as a powerful measure of implementation fidelity.

Researchers from the National Implementation Science Network (NIRN) developed OTISS in response to the lack of validated fidelity measures available for evidence-based instructional practices (Fixsen, Ward, Ryan Jackson & Chapparo, 2015). The purpose of OTISS is to assess the quality of systems and supports available to help teachers use best practices for instruction. OTISS is a brief, 10-minute walkthrough observation of instruction. It is content free and applicable to any grade level and any part of an instructional activity. To maximize the benefit to students, OTISS observations are designed to be frequent, relevant, and actionable. John Hattie's (2009) meta-analytic work informed the development of OTISS. Consequently, inclusion criteria for OTISS items included teacher behaviors that would happen (1) at any time during instruction, (2) consistently during instruction, and (3) at any grade level or content area.

¹ Corresponding author: mlenard@wcpss.net

Purpose & Research Questions

The purpose of this work is to establish the predictive power and reliability of OTISS. To date, NIRN has established the face and construct validities, sensitivities, and bias reviews of OTISS. In spring 2018, we will present the results of the following research questions to evaluate the tool's validity and reliability:

1. Which effective instructional practices predict student achievement in elementary reading and math?
2. Are OTISS scores consistent (reliable) over time?
3. Are there differences in teachers' use of effective instructional practices across different levels of schooling and thresholds of poverty?
4. Do teachers' use of effective instructional practices vary over time and across content areas

In this proposal, we present our findings for the first research question, which explores the predictive power of OTISS ratings (see Table 1 for the seven items) on student achievement outcomes, including Dynamic Measures of Basic Early Literacy (DIBELS) measures and North Carolina's End-of-Grade Tests.

Setting

This work constitutes a research-practice partnership between the Wake County Public School System (Wake County) and NIRN, based at the University of North Carolina at Chapel Hill. Wake County is the largest school district in North Carolina and the 15th largest in the nation. The two entities collaborated when Wake County required an implementation fidelity tool for use in its randomized controlled trial of multi-tier system of supports (MTSS).² NIRN's OTISS tool was deployed in the district's 44 MTSS treatment schools and administered throughout the 2016-17 school year.

Participants

Over the course of 2016-17, school-based observers established inter-rater reliability and conducted roughly 2,500 unique observations across the 44 MTSS schools. Observers were instructed to focus on the core content areas of math and reading (though they did observe teachers of non-core subjects) and individual teachers in our file were observed anywhere from one to five times, with the average teacher receiving 1.8 observations. Ratings on OTISS range from 0 to 2 (with a not-observed option). Table 2 summarizes mean OTISS ratings as well as descriptive statistics for our achievement outcomes and controls for individual student characteristics. Our analytic sample consists of roughly 15,000 unique students linked to their teacher of record.

² This work is generously supported by the Laura and John Arnold Foundation through its *Low-Cost Randomized Controlled Trials to Drive Effective Social Spending* program.

Research Design

To estimate the predictive power of aggregate and individual OTISS ratings on measures of student achievement, we fit a three-level model of the following form:

$$OUTCOME_{ijk} = \beta_{0ijk} + \beta_1 OTISS_{ijk} + \beta_3 PRIOR_{ijk} + \beta_4 X + \varepsilon_{ijk}$$

Where *OUTCOME* represents standardized achievement for student *i* in classroom *j* in school *k* as a function of OTISS rating, prior achievement on that same outcome measure, and a vector of student level covariates. We also fit two-level models, omitting the classroom-level contribution.

Data Collection and Analysis

DIBELS and End-of-Grade Test scores data are drawn from Wake County administrative records. Student-level data are maintained through Wake County's student information system. OTISS ratings are maintained by the State Implementation and Scaling-up of Evidence-based Practices (SISEP) Center, a national technical assistance center funded by the U.S. Department of Education. As noted above, we fit two- and three-level models in order to estimate the predictive power of OTISS on student achievement.

Results

Tables 3-6 summarize findings for our first research question, above. The key finding is that on the EOG reading test, OTISS ratings predict higher EOG outcomes, which are statistically significant in the case of two of the seven ratings and in the total score ($0.043 \leq SD \leq 0.060$; $p < .05$). The results are less conclusive for student performance on the EOG math test and on the two DIBELS measures included here.

Conclusions

This proposal includes findings from the first of four research questions we would present this spring. Our goal is to highlight the validity and reliability of OTISS, a contemporary instructional walkthrough instrument, which can inform the implementation fidelity of interventions that lack such measures. Specifically, we find that variation in select classroom practices—demonstrating tasks and providing multiple opportunities for practice—significantly predicts reading outcomes for elementary school students on North Carolina's state summative assessment. The results suggest that OTISS may constitute an additional resource for teacher and practitioners exploring program fidelity for new or established literacy interventions.

References

Crosse, S., Williams, B., Hagen, C. A., Harmon, M., Ristow, L., DiGaetano, R., . . . Derzon, J. H. (2011). Prevalence and implementation fidelity of research-based prevention programs in public schools: Final report. Washington, DC: U.S. Department of Education.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18(1), 23-45.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327-350. doi: 10.1007/s10464-008-9165-0

Dobson, L., & Cook, T. (1980). Avoiding Type III error in program evaluation: results from a field experiment. *Evaluation and Program Planning*, 3, 269 - 276.

Fixsen, D., Ward, C.S., Ryan Jackson, K., & Chapparo, E. (2015). *The Observation Tool for Instructional Supports and Systems*. University of North Carolina at Chapel Hill

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Naleppa, M. J., & Cagle, J. G. (2010). Treatment fidelity in social work intervention research: A review of published studies. *Research on Social Work Practice*, 20(6), 674-681. doi: 10.1177/1049731509352088

Table 1. *Example of Observation Tool for Instructional Supports and Systems (OTISS), Teacher Copy*

Note: The purpose of the OTISS observation and data collection protocol is **to assess the quality of systems and supports available to assist teachers** use best practices for instruction. *In addition, today's observation is for training purposes and to inform the development of OTISS.*

Category Rating ¹	Category of Adult Behavior The Instructor...	Operationalized Definitions of Instruction Behavior
2 1 0 NA	1. Provides clear instruction	“Provides clear instruction” is operationalized as the teacher stating a) what students should learn and b) why the learning concept is important.
2 1 0 NA	2. Demonstrates instructional tasks	“Demonstrates instructional tasks” is operationalized as the teacher modeling skills a) explicitly (e.g. step by step), b) consistently (e.g. repeatedly), and c) with examples and/or non-examples.
2 1 0 NA	3. Engages students in meaningful interactions with content	“Engages students in meaningful interactions with content” is operationalized as the teacher using a variety of materials and strategies to a) generate new knowledge, b) extend critical thinking, and c) promote reflection by all students on their own learning, effort, and understanding.
2 1 0 NA	4. Provides prompt and accurate feedback	“Provides prompt and accurate feedback” is operationalized as the teacher communicating verbal and nonverbal positive and encouraging responses appropriate to a student's a) effort, b) behavior, or c) engagement with academic content.
2 1 0 NA	5. Adjusts to students' responses to instruction	“Adjusts to students' responses to instruction” is operationalized as the teacher demonstrating flexibility in a) pacing the lesson, b) redirecting off-task behavior, or c) incorporating additional practice based on students' responses.
2 1 0 NA	6. Provides multiple opportunities for students to practice	“Provides multiple methods and opportunities for students to practice” is operationalized as the teacher a) checking for understanding and b) using a variety of methods and activities to engage students through individual and group responses.
2 1 0 NA	7. Adjusts to students' engagement with instruction	“Adjusts to students' engagement with instruction” is operationalized as the teacher purposefully a) adjusting the physical environment, b) access to materials, or c) the manner of instruction in response to an

¹ Note: 2=Fully Observed, 1=Partially Observed, 0=Not Observed, NA = No Opportunity to Observe

Table 2. Summary Statistics

<i>OTISS Items</i>	N	Mean	SD	Min	Max
Overall score	13,969	1.322	0.427	0	2
Provides clear instruction	14,841	0.885	0.625	0	2
Demonstrates instructional tasks	14,784	1.158	0.668	0	2
Engages students in meaningful interactions with content	15,371	1.293	0.589	0	2
Provides prompt and accurate feedback	15,472	1.585	0.568	0	2
Adjusts to students' responses to instruction	15,165	1.510	0.585	0	2
Provides multiple opportunities for students to practice	15,279	1.368	0.581	0	2
Adjusts to students' engagement with instruction	15,377	1.476	0.611	0	2
<i>Achievement</i>	N	Mean	SD	Min	Max
DIBELS Composite Prior	13,300	239.6	155.5	0	809
DIBELS Composite	13,259	322.6	147.7	0	818
DIBELS ORF Prior	8,484	104.7	41.1	0	265
DIBELS ORF	10,784	117.7	45.2	0	292
EOG Math Prior	4,837	454.3	10.0	421	473
EOG Math	7,913	454.1	10.3	424	475
EOG Reading Prior	4,799	446.9	10.7	406	468
EOG Reading	7,841	449.0	10.9	409	472
<i>Student Characteristics</i>	N	Mean	SD	Min	Max
Asian	15,651	0.143	0.350	0	1
Hispanic	15,651	0.146	0.353	0	1
Black	15,651	0.178	0.383	0	1
Multiracial	15,651	0.033	0.180	0	1
White	15,651	0.496	0.500	0	1
Gifted/talented	15,651	0.108	0.310	0	1
Male	15,651	0.506	0.500	0	1
Students with Disabilities	15,651	0.095	0.293	0	1
Limited English Proficient	15,651	0.086	0.280	0	1

Table 3. *Contribution of OTISS Items to DIBELS Composite*

	2-Level Model	3-Level Model	<i>N</i>	<i>R</i> ²
Overall score	-0.018 (0.014)	-0.022 (0.027)	11181	0.636
Provides clear instruction	0.003 (0.009)	0.010 (0.018)	11903	0.634
Demonstrates instructional tasks	-0.025*** (0.008)	-0.024 (0.017)	11903	0.641
Engages students in meaningful interactions with content	-0.007 (0.009)	-0.008 (0.019)	12290	0.637
Provides prompt and accurate feedback	-0.004 (0.009)	0.010 (0.019)	12404	0.636
Adjusts to students' responses to instruction	-0.013 (0.009)	-0.012 (0.019)	12141	0.636
Provides multiple opportunities for students to practice	0.006 (0.009)	0.001 (0.019)	12220	0.636
Adjusts to students' engagement with instruction	-0.021** (0.009)	-0.022 (0.018)	12294	0.637

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 4. *Contribution of OTISS Items to DIBELS Oral Reading Fluency*

	2-Level Model	3-Level Model	<i>N</i>	<i>R</i> ²
Overall score	0.004 (0.012)	0.000 (0.019)	6834	0.807
Provides clear instruction	0.006 (0.008)	0.007 (0.013)	7306	0.807
Demonstrates instructional tasks	-0.015** (0.008)	-0.017 (0.012)	7293	0.808
Engages students in meaningful interactions with content	0.019** (0.008)	0.021* (0.013)	7598	0.808
Provides prompt and accurate feedback	-0.007 (0.008)	-0.007 (0.013)	7662	0.808
Adjusts to students' responses to instruction	-0.002 (0.008)	-0.002 (0.013)	7444	0.809
Provides multiple opportunities for students to practice	0.021** (0.008)	0.014 (0.013)	7546	0.809
Adjusts to students' engagement with instruction	-0.009 (0.008)	-0.014 (0.012)	7596	0.808

Standard errors in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

Table 5. *Contribution of OTISS Items to End-of-Grade Math Performance*

	2-Level Model	3-Level Model	<i>N</i>	<i>R</i> ²
Overall score	0.007 (0.018)	-0.005 (0.036)	4235	0.756
Provides clear instruction	-0.012 (0.012)	-0.003 (0.023)	4476	0.757
Demonstrates instructional tasks	0.001 (0.011)	-0.004 (0.022)	4465	0.758
Engages students in meaningful interactions with content	0.028** (0.012)	0.014 (0.025)	4731	0.760
Provides prompt and accurate feedback	-0.001 (0.013)	-0.004 (0.024)	4732	0.758
Adjusts to students' responses to instruction	-0.019 (0.012)	-0.019 (0.024)	4595	0.759
Provides multiple opportunities for students to practice	0.017 (0.013)	0.009 (0.025)	4683	0.4683
Adjusts to students' engagement with instruction	-0.012 (0.011)	-0.017 (0.022)	4731	0.759

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 6. *Contribution of OTISS Items to End-of-Grade Reading Performance*

	2-Level Model	3-Level Model	<i>N</i>	<i>R</i> ²
Overall score	0.067*** (0.019)	0.060** (0.026)	4204	0.719
Provides clear instruction	0.014 (0.013)	0.016 (0.017)	4444	0.719
Demonstrates instructional tasks	0.047*** (0.012)	0.043*** (0.016)	4434	0.720
Engages students in meaningful interactions with content	0.033** (0.013)	0.027 (0.018)	4698	0.722
Provides prompt and accurate feedback	0.030** (0.013)	0.027 (0.018)	4699	0.721
Adjusts to students' responses to instruction	0.028** (0.013)	0.023 (0.017)	4562	0.722
Provides multiple opportunities for students to practice	0.047*** (0.013)	0.043** (0.018)	4650	0.724
Adjusts to students' engagement with instruction	0.026** (0.012)	0.027* (0.016)	4698	0.722

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$