(2016) In T. Fukawa-Connelly, N. Infante, K. Keene, and M. Zandieh (Eds.), Proceedings of the 19th Annual Conference on Research in Undergraduate Mathematics Education, Pittsburgh, Pennsylvania.

# How "Good" is "Good Enough"? Exploring Fidelity of Implementation for a Web-based Activity and Testing System in Developmental Algebra Instruction

Shandy Hauk, WestEd
Katie Salguero, WestEd
Joyce Kaser, WestEd

*A web-based activity and testing system (WATS) has features such as adaptive problem sets, videos, and data-driven tools for instructors to monitor and scaffold student learning. Central to WATS adoption and use are questions about the implementation process: What constitutes "good" implementation and how far from "good" is "good enough"? Here we report on and illustrate our work to provide structure for such examination. The context is a study about implementation that is part of a state-wide randomized controlled trial examining student learning in community college algebra when a particular WATS suite of tools is used. Discussion questions for conference participants dug into the distinctions among intended, enacted, and achieved curriculum and the processes surrounding these as well as the challenges and opportunities in researching fidelity of implementation in the community college context, particularly the role of instructional practice as a contextual component of the research.*

*Key words:* Fidelity of Implementation, College Algebra, Research Tools

## Background

"How good is good enough?" has plagued humankind since the early cave dwellers wondered if killing three bison would get the family through the cold winter months. Even today, with our technological advances, we still ask questions such as "Do I have enough money for retirement?" "Have I practiced enough hours?" or "Is what I'm doing good enough?"

This ubiquitous question plagues social science researchers who are assessing the whats, whys, and hows of an intervention. Did the instructors have enough support to adequately implement the new curriculum? Were the materials adequate to provide enough practice hours for students? Was the instruction sufficient to prepare students to pass the final exam? Oh, if there were only an answer!

### Study Context

We chose to attempt to answer this question of "good enough" in the implementation of a large project investigating relationships among student achievement and varying conditions of implementation for a web-based activity and testing system (WATS) used in community college algebra. We selected an implementation research approach that we had used previously and found to be helpful. In the new study we hope to replicate and to refine our earlier experience. Implementing the WATS is part of a statewide, randomized controlled trial examining student learning in community college algebra. WATS tools include adaptive problem sets, instructional videos, and data-driven tools for instructors to use to monitor and scaffold student learning. The WATS is accessed on the internet and is designed primarily for use as replacement for some in-class individual seatwork.

### Research Questions

In what ways does a program-in-operation have to match the program-as-intended to be successful? Well, we have to identify what "success" means and also to identify alignment

between intended and enacted implementation. Thus, two major research questions drive our attempt to answer the "good enough" question:

(1) What is the nature of alignment between how the program is implemented and how the developer/publisher envisioned it (i.e., what is the fidelity of implementation)?
(2) What are the relationships among varying conditions of implementation (differing degrees of fidelity) and the extent to which students are achieving the desired results?

## Conceptual Framework

The theoretical basis for our approach lies in program theory, "the construction of a plausible and sensible model of how a program is supposed to work" (Bickman, 1987, p. 5). Having such a model in place allows researchers to conjecture and test causal connections between inputs and outputs, rather than relying on intuition or untested assumptions. As in many curricula projects, developers of the program in our study did pay attention to learning theory in determining the content in the web-based system, but the same was not true for determining implementation processes and structures. The pragmatic details of large-scale classroom use were under-specified. Developers articulated their assumptions about what students learned as they completed activities, but the roles of specific components, including the instructor role in the mediation of learning, were not clearly defined.

As Munter and colleagues (2014) have pointed out, there is no agreement on how to assess fidelity of implementation. However, there is a growing consensus on a component-based approach to measuring its structure and processes (Century & Cassata, 2014). *Fidelity of implementation* is the degree to which an intervention or program is delivered as intended (Dusenbury, Brannigan, Falco, & Hansen, 2003). Do implementers understand the trade-offs in the daily decisions they must make "in the wild" and the short and long-term consequences on student learning as a result of compromises in fidelity? Century and Cassata's (2014) summary of the research offers five core components to consider in fidelity of implementation: Diagnostic, Procedural, Educative, Pedagogical, and Student Engagement (see Table 1).

Table 1. *Components and Focus in a Fidelity of Implementation Study*

| *Components* | *Focus* |
|---|---|
| Diagnostic | These factors say what the "it" is that is being implemented (e.g., what makes this particular WATS distinct from other activities). |
| Structural-Procedural | These components tell the user (in this case, the instructor) what to do (e.g., assign intervention *x* times/week, *y* minutes/use). These are aspects of the *expected* curriculum. |
| Structural-Educative | These state the developers' expectations for what the user needs to know relative to the intervention (e.g., types of technological, content, pedagogical knowledge are needed by an instructor). |
| Interaction-Pedagogical | These capture the actions, behaviors, and interactions users are expected to engage in when using the intervention (e.g., intervention is at least *x* % of assignments, counts for at least *y* % of student grade). These are aspects of the *intended* curriculum. |
| Interaction-Engagement | These components delineate the actions, behaviors, and interactions that students are expected to engage in for successful implementation. These are aspects of the *achieved* curriculum. |

## Method

The components in Table 1 are operationalized through a rubric, the guide for collecting and reporting data in our implementation study. A rubric is a "document that articulates the expectations for an assignment by listing the criteria, or what counts, and describing the levels of quality from excellent to poor" (Andrade, 2014). Each component has several factors that define the component. The project's research team has developed a rubric for fidelity of implementation, identifying measurable attributes for each component (for example, see Table 2 for some detail on the "educative" component).

Table 2. *Example of rubric descriptors for levels of fidelity, Structural-Educative component.*

| ***Educative:*** These components state the developers' expectations for what the user needs to know relative to the intervention. | | | |
|---|---|---|---|
| | **High Level of Fidelity** | **Moderate Fidelity** | **Low Level of Fidelity** |
| **Users' proficiency in math content** | Instructor is proficient to highly proficient in the subject matter. | Instructor has some gaps in proficiency in the subject matter. | Instructor does not have basic knowledge and/or skills in the subject area. |
| **Users' proficiency in TPCK** | Instructor regularly integrates content, pedagogical, and technological knowledge in classroom instruction. Communicates with students through WATS. | Instructor struggles to integrate CK, PK, and TK in instruction. Occasionally sends digital messages to students using WATS tools. | Instructor CK, PK, and/or TK sparse or applied in a haphazard manner in classroom instruction. Rarely uses WATS tools to communicate with students. |
| **Users' knowledge of requirements of the intervention** | Instructor understands philosophy of WATS resources (practice items, "mastery mechanics," analytics, and coaching tools), | Instructor understanding of the philosophy of WATS tool has some gaps. NOTE: Disagreeing is okay, this is about instructor knowledge of it. | Instructor does not understand philosophy of WATS resources. NOTE: Disagreeing is okay, this is about instructor knowledge of it. |
| **Users' knowledge of requirements of the intervention** | Instructor understands the purpose, procedures, and/or the desired outcomes of the project (i.e., "mastery") | Instructor understanding of project has some gaps (e.g., may know purpose, but not all procedures, or desired outcomes). | Instructor does not understand the purpose, procedures, and/or desired outcomes. Problems are typical. |

## Results

Our focus here is two-fold. We first offer the preliminary results of rubric refinement from data collected through observation, interview, and teacher self-report in weekly surveys (also known as "teaching logs"). These results were shared on the poster (and handouts) at the conference. Then we summarize the highlights of the conversations about researching fidelity of implementation that emerged at the conference.

**Defining and Refining Measures for the Fidelity of Implementation Rubric**

The ultimate purpose of a fidelity of implementation rubric is to articulate how to determine what works, for whom, under what conditions. In addition to allowing identification of alignment between developer expectations and classroom enactment, it provides the opportunity to discover where productive adaptations may be made by instructors, adaptations that boost student achievement beyond that associated with an implementation faithful to the developers' view.

The example on the poster was for the procedural component from our WATS intervention (see Table 3, next page). The Structural-Procedural components tell the user what needs to be done (e.g., makes assignments for students to complete using the WATS tool). The table has four rows of expectations. Columns define high, moderate, and low fidelity followed by data sources and notes on the measures used.

We employ a mixed-method, feedback design to capture and communicate about fidelity of implementation. A feedback design for refining an intervention can be driven by qualitative research and supported by quantitative snapshots of student performance, teacher understandings, and systemic growth. Or vice versa. Our rubric (Table 3) lists primary, secondary and tertiary sources of data for gathering information about the four items on the procedural component of the fidelity rubric. These sources are WATS Application programming interface (API) – this provides data from the digital audit trail of WATS usage, occasional classroom observations for some instructors with an associated instructor interview, instructor self-report (through logs and surveys), and student survey. These measures were selected based on available sources and constraints on project time and funding.

We always dance between what we want to know about an intervention and what we are able to measure. Instructor self-report logs are highly useful as they can document what is happening with implementation. For example, logs can tell us how many times an instructor mentioned or used the intervention. And that accretion across weeks gives the area under the curve of what's going on across time, contributing to the big picture, of implementation.

In using the rubric, we assign a number to each level of fidelity. This can be as simple as a 3 for a high level of fidelity, 2 for a moderate level of fidelity, or a 1 for a low level; or the items can be weighted. Note on Table 3 under "amount of instruction – mindset lessons" we will know instructors' use of mindset lessons through logs and an interview question and can then assign a high, moderate, or low level of fidelity to the item (see Table 2, Notes on Metrics).

The score for the intervention will be the total number of points assigned in completing the rubric as a ratio of the total possible, across all instructors. It will also be possible to create a fidelity of implementation score on each row for each instructor – these data will be used in statistical modeling of the impact of the intervention as part of a "specific fidelity index" (Hulleman & Cordray, 2009). We first total points for the item, then the component, and finally all components for a single score as an index of implementation.

We anticipate having data that allow us to answer several questions related to "good enough." For example, for Research Question 1:

- To what extent did the instructors assign WATS activities?
- To what extent did the instructors encourage students to complete the WATS activities?
- To what extent were the mindset lessons implemented?
- How frequently was WATS assigned?

And, for Research Question 2: What is the relationship between level of mastery students achieved and number of WATS activities students completed or number of mindset activities students experienced?

**Table 3.** *Structural-Procedural.* These components tell the user what to do regarding instruction.

| | *High Level of Fidelity* | *Moderate Level of Fidelity* | *Low Level of Fidelity* | *Primary Data Source* | *Second Data Source* | *Third Data Source* | *Notes on Metric(s)* |
|---|---|---|---|---|---|---|---|
| *Assigns WATS* | Instructor assigns all target WATS activities. | Instructor assigns 80% to 99% of target activities. | Instructor assigns fewer than 80% of target activities. | WATS API | Logs | Instruc. Interview | Source of boundary conditions: Developer, who said: "best guess at what a minimally effective dose is. Really, it has to be all." |
| *Amount of instruction - Promotes Completion* | Regularly encourages students to complete the mission. | Sometimes encourages students to complete the mission. | Rarely or never encourages students to complete the mission. | Logs | Instruc. interview | Obser-vation | *High & Moderate*: >2/3 of instructors say:<br>Regularly: >75% of weeks (High)<br>Sometimes: >35% of wks (Mod.)<br>*Low*: <2/3 instructors OR<br>Rarely: <35% of weeks |
| *Amount of instruction - Mindset Lessons* | Conducts 3 mindset lessons. | Conducts 2 mindset lessons. | Conducts at most 1 mindset lesson. | Logs | Instruc. Interview | N/A | Instructor Interview Q4: What plans do you have for mindset activities with students?<br>Logs: What mindset lessons did your class do this week? |
| *Frequency of instruction* | WATS assigned at least 12 of 15 weeks in the term. | WATS assigned at least 7, up to 11 weeks in the term. | WATS assigned in 6 or fewer weeks in the term. | Logs | WATS API | Student survey | *High*: At least 2/3 of Instructors report WATS use in at least 75% of logs.<br>*Moderate*: At least 2/3 of Instructors report use in 50 to 75% of logs.<br>*Low*: not high or moderate. |

**At the Conference: Poster Conversations**

The factors included in the poster were meant as a starting point for conversation. The poster shared the theory behind the protocol and was a touchstone for gathering ideas from RUME attendees on dissemination that might be productive as we move forward into the full study (2015 is a "practice" year for the study). Here we summarize the highlights of the conversations at the poster.

*Participant comment:* I never thought about this before, that somebody might pick up an activity that I designed and use it in a counter-productive way. Why would they even try it if they didn't think like I did about how to use it?

*Response:* There are myriad of reasons why someone would change the way they conduct an activity that another had designed. They can be from differences in content, pedagogical, or pedagogical content knowledge, or due to limitations of time or resources or even relate to someone wanting to "brand" the activity as their own. Just remember: what **you** call counter-productive may be seen as a helpful tweaking to someone else. The key lies in the impact of the change on the desired result (e.g., student learning gains).

*Participant question:* How do you do curriculum development when multiple people develop something for use by multiple people, including some who are not in the room?

*Answer:* Very skillfully! Your start with using data about your intended audience. As you design, you determine what you think will be your fidelity factors – the ones that drive your anticipated results. Next you implement your activity and then gather data to determine results and confirm the role of your fidelity factors. Further testing can show what happens when you vary a fidelity factor like contact time or dosage. These are all excellent opportunities to document "good enough." Remember you are dealing with human beings. Keep in mind the idea of "close approximation."

*Participant question:* How do you decide what the "it" is that is being implemented?

*Answer.* Excellent question. The "it" is the intervention, the project, the curriculum. Determining the "it" is answered in part by asking about a series of diagnostic factors that are part of our model. We start by interviewing developers, asking these diagnostic questions. One of the questions is how the intervention, project, or curriculum differs from others that are similar. Then we layer this information with observation of the training that developers give to faculty and the kinds of questions faculty ask about using the intervention during the training. You might think that determining the "it" is easy – sometimes yes, sometimes no. Unless you zero in on what the "it" is, you will never get to the level of specificity required to evaluate fidelity factors.

*Participant question:* I like to think of a three-way overlapping Venn diagram for an intervention: the intended curriculum, the implemented curriculum, and the achieved curriculum. Can your framework relate to this concept?

*Answer:* Our model is a fourth party that attempts to take in perspectives of all these aspects curricula. It can connect them as an important way of monitoring for efficacy (also, see the notes in the second column of Table 1, above).

*Participant comment:* I am surprised that a component is that instructors might need certain types of knowledge before they are ready to use a particular type of intervention.

*Response:* Usually something about an intervention is new. Maybe new content. Maybe new pedagogy. The instructor may not have learned whatever is required to carry out the

intervention. One of the major reasons interventions fail is that participants are asked to do new things (such is the nature of interventions) for which they are given little or no training.

*Participant question:* I like the idea of descriptions of performance at the high, medium, and low levels. Can you develop materials that incorporate such descriptions on the front end?
*Answer:* Sure you can. Such descriptions can be used at each level from the beginning of an activity or program through the implementation and finally for the evaluation at the end.

*Participant question:* Does your framework help increase equity in any respect?
*Answer:* The specificity of what fidelity of implementation requires we include in the rubrics is an opportunity for us to address potential challenges to equity and inclusion in the implementation of an intervention. How to make college math accessible to all students is a theme of the work in in the WATS system we are studying. Investigating fidelity of implementation allows us to identify how curriculum and its implementation play a part in that accessibility process.

*Participant question:* Where is it explicit to a user what the developer's intentions are?
*Answer:* Sometimes the developers will tell you outright in the introductory material. Other times the intent is buried in the content, and you have to unearth it. Sometimes developers are very cognizant of their intentions; other times, oblivious. Regardless of level of transparency, intentions are always there.

*Participant question:* As a classroom instructor, where in the rubrics is my relationship with the WATS online resource? My perspective about its use in teaching and learning?
*Answer:* Yes, that's something we are wrestling with as we develop the details of the Educative rubric (Table 2). Right now, the rubric looks at the degree of knowledge instructors have about the intended relationship (e.g., about the philosophy behind the WATS tool), not at the alignment of the instructor's view with that perspective. We agree success of implementation may depend on how someone sees the resource, but is it necessarily an aspect of being faithful to the intentions of the tool? For an instructor, the resource can be a partner, or a distinctly separate support for teaching, or even an obstacle. The Concerns Based Adoption Model provides some ideas that we are pursing (Hall & Hord, 2014).

*Participant comment:* It's a new idea to me that implementation could be a major field of study.
*Response:* It has grown exponentially over the past 20 years, and we have learned much about the implementation process. You have probably heard the cliché, "We tried that once and it didn't work." What actually happens most often is a failure in implementation. Even the best ideas will collapse with insufficient or faulty implementation.

## Implications for Practice

By definition, high fidelity implementation of an instructional tool is use that results in greater learning gains than non-use. Instructors and students are better equipped to implement with high fidelity when they have answers to questions like: What are the characteristics of good implementation? Among preferred actions in implementation, which are the highest priority? What are the trade-offs and consequences of making particular decisions about use of the tool?

Answers to these questions provide data for determining what is "good enough" and help users make the best decisions for program efficacy. As the field moves forward, we seek

effective ways to communicate implications to college instructors, department chairs, as well as stakeholders in the larger public arena.

**Acknowledgement**

This project is supported by a grant from the U.S. Department of Education, Institute of Education Sciences (IESR305A140340). Any opinions, findings, conclusions or recommendations are those of the authors and do not necessarily reflect the views of the Federal Government.

**References**

Century, J., & Cassata, A. (2014). Conceptual foundations for measuring the implementation of educational innovations. *Treatment Integrity: A Foundation for Evidence-Based Practice in Applied Psychology*, 81-108.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health education research*, *18*(2), 237-256.

Hall, G. E., & Hord, S. M. (2014). *Implementing change: Patterns, principles, and potholes (4th ed.)*. Upper Saddle River, NJ: Pearson.

Hulleman, C. S., & Cordray, D. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Intervention Effectiveness*, 2(1), 88-110.

Munter, C., Garrison Wilhelm, A., Cobb, P., & Cordray, D. S. (2014). Assessing fidelity of implementation of an unprescribed, diagnostic mathematics intervention. *Journal of Research on Educational Effectiveness*. *7*(1), 83-113.