



# A case study of pragmatic measure development of the Treatment Integrity for Elementary Settings self-report measure for teachers

Implementation Research and Practice  
Volume 5: Jan-Dec 2024 1–6  
© The Author(s) 2024  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/26334895231220262  
journals.sagepub.com/home/irp



Emma R. Dear<sup>1</sup> , Bryce D. McLeod<sup>1</sup> , Nicole M. Peterson<sup>2</sup>,  
Kevin S. Sutherland<sup>3</sup>, Michael D. Broda<sup>3</sup>, Alex R. Dopp<sup>4</sup> ,  
and Aaron R. Lyon<sup>5</sup> 

## Abstract

### Introduction

Due to usability, feasibility, and acceptability concerns, observational treatment fidelity measures are often challenging to deploy in schools. Teacher self-report fidelity measures with specific design features might address some of these barriers. This case study outlines a community-engaged, iterative process to adapt the observational Treatment Integrity for Elementary Settings (TIES-O) to a teacher self-report version designed to assess the use of practices to support children's social-emotional competencies in elementary classrooms.

### Method

Cognitive walkthrough interviews were conducted with teachers to improve the usability of the teacher self-report measure, called the Treatment Integrity for Elementary Schools–Teacher Report (TIES-T). Qualitative content analysis was used to extract themes from the interviews and inform changes to the measure.

### Results

Increasing clarity and interactive elements in the measure training were the dominant themes, but suggestions for the measure format and jargon were also suggested.

### Conclusion

The suggested changes resulted in a brief measure, training, and feedback system designed to support the teacher's use of practices to support children's social-emotional competencies in elementary classrooms. Future research with the TIES-T will examine the score reliability and validity of the measure.

**Plain Language Summary:** Collecting observational data in schools is challenging, so developing teacher self-report measures and involving teachers in the design process is important to help make them easier to use. This paper reports on the development of a teacher self-report measure designed to collect information about the instructional practices teachers deliver to promote positive student behavior.

<sup>1</sup>Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA

<sup>2</sup>Neag School of Education, University of Connecticut, Storrs, CT, USA

<sup>3</sup>School of Education, Virginia Commonwealth University, Richmond, VA, USA

<sup>4</sup>Department of Behavioral and Policy Sciences, Rand Corporation, Santa Monica, CA, USA

<sup>5</sup>Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA

### Corresponding author:

Bryce D. McLeod, Department of Psychology, Virginia Commonwealth University, 806 West Franklin Street, PO Box 842018, Richmond, VA 23284-2018, USA.

Email: bmcLeod@vcu.edu



## Keywords

measurement development, user-centered design, treatment fidelity

## Introduction

Pragmatic (i.e., practical, brief, accessible, acceptable, technically accurate; Glasgow & Riley, 2013) treatment fidelity measures (i.e., measures that assess if an evidence-based program [EBP] is delivered as designed) can support efforts to implement and sustain EBPs in schools. Yet few pragmatic treatment fidelity measures exist for schools (McLeod, Sutherland, et al., 2022). Observer-report fidelity measures are often considered the “gold standard” (McLeod, Sutherland et al., 2022) but are costly to deploy in schools. Self-report fidelity measures may provide a cost-effective, teacher-centered alternative. However, new approaches to measure development are needed to produce pragmatic self-report fidelity measures.

Usability, or the ease with which a consumer can interact and derive the desired outcome from a product, is a principle of human-centered design (HCD; Lyon et al., 2021). To develop a pragmatic measure, optimizing usability by including end-users, such as teachers, in the development process is essential. Soliciting input from end-users during measure development can increase usability by (a) ensuring items are understandable (Haynes et al., 1995; Ware et al., 2003), and (b) improving feasibility by identifying potential problems (i.e., usability issues) that might interfere with implementation (Lavery et al., 1997).

**Table 1.**  
*TIES-T Item Definitions*

Item	Item definition
Behavior-specific praise	The teacher provides a positive verbal statement(s) of approval in response to appropriate behavior from students.
Precorrection	Before a problem behavior is likely to occur, the teacher proactively reminds students of desired behavior or responses.
Rules	The teacher explicitly uses prescribed guidelines to teach or remind students the rules of the classroom.
Routines	The teacher explains or provides structure for students in daily activities, transitions, or expectations via rituals.
Opportunities to respond	The teacher prompts a specific student response through verbal statements, gestural symbols, or visual cues to promote target behavior.

This article describes a collaborative process involving experts and teachers to revise the Treatment Integrity for Elementary Schools–Teacher Report (TIES-T; Sutherland et al., 2019) to develop a pragmatic fidelity measure teachers can use to monitor the delivery of evidence-informed practices. The TIES-T is part of a suite of observer- and self-report measures designed to capture teachers’ use of practices that support children’s social and emotional competencies.

## Method

### Step 1: Adaptation of the TIES-Teacher

The TIES-T was created to capture universal practices focusing on student behavior during the school day. Expert input and initial teacher interviews informed changes to the TIES-T to support clarity, ease of use, and efficiency. Following these steps, the TIES-T assessed the usage of five practices and student responsiveness. See Table 1 for item definitions.

### Item Selection

We selected five practices from the original 18-item TIES–Observer report (TIES-O) as the focus of the current study for two reasons: First, the project aims to develop a method for producing a pragmatic teacher-report measure that can be adapted for use in different schools, not validate a specific set of items; Second, we aimed to select practices requiring respondents to report on practices used in different classroom contexts (e.g., practices more likely to be used during group instructions, transitions, or small group activities) because these place different demands on the respondents. To select the items, a group of researchers with expertise in school-based research, EBPs, and self-report measurement rated each of the original TIES-T 18 items on their feasibility (i.e., extent to which a practice can be easily deployed in this setting), effectiveness (i.e., extent to which a practice will be successful when deployed), and flexibility (i.e., extent to which a practice can be adaptable) using a 1–5 scale (1 = *not at all*, 5 = *completely*). Observational data collected using the TIES-O in business-as-usual ( $n = 163$ ) classrooms from an efficacy trial (Behavioral, Emotional, and Social Training (BEST) in Competent Learners Achieving School Success (CLASS), Sutherland et al., 2020) were also considered. Together, this information led to selecting five items (Behavior-Specific Praise, Precorrection, Rules, Routines, and Opportunities to Respond) that were rated high on the feasible, effective, and flexible dimensions by experts and were observed in 12.9%–89.6% of observations.

### **TIES-T Scoring Strategies**

A 5-point Likert scale that prompts teachers to rate practices from *not at all used* to *always used* was used to rate each item. A 5-point scale was used to promote clarity in the rating process for teachers (McLeod, Sutherland, et al., 2022). In producing a rating, teachers are prompted to consider the frequency (number of times) and the thoroughness (depth, complexity, or persistence) of their practice use.

### **Training**

A self-guided, video-based training was created to instruct teachers on how to use the TIES-T. Informed by training used for the TIES-O, the training leads teachers through videos, interactive activities, and knowledge checks. The training is divided into three parts: (a) the TIES-T's purpose, (b) the operational definitions of each practice (including student responsiveness), and (c) the information on how to rate each practice.

## **Step 2: Cognitive Walkthrough Interviews**

Similar to the previous phase, HCD principles were used to guide the design of the cognitive walkthroughs, which positioned the experience of the end-user (teacher) as the primary data source. The cognitive interviews were conducted to gauge teachers' perceptions of the system (training and self-reporting) and screen for user errors (Willis, 2004). Each interviewee was given a list of specific reflection questions after interacting with each component of the TIES-T system. When appropriate, the interviewer verbally probed the participants to understand their reactions to the TIES-T and the rationale for their ratings. Their responses, suggestions, and areas of confusion or uncertainty were recorded for theme extraction. A post-baccalaureate research assistant conducted the interviews virtually and did not have existing relationships with the teachers.

### **Participants**

Teacher participants were recruited from two local school districts on the East Coast of the United States. Participants included eight elementary school teachers who taught kindergarten to fifth grade. The participants averaged 11.25 years of teaching experience ( $SD = 9.08$ ; range 2–30). Over half ( $n = 5$ ) had experience teaching in Title I public schools. Our Institutional Review Board reviewed the study and deemed that it did not require informed consent. Teachers opted into the interview process and were paid as consultants for their time.

### **Procedure**

At the start of the interview, teachers were encouraged to imagine themselves within their classrooms. A list of testing scenarios (e.g., “after completing a math lesson,

you are reflecting on your usage of these practices and how your students responded”) was developed to use for the cognitive interviews. As part of the scenario, they completed the self-guided training and the TIES-T while considering the following questions: (a) To what extent do you understand the purpose of the TIES-T? (b) To what extent is the TIES-T training helpful? and (c) How easy is it to use the TIES-T to rate your practice use with students? The interviewer then used a “Think Aloud” process to probe teachers to articulate their reactions and decision-making process in response to the content to determine participant understanding (Ericsson & Simon, 1993; Pressley & Afflerbach, 1995). Feedback related to the TIES-T design and formatting, the TIES-T supporting materials, and the TIES-T training were solicited. Following the completion of each task, the teachers were asked to rate how understandable, time-burdensome, difficult, and useful they found a task on a 6-point scale (1 = *not at all* to 6 = *completely*). Participants also answered questions about each task, such as their perceptions of teacher professional development systems, identifying the most challenging task, and how the system compares to other professional development training they have received. Interviews ran from 90 to 120 min. Findings from the quantitative and qualitative data collection were synthesized to inform improvements to the TIES-T measure for optimal utility in schools. To identify appropriate changes to make to the TIES-T we considered the frequency (i.e., how often the change was suggested), impact (i.e., how severe is the consequence of not implementing the change), scope (i.e., number of users a suggestion would impact), and complexity (i.e., how straightforward a suggestion is to address; Lyon et al. 2021) of the recommendations. Iterative changes were made to the measure between interviews until saturation was achieved after eight interviews, and no further feedback was received. See Table 2.

## **Results**

Qualitative content analysis was used to extract themes from the teacher interviews (Krippendorff, 2004). All interviews were recorded and initially coded by a research staff member and then checked by a second team member. Common themes and response errors emerged through coding the cognitive walkthroughs, structured interview questions, and the quantitative rating scale. Average ratings for the quantitative usability scales were calculated. The quantitative results indicated that the TIES-T was well received. Teachers rated the TIES-T and the training as highly understandable ( $M = 5.90$ ,  $SD = 0.40$ ) and useful ( $M = 5.90$ ,  $SD = 0.30$ ), and minimally time-burdensome ( $M = 1.00$ ,  $SD = 0.16$ ) and difficult ( $M = 1.30$ ,  $SD = 0.74$ ).

### Training Feedback

Overall, teacher feedback regarding training comprised two themes: (a) clarity and (b) interactive elements. A complete list of the changes made is provided in Table 2. As indicated by their usability scores, teachers reported that the training gave a good overview of the project's purpose, how the system can be used in the classroom, and the knowledge and practice needed to rate their classroom accurately. Teacher feedback was focused on elements that could be modified or added to increase the likelihood a teacher could accurately reflect on the practices they use in their classroom.

### Measure Feedback

Teachers found the first draft of the self-report measure concise and valuable. Some voiced concerns about remembering the "coding caveats" (e.g., rate all students in the class, rate practices independently, focus on behavior), so these reminders were added to the top of the measure. Teachers cited time (i.e., finding the time to complete the training and rate their classroom), and difficulty making rating decisions (i.e., usage and student responsiveness) as the most significant barriers to using the TIES-T system in their classroom.

## Discussion

Pragmatic fidelity measures can support EBP implementation and sustainment in schools (McLeod, Cook, et al.,

2022). This study described a measure development approach with teachers informed by HCD principles intended to produce a pragmatic self-report fidelity measure. Teachers rated the training and TIES-T as understandable, feasible, and minimally time-consuming. Teacher feedback mainly focused on ways to improve the clarity of TIES-T and the importance of adding interactive elements to the training.

Teachers noted that a brief, online, hour-long training was feasible for their work week. Features of the training, such as the ability to complete 20-min segments at a time, offered a preferred alternative to in-person training, highlighting the importance of providing short, flexible training options to practitioners who typically have busy schedules that make attending in-person training difficult. Teachers also felt the length of the training was adequate to prepare them to use the TIES-T. However, future work must be done to determine if the TIES-T training is effective.

Given the length of the TIES-T, teachers indicated that using the measure during their workday was feasible. Given the length of the TIES-T, teachers indicated that using the measure during their workday was feasible. They felt the brevity and clarity of the TIES-T would make it possible to use in their classrooms. This feedback helps to underscore the importance of developing "pragmatic" measures for teachers to use in their classrooms (Glasgow & Riley, 2013; Lewis et al., 2021).

**Table 2.**

*Summarized Teacher Informed Changes to the TIES-T System.*

Project component	Change implemented	Description
Training feedback	Reduce jargon in practice definitions	<ul style="list-style-type: none"> <li>Substituting "observer" for "project facilitator."</li> <li>Substituting "observation" for "classroom visit."</li> </ul>
	Clarify practice definitions	<ul style="list-style-type: none"> <li>Shorten practice definitions and bold key phrases</li> <li>Praise improved definition: "The teacher provides a positive verbal statement(s) of approval in response to appropriate behavior from students."</li> </ul>
	Reflection questions	<ul style="list-style-type: none"> <li>Added additional interactive activity</li> <li>Teachers brainstorm how a concept applies to their classroom ("What does problem behavior usually look like in your classroom?")</li> </ul>
	Example and non-example activity	<ul style="list-style-type: none"> <li>After reviewing the practice definitions, teachers are presented with a list of short scenarios and must decide if it fits the practice definition and why.</li> </ul>
	Explain examples and non-examples	<ul style="list-style-type: none"> <li>Voiceover explanation of the "why" behind examples.</li> <li>Directly reference practice definition in reasoning.</li> </ul>
	Scenario rating activity	<ul style="list-style-type: none"> <li>Once learning the rating scales, teacher reads a scenario describing a full academic block.</li> <li>They then must decide how the teacher should rate this situation</li> <li>They then hear the correct rating and justification.</li> </ul>
Measure feedback	Practice rating activity	<ul style="list-style-type: none"> <li>Teacher practices using the measure by predicting what scores they may rate in their classroom for a particular subject.</li> </ul>
	Edit practice definitions Rating reminders summary	<ul style="list-style-type: none"> <li>Shortened practice definition list</li> <li>Summarized one-line reminders to avoid common biases: reflect on practice usage with all students only rate practices to promote behavioral competencies within your academic subject or time block</li> </ul>

Teachers provided important feedback for the training. They noted that reducing jargon and succinct definitions and explanations would allow teachers to spend less time trying to understand the concepts and more time trying to understand how to apply them to their classroom. This is consistent with recommendations to engage end-users in the design process to ensure wording is understandable (Haynes et al., 1995; Ware et al., 2003). The addition of interactive training elements was seen as important to enhance learning. This recommendation is consistent with adult learning principles that use multiple learning approaches to optimize learning (McLeod et al., 2015).

This study illustrated how HCD principles could engage end-users in modifying a measure for use in schools. Soliciting end-user feedback represents the first step in developing pragmatic measures for use in schools. The collaborative, iterative review process described in this article informed by HCD principles, may help advise future efforts to develop pragmatic measures for community settings. This process can be used early in the development process to help optimize the design of the measure for the context in which it will be deployed.

To establish if the TIES-T is a pragmatic measure more work is needed. First, pragmatic measures must demonstrate score reliability and validity (Glasgow & Riley, 2013), so future work will need to investigate the initial psychometric properties of the measure. Second, though the cognitive walkthroughs are useful, it is possible that the teachers may produce different ratings if they actually used the TIES-T in their classroom. Future work will therefore need to collect feasibility ratings from teachers who have used the TIES-T in their classrooms along with recommendations for ways to improve usability.

Another direction for future research is investigating the impact organizational factors have on using innovations, like the TIES-T. Teachers voiced concerns over how contextual factors, such as school or organizational climate, may impact a teacher's ability or willingness to complete the measure, particularly regarding unsupportive administration. This is consistent with the research literature that suggests these constructs may affect an individual's ability to implement an intervention with fidelity (Aarons et al., 2015; Glisson & Durick, 1988).

In summary, this article describes the collaborative process informed by HCD principles we used to iteratively adapt the TIES-T. This process models a methodology applicable across fields to initiate the development of self-report fidelity measures from validated observer measures with a community-centered focus. The feedback we received was used to increase the acceptability, appropriateness, and feasibility of the TIES-T. Next steps in this research will include piloting the training and the TIES-T to estimate preliminary score reliability and validity and seek additional feedback regarding acceptability, appropriateness, and feasibility.





## Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Dr. Aaron R. Lyon serves as an Associate Editor for *Implementation Research and Practice*. As such, Dr. Lyon was not involved in any aspect of the peer review process for this manuscript. All other authors declare no conflicting interests.

## Funding

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Preparation of this article was supported in part by a grant from the Institute of Education Science (R305A210168; Bryce D. McLeod).

## ORCID iDs

Emma R. Dear  <https://orcid.org/0000-0003-2397-365X>  
 Bryce D. McLeod  <https://orcid.org/0000-0002-0996-0492>  
 Alex R. Dopp  <https://orcid.org/0000-0002-2522-6546>  
 Aaron R. Lyon  <https://orcid.org/0000-0003-3657-5060>

## References

- Aarons, G. A., Ehrhart, M. G., Farahnak, L. R., & Hurlburt, M. S. (2015). Leadership and organizational change for implementation (LOCI): A randomized mixed method pilot study of a leadership and organization development intervention for evidence-based practice implementation. *Implementation Science, 10*(1), 1–12. <https://doi.org/10.1186/s13012-014-0192-y>
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*. The MIT Press.
- Glasgow, R. E., & Riley, W. T. (2013). Pragmatic measures: What they are and why we need them. *American Journal of Preventive Medicine, 45*(2), 237–243. <https://doi.org/10.1016/j.amepre.2013.03.010>
- Glisson, C., & Durick, M. (1988). Predictors of job satisfaction and organizational commitment in human service organizations. *Administrative Science Quarterly, 33*(1), 61–81. <https://doi.org/10.2307/2392855>
- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238–247. <https://doi.org/10.1037/1040-3590.7.3.238>
- Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research, 30*(3), 411–433.
- Lavery, D., Cockton, G., & Atkinson, M. P. (1997). Comparison of evaluation methods using structured usability problem reports. *Behaviour & Information Technology, 16*(4–5), 246–266. <https://doi.org/10.1080/014492997119824>
- Lewis, C. C., Mettert, K. D., Stanick, C. F., Halko, H. M., Nolen, E. A., Powell, B. J., & Weiner, B. J. (2021). The psychometric and pragmatic evidence rating scale (PAPERS) for measure development and evaluation. *Implementation Research and Practice, 2*, 263348952110373. <https://doi.org/10.1177/26334895211037391>
- Lyon, A. R., Coifman, J., Cook, H., McRee, E., Liu, F. F., Ludwig, K., Dorsey, S., Koerner, K., Munson, S., &

- McCauley, E. (2021). The cognitive walkthrough for implementation strategies (CWIS): A pragmatic method for assessing implementation strategy usability. *Implementation Science Communications*, 2(1), 1–16. <https://doi.org/10.1186/s43058-020-00099-1>
- McLeod, B. D., Cook, C. R., Sutherland, K. S., Lyon, A. R., Broda, M., & Bedias, R. (2022). A theory-informed approach to locally managed learning school systems: Integrating treatment integrity and youth mental health outcome data to promote youth mental health. *School Mental Health*, 14, 88–102. <https://doi.org/10.1007/s12310-021-09413-1>
- McLeod, B. D., Sutherland, K. S., Broda, M., Granger, K. L., Cecilione, J., Cook, C. R., Conroy, M. A., Snyder, P., & Southam-Gerow, M. A. (2022). Examining the correspondence between teacher- and observer-report treatment integrity measures. *School Mental Health*, 14(1), 20–34. <https://doi.org/10.1007/s12310-021-09437-7>
- McLeod, B. D., Wood, J. J., & Klebanoff, S. (2015). Advances in evidence-based intervention and assessment practices for youth with an autism spectrum disorder. *Behavior Therapy*, 46(1), 1–6. <https://doi.org/10.1016/j.beth.2014.07.004>
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Lawrence Erlbaum Associates.
- Sutherland, K. S., Conroy, M. A., & Granger, K. (2020). BEST in CLASS: A Tier-2 program for children with and at risk for emotional/behavioral disorders. In T. W. Farmer, M. A. Conroy, E. M. Z. Farmer, & K. S. Sutherland (Eds.), (Eds.), *Handbook of research on emotional and behavioral disorders: Interdisciplinary developmental perspectives on children and youth* (pp. 214–226). Routledge/Taylor & Francis Group.
- Sutherland, K. S., Conroy, M. A., McLeod, B. D., Granger, K., Nemer, S. L., Kunemund, R. L., Johnson, A., & Miles, C. (2019). Adapting an evidence-based early childhood tier 2 program for early elementary school. *The Elementary School Journal*, 119(4), 542–561. <https://doi.org/10.1086/703103>
- Ware, J. E. (2003). Conceptualization and measurement of health-related quality of life: Comments on an evolving field. *Archives of Physical Medicine and Rehabilitation*, 84, S43–S51. <https://doi.org/10.1053/apmr.2003.50246>
- Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications.