What Should a Restorative Classroom Look and Sound Like? Content Validation of a

Direct Observation Tool

Claudia G. Vincent, Heather McClure, Rita Svanks, Erik Girvan, and John Inglish

University of Oregon

Darren Reiley and Scott Smith

Center for Dialogue and Resolution

Citation:

Peer Review Process:

https://www.emeraldgrouppublishing.com/journal/jrit

Structured Abstract

Purpose: Our study focused on identifying measurable constructs of a restorative classroom and appropriate metrics to measure those constructs through content validity analysis of a direct observation tool. The tool was designed to assess restorative practices implementation in the classroom in the context of professional development supporting teachers in a fundamental reorientation towards non-punitive discipline.

Study Design/Methodology/Approach: We administered a 30-item survey to a panel of 14 experts in restorative practices implementation in schools asking them to provide quantitative and qualitative feedback on the tool's content, metrics, and utility for building teachers' skill and confidence in promoting a restorative classroom. We calculated item-level content validity indeces and scale-level content validity indeces. To interpret findings, we applied acceptability criteria recommended in the literature. We used qualitative coding to analyze qualitative responses and contextualize quantitative findings.

Findings: Quantitative results indicated that the tool's structure and measures of teacher behavior were acceptable. The student behavior scale did not meet the acceptability criterion. Qualitative feedback indicated that observation and later co-reflection on teachers' use of specific restorative skills was deemed helpful to teacher implementation of restorative practices. Observations of student behaviors, however, needed to be broadened to emphasize student voice and agency and the quality of student interactions.

Originality: Novel approaches to measurement are needed to facilitate teacher implementation of restorative practices as schools adopt those practices to promote equitable student agency, engagement, and belonging in a pivotal shift from existing punitive discipline systems.

Key words: restorative practices, implementation, direct observation, content validity

What Should a Restorative Classroom Look and Sound Like? Content Validation of a Direct Observation Tool

School personnel are looking toward restorative practices (RP) as a shift away from punitive discipline and towards relationship- and trust-focused interactions foundational to student engagement and equitable student discipline outcomes (Lodi *et al.*, 2021). Derived from indigenous approaches to community building and conflict resolution through acknowledging the impact of a person's actions on the entire community and making amends (Zehr, 2015), RP was introduced into schools in Australia in the 1990s (Blood and Thorsborne, 2005) and soon after in the United States (Gregory *et al.,* 2018).

Restorative Practices

RP is derived from Restorative Justice which focuses on victims' and offenders' needs and obligations following harm and offers an alternative to retributive justice in judicial settings (Zehr, 2015). In schools, RP focus on teachers and students working together to understand harmful behavior, its impact, and accountability as an alternative to punitive discipline focusing on rule violations and consequences (Gregory *et al.*, 2018).

Conceptualized along a multi-tiered support continuum, RP consists of trust- and relationship-building at Tier 1, relationship affirmation at Tier 2, and conflict resolution at Tier 3 (Vincent *et al.* 2016). Tier 1 practices include active listening (Amstutz, 2015), using affective language to communicate impact of behavior on self and others, (Amuseghan, 2009) and participating in and shared ownership of community-building circles (Evanovich *et al.*, 2020). Tier 2 and 3 practices include restorative chats and restorative circles to address conflict and encourage accountability (Amstutz, 2015). Tier 1 relationship and community building practices are foundational to RP implementation (Rideout et al., 2010).

The Evidence Base and Measurement Challenges

Non-experimental studies document promising associations between RP implementation and desirable outcomes, but also highlight measurement challenges. Examining RP in schools across the United States, Guckenburg and colleagues (2016) associated RP with improvements in school climate and teacher-student relationships based on survey and interview data, but noted that participants' responses might be based on divergent definitions of restorative practices. Gregory and colleagues (2018) found reductions in racial disparities in discipline based on suspension data correlated with students' participation in restorative circles or conferences, but noted that findings were limited due to unknown fidelity of implementation of restorative interventions. Based on interviews with high school students and personnel, Ortega and colleagues (2016) associated RP with positive peer relationships, but noted that their results were collected with unvalidated measures. Exploring the relationship between RP implementation and disciplinary equity in one school district, Davison and colleagues (2022) found few reductions in racial disparities and noted that unknown variability in implementation across schools and classrooms might be a factor.

The results of randomized controlled trials were mixed and also highlighted measurement challenges. Augustine and colleagues (2018) found causal relationships between RP and decreases in suspension rates and racial disparities in discipline based on staff surveys, interviews with staff and RP trainers and direct observation of trainings consisting largely of field notes on student and teacher behaviors. They stressed the challenge of measuring RP fidelity of implementation, because RP tend to be adapted to individual school contexts, cultures, and personal teaching styles. Acosta et al (2019) found no significant intervention effect based solely on student survey data, although students' perceptions of RP were associated with

improved school climate and connectedness. The authors acknowledged the challenge of collecting observational data on RP components that tend to be implemented in an impromptu fashion, and called for future research on observation measures that could strengthen the evidence base for RP. Huang et al (2023) found no relationship between RP implementation and the likelihood of suspension for students from varying racial/ethnic backgrounds. The authors attributed their findings partially to insufficient measurement of classroom RP practices and called for direct classroom observations.

Existing Direct Observation Measures

Tools to observe teachers' RP implementation, or the quality of classroom relationships more generally, exist; however, they do not produce consistent results. For example, Gregory and colleagues developed *RP Observe* (Gregory *et al.*, 2014) to assess the extent to which teachers implemented proactive classroom circles and restorative circles or conferences. Key constructs measured include the structure of circles and conferences to ensure all participants' emotional safety, support and belonging, the presence of student voice, and opportunities for social-emotional learning. The *Classroom Assessment Scoring System* (CLASS, Pianta *et al.*, 2012) assesses the quality of teacher-student interactions and relationships in general. It measures emotional support, classroom organization, and instructional support; each domain has been found to have good reliability (α = .77, .82, and .73 respectively) in secondary classrooms (Allen *et al.*, 2013). Although in theory related, in practice, *RP Observe* scores have been found not to correlate with *CLASS* scores, indicating that the two measures assess unique constructs (Gregory *et al.*, 2014). Fisher (2020) developed the *Restorative Practices Classroom Observation* tool to measure teacher student interactions. The tool yielded good inter-rater reliability in secondary school classrooms (ICC = .755 for teacher observations and .934 for

student observations) and acceptable internal reliability ($\alpha = .85$ for negative interactions and .50 for positive interactions). Based on the measure's limited success in assessing the impact of RP training on teacher and student behavior, the author identified the wide variability in RP practices as a measurement challenge.

While researchers seem to agree on the need for direct observations of RP classroom practices, there seems to be a lack of consensus on what observable teacher and student behaviors define a restorative classroom (Guckenburg *et al.*, 2015). To address this challenge, we conducted a content validation study of a recently developed Tier 1 RP classroom observation tool. The primary purpose of the tool is to assess fidelity of implementation; the secondary purpose of the tool is to support implementation efforts through providing performance feedback to teachers. Joyce and Showers (2002) noted that fidelity of implementation of learned practices increases substantially if teachers receive coaching and feedback. The tool was developed together with professional development in *School-wide Positive and Restorative Discipline.* The majority of the training focused on Tier 1 practices, namely relationship-building through active listening, use of affective language, reframing, and conducting community-building circles. Teachers were given access to a *Circle Planning Tool* to prepare proactive circles in their classrooms. More detail on the training materials' development and their delivery is available in Vincent *et al.* (2021a) and Vincent *et al.* (2021b).

SWPRD Fidelity of Implementation Classroom Observation Tool

Given the importance of strong implementation of Tier 1 RP (Rideout *et al.*, 2010), and the literature's recommendation that RP should change both teacher and student behavior (Gregory and Evans, 2020), the tool offers a comprehensive assessment of teacher-student interactions during an entire class period. It is intended to be completed by a direct observer in

collaboration with the teacher whose classroom is observed. Observations should be scheduled during a class period when the teacher plans to conduct a proactive classroom circle. Observation sessions identifiers include the teacher's and observer's name, date, class period, grade level, number of students present, and subject taught. The tool consists of three sections: (1) Pre-observation meeting, (2) Observation of the class period, and (3) Post-observation teacher debrief meeting.

During the pre-observation meeting (section 1) the observer reviews the *Circle Planning Tool* completed by the teacher and rates the extent to which the planned circle's goal and learning objective have been identified, clarity of the opening statement, and the circle prompts' relation to the school's overall behavioral expectations and to the goal and learning objectives for the circle. The observer asks the teacher how many and what types of circles (e.g. relationship building, academic instruction, defining behavioral expectations) he/she/they have led during the current term. The observer also assesses the physical classroom environment, that is, whether behavioral expectations/values and circle guidelines are posted in the classroom, and whether the furniture placement is conducive to the circle process. A total of 10 minutes is allocated to the pre-observation meeting.

Section 2 (observation of the class period) is completed by the observer alone. It consists of four domains: (a) proactive circle: teacher behaviors, (b) proactive circle: student behaviors, (c) remaining class period: teacher behaviors, (d) remaining class period: student behaviors. The observer records key teacher behaviors associated with circle keeping (e.g., stating the purpose of the circle, stating the circle guidelines, adhering to the circle guidelines, intervening when guidelines are violated, stating prompts, providing a closing statement) on a 3-point Likert-type scale ranging from "did not display" to "displayed partially" to "displayed thoroughly." The

observer also tallies the number of students who engage in recommended circle behaviors (e.g., responding to the prompt, speaking when they have the talking piece) as well as students who violate circle guidelines (e.g. speaking when they do not have the talking piece, commenting on others' statements). Finally, the observer records how much time is spent on setting up the circle and transitioning to the next classroom activity. For the remaining class period, the observer assesses the extent to which the teacher models restorative practices, such as active listening, affective language to respond to positive and negative student behaviors, reframing, and referring back to the circle discussion on a 3-point Likert-type scale ranging from "did not display" to "displayed partially" to "displayed thoroughly." The observer also tallies the number of students engaging in those same practices. Tables 3 and 4 list the observed behaviors.

Section 3 (post-observation teacher debrief meeting) of the tool is completed by the observer in collaboration with the teacher. The observer asks the teacher how he/she/they felt about the proactive circle (e.g., what went well, what was challenging), and how he/she/they felt about the remaining class period (e.g., what went well, how challenges might be prevented in the future). Finally, the observer asks what support the teacher needs to address challenges with RP implementation in the classroom. Ten minutes are allocated to the post-observation meeting.

Direct observation is considered the most direct approach to measuring the association between an intervention and it's intended outcomes (Lewis *et al.*, 2014), and therefore critical to establishing evidence-based pratice. To maximize the usefulness of direct observation data, direct observation instruments need to capture key constructs in a meaningful and easy to interpret metric (Sanetti and Collier-Meek, 2014).

Content Validation

Oluwatayo (2012) identifies examination of content validity as a critical initial step in the

psychometric evaluation of measures used in education research. Establishing what observable behaviors are associated with RP and how they should be measured is precisely what is currently needed to strengthen the evidence base for RP (Acosta *et al.,* 2019; Augustine *et al.,* 2018; Huang *et al.,* 2023).

Content validity indicates the extent to which an instrument measures constructs of interest (Yusoff, 2019). Content validity is commonly determined through an expert panel reviewing the tool's items for their relevance, necessity, and usefulness, and the overall tool for its adequacy and coverage of the construct of interest (Oluwatayo, 2012). Agreement among the experts is calculated to indicate individual items' and subscales' content validity (Yusoff, 2019). Consistent with this guidance, our study was driven by the following research questions:

1. Is the overall structure of the tool appropriate to assess its content?

2. Is the content of the Pre-observation section of the tool appropriate?

3. Are the teacher behaviors to be assessed during proactive circle relevant, necessary, and useful?

    a. Is the scoring metric appropriate and useful?

4. Are the student behaviors to be assessed during proactive circle relevant, necessary, and useful?

    a. Is the scoring metric appropriate and useful?

5. Are the teacher and student behaviors during the remaining class period appropriate?

6. Is the content of the Post-observation section of the tool appropriate?

7. What is the overall assessment of the tool?

Method

Participants

Fourteen experts with expertise in restorative practices implementation in schools from school districts in the Pacific Northwest and the Mountain Region, and a scholar from the Mid-Atlantic Region participated. Ten experts identified as female and four as male. One identified as Hispanic/Latino, 11 as non-Hispanic/Non-Latino, and 2 preferred not to identify their ethnicity. One identified as American Indian/Alaska Native, one as African-American/Black, 13 as Caucasian/White, and one preferred not to answer. Respondents could choose more than one racial background. One expert completed postgraduate work and 13 had a graduate degree. Current job titles included Advisor for Affinity Unions, Equity and Inclusion Leader, Restorative Justice Practitioner, Cultural Service Coordinator, Assistant Principal, Dean of Students, Director of Restorative Justice, District Restorative Practices Coordinator, Principal, Professor, Program Specialist, School Counselor/Teachers, Special Programs Administrator, and Teacher. Experience in the current position ranged from 1-2 years (4 respondents), to 3-5 years (4 respondents), 6-10 years (4 respondents) and over 10 years (2 respondents). Experts who were school personnel taught elementary grades (3 respondents), middle school grades (1 respondent) and high school grades (6 respondents). School personnel participants taught English, Social Studies, Science, Music, Theater, and Alternative Education.

Measure

Experts completed a 30-item *Content Validity Survey* specifically designed for the current study. The survey introduced the respondent to the larger context in which the tool was developed, namely the SWPRD teacher training. Experts were asked to familiarize themselves with the direct observation tool as well as the Circle Planning Tool, and then complete the survey based on their expertise and professional judgement. The survey consisted of five parts: (1)

Overall structure of the tool (6 items), (2) Pre-observation meeting (5 items), (3) Observation (class period) (14 items), (4) Post-observation meeting (3 items), and (5) Overall assessment (2 items). Of these 30 items, 19 asked experts to provide a rating on a Likert scale ranging from 1 = strongly disagree/very unlikely/not at all relevant/necessary/useful to 4 = strongly agree/very likely/very relevant/necessary/useful. When rating the relevance, necessity and usefulness of observing specific teacher and student behaviors, experts were provided the following definitions:

- Relevance: The behavior describes or relates to something I consider restorative practice.

- Necessity: The behavior describes or relates to something essential to competent execution of a restorative practice.

- Usefulness: The behavior describes or relates to something that would be helpful for teachers to receive feedback on as they are learning to implement restorative practices.

Eleven items asked experts to provide write-in responses.

Study Procedure

After obtaining approval from our Human Subjects Institutional Review Board (IRB), we recruited experts in February of 2021. After experts indicated their interest in participating in the study in response to a recruitment email, they were sent a Qualtrics link to an online consent form approved by the IRB. Once they provided informed consent, they were sent the materials to review via email: The *SWPRD Fidelity of Implementation Classroom Observation Tool* and the *Circle Planning Tool* together with a link to the *Content Validity Survey*. Experts were given four weeks to review the materials and complete the on-line *Content Validity Survey*. Not all experts responded to all items. The study concluded in April 2021.

Analytical Procedures

To analyze the quantitative data, we followed Yusoff's (2019) recommendations. We

calculated item-level content validity indeces (I-CVI) by dividing the number of experts who

rated an item 3 or 4 by the total number of expert responses. We calculated scale-level content

validity indeces by calculating the average of the I-CVIs for the scale (S-CVI/Ave). To interpret

findings, we used Yusoff's (2019) acceptability criterion of ≥.78, based on an expert panel

consisting of at least nine experts.

We used Dedoose to code all write-in responses. Two coders first identified simple codes

of "meaningful to the creation of a restorative classroom," "meaningful to the fidelity monitoring

of a restorative classroom" and two codes reflecting "not meaningful" to the creation of a

restorative classroom and fidelity monitoring, respectively. Second, we completed selective

coding, meaning we used the developed coding framework to re-review the write-in responses

and identify any content we might have missed and adjust the coding framework as necessary.

Third, we used "memo-ing" to link theory or umbrella concepts to the ideas presented by the

participants. Finally, we reviewed draft write-ups to cross-check and validate our codes and

deepen our interpretation of the data (Charmaz, 2006; Corbin & Strauss, 1990).

Results

Section 1 of the *Content Validity Survey* asked experts to rate the overall structure of the

tool. It consisted of 4 quantitative and 2 qualitative items. Quantitative results are presented in

Table I.

<Insert Table I about here>

I-CVIs ranged from .69 to 1.0, with a S-CVI of .87. Experts rated the tool's print-layout lowest,

and the logical sequencing of the tool's sections highest. Qualitative responses focused on

changes to the form's layout that would help the observer capture the data more efficiently and

accurately for each of the three sections. For example, two expert reviewers suggested fitting the

teacher behavior measures and student behavior measures next to each other on one page and reducing the number of the tool's pages overall. In addition, one expert appreciated the shortness of the pre-meeting and suggested that it be reduced from 10 to 5 minutes. Five of the experts recognized the challenge of capturing student behaviors and recommended simplifying the types of behaviors observed, One expert appreciated the teacher-centered post-observation debrief and the opportunity for a rich discussion. Experts also recommended more detailed information about the students present (e.g., gender, ethnicity, disabilities, pronoun preference) and type of classroom (special education, emergent bilingual class, or advisory homeroom).

Experts next rated the Pre-observation meeting items. Table II summarizes the results.

<Insert Table II about here>

I-CVIs ranged from .77 to .79 with an S-CVI/Ave of.78.. Write-in items encouraged experts to comment on what items should be omitted or added to evaluate the teacher's use of the Circle Planning Tool and to evaluate the physical classroom environment. One expert appreciated the sample prompts the tool provides. Five experts focused on ways to ensure that teachers understood each planning item by recommending additional explanations and examples. In commenting on the classroom environment items, experts were uncertain about the value of observing furniture placement; one expert commented: "I never would have been able to place my furniture in a way that was conducive to circles, so we just stood around the desks or moved into the hallway…"

Observation of the classroom period focused on teacher and student behavior during a proactive circle and the remaining class period. Experts first rated the relevance, necessity, and usefulness of observing teacher behaviors during a proactive circle, and then the appropriateness of the scoring metric. Results are summarized in Table III.

<Insert Table III about here>

Expert ratings of the relevance, necessity, and usefulness of observing teacher behaviors were high, with an S-CVI/Ave of .91 for necessity, .92 for relevance, and .93 for usefulness. Experts' rating of the appropriateness of the scoring metric (i.e., a 3-point scale ranging from "did not display" to "displayed partially" to "displayed thoroughly") was lower at .75.

Experts provided additional comments on the Proactive Circle Teacher Behaviors Scale. Experts voiced concerns about the lack of emphasis on student voice. Described one reviewer, "my belief is that a restorative circle (especially proactive) should encourage and build (the) capacity of students to introduce prompts...Without this, student participation is 'managed' and does not build (the) capacity of the students to own their circle." Another reviewer agreed saying, "the heart of the work" would ask about student-designed and facilitated circles and the sharing of "highly relevant youth experience of political events, equity, -'isms, social media, etc." Another expert stated "teachers' level of regulation, comfortability with conflict, and use of equitable practices in all settings is what will make for a wholly restorative space." In this context, experts recommended changing the language in the items from "adhered to circle guidelines to "follows circle guidelines as appropriate," a rewording that recognizes teachers' abilities to capitalize on teachable moments that might arise when students deviate from formal guidelines.

Experts next rated the metric to assess student behavior during a proactive circle, and the relevance, necessity, and usefulness of observing the listed student behaviors during a proactive circle. Results are summarized in Table IV.

<Insert Table IV about here>

Ratings for the metric to observe student behaviors were .77 and .79, with an S-CVI/Ave of .78.

When asked if tallying occurrences of a behavior or the number of students engaging in a behavior would be more appropriate, four experts recommended occurrences and nine recommended number of students. Experts pointed to the importance of identifying qualities over quantities of interaction. For instance, a simple count of students who respond to the prompt may miss students who are listening keenly and opt to pass; experts emphasized that for students' agency to be truly respected, their participation must be broadened to include active listening, which can be difficult to observe and quantify. Wrote one expert, "Sometimes a kid will shout out an answer at the wrong time, but it is helpful to the discussion. That shouldn't count equal to a kid swearing and storming out of the classroom." An expert suggested adding a subjective overall evaluation of the quality of the circle as "it would be helpful to (know)... how productive the circle time was. A circle might have very few interruptions or distractions because nobody is engaged in what's happening, even if they answer when prompted." Finally, though one expert thought that quantifying student behaviors during a circle would be less vulnerable to bias, another noted that this approach does not allow the observer, upon later reflection, to know whether the number of behaviors were due to a single struggling student or one behavior exhibited by many students. Experts rated the relevance, necessity, and usefulness of the student behaviors much lower than the teacher behaviors. The S-CVI/Ave for relevance was .64, for necessity .60, and for usefulness .72.

Experts then reviewed a number of teacher and student behaviors exhibited during the remaining class period and rate the extent to which those behaviors promote a restorative classroom. Experts also rated the metrics to be used to rate the teacher behaviors ("did not display," "displayed partially," "displayed thoroughly") and the appropriateness of tallying the number of students displaying the behaviors. Table V presents outcomes.

<Insert Table V about here>

Ratings were high with I-CVIs ranging from .71 to .93, and an S-CVI/Ave of .86.

Experts could comment on what teacher and student behaviors should be omitted or added; two experts noted that teachers' references to the circle were not necessary for the circle to have had high fidelity. Instead, experts suggested that it would be more useful to note the degree to which students talk and engage following a circle (compared with "teacher talk") and to observe teachers' use of restorative communication skills. These skills included teachers valuing each student's contribution to the class, using "empathetic" language with responses characterized by "What I hear you saying...", and giving a positive welcome when students arrive to the classroom using students' name and pronouns, and a positive exit from the class with encouragement for the next circle.

Experts queried whether a tally system of student behaviors provided useful insight into whether a classroom was restorative. Instead, they pointed to the utility of noting whether general interactions took place, such as whether students generally demonstrated good listening for the duration of the class, whether they generally exhibited behavior that was helpful/respectful toward their classmates and teacher, whether students generally waited for a turn to speak and did not interrupt one another for the rest of class, and kept distractions and side conversations to a minimum. Another expert asked whether students were being taught how to actively listen, reframe, or use affective language. If not, it seemed less useful to assess those skills.

Experts suggested alternative scoring metrics and said that "Displays Partially" could be replaced with "Displays Occasionally" and "Displays Frequently" to allow for more nuance between "never" and "always". One reviewer noted that the class structure following the circle

might not allow for extensive student discussion thus making impossible observations of

students' restorative communication skills. Finally, reviewers noted the importance of

longitudinal observations to capture the potential of circles to contribute to changes in classroom

culture with greater emphases on student voice/agency over time.

Experts rated the content of the Post-Observation Meeting section of the tool high (I-CVI

= 1.00), but the time allotted (i.e.,10 minutes) low (I-CVI = .57. The S-CVI/Ave was .79. Table

VI summarizes these outcomes.

<Insert Table VI about here>

When asked what questions should be omitted or added to this section, four experts

recommended that questions should be less generic. On the question of, "Did the circle

accomplish its intended objective," three experts suggested digging more deeply into the

observable outcomes and impact of the circle. Experts also suggested rephrasing questions so

that teachers can reflect on the next steps and needed supports.

Finally, experts provided an overall assessment by rating teachers' likelihood to

participate in an observation guided by the tool. Results are provided in Table VII.

<Insert Table VII about here>

The I-CVI of .79 indicated overall acceptability. Qualitative responses focused on the tool's

usefulness for teachers' ongoing learning to promote a restorative classroom. Experts' final

feedback focused on simplifying the tool's design so that observations of teacher and student

behavior during and following a circle could serve as a starting point for discussions between

observer and teacher about lessons learned and next steps. For instance, experts recommended to

focus on student behaviors related to "build(ing) relationships and community" rather than on

impacts, as these were geared more toward responsive circles rather than proactive ones. Experts

emphasized the need for items to reflect student agency rather than behavioral compliance. Asked one reviewer, "Is a circle not a good circle if students speak out of turn? Or if they don't respond directly to the prompt but are instead inspired to share something in response to what another student brought up?" Reviewers recommended against favoring student rule adherence and verbal participation. Two experts emphasized the importance of extensive coaching and relationship-building between the observer and teacher prior to using this tool so that teachers are eager to engage in a reflective learning process with the observer rather than feel they are being evaluated and set up for failure. Finally, reviewers expressed concern that there might be little, if any, time to debrief before or after a class meeting.

Discussion

Existing studies of RP ascribed measurement challenges to a lack of consensus on observable teacher and student behaviors. Our study gathered expert feedback on what observable teacher and student behaviors are associated with Tier 1 RP practices as a first step towards developing a direct observation tool.

The quantitative outcomes of our study suggest that most portions of our tool had overall adequate content validity. Experts' ratings of the tool's overall structure, pre-observation meeting, proactive circle teacher behaviors, teacher and student behaviors during the remaining class period, post-observation meeting, and overall assessment either met or exceeded the acceptability criterion of .78 recommended by Yusoff (2019). The only scale that did not meet the acceptability criterion was student behaviors during proactive circles. Qualitative feedback indicated that experts felt students' adherence to circle rules should not be a measure of success. Students should be allowed to engage with the circle process more freely to find their level of comfort with and benefit from it. A student speaking up without having the talking piece might

demonstrate engagement; a student opting not to speak might be actively listening to others'

comments. These findings underscore the complexity of measuring student engagement in

intentional community building. To address this measurement challenge, qualitative feedback

from experts strongly recommended a heightened emphasis on student voice and agency through

focusing on the quality of circle interactions among participants.. This suggests that qualitative

feedback from skilled observers might be a better measure than quantitative scales.

Experts indicated that the tool, with revisions, might be useful for teachers to further

reflect upon and improve their work to foster a restorative classroom climate. In general,

reviewers commented that observing teachers' use of specific skills (e.g., active listening,

affective language) could prove useful to teachers, though observations of restorative behaviors

among students clearly needed to be broadened. In post-circle class time, experts noted the

importance of teachers modeling restorative practices and relating in a restorative manner,

including expressing warmth and clear direction during transitions, such as when students were

exiting (or entering) the classroom, a time of heightened vulnerability for students in terms of

self-regulation. Expanding the tool to capture measurement of these teacher behaviors might be

useful.

Finally, experts noted the value of a brief post-observation chat marked by the transparency of

the observer regarding their own observations and authentic curiosity regarding the teachers'

behavior and choices during and following the class circle. Experts noted that skills such as the

use of affective language can be difficult to master. This post-observation period might provide

the observer with a valid opportunity to double-check observations, and for teacher and observer

to mutually identify successes and challenges, problem-solve and identify needed support, and

consider next steps in the implementation of difficult skills. Thus, experts clearly resonated to the

two uses of the tool: They noted that the tool can serve as a fidelity measure in the context of a research study. Alternatively, it can be used to support teachers' use of RP in the classrooms.

While the results of our study provided overall support for the content validity of our tool, they also reflected challenges associated with direct observation of classrooms in general (Lewis *et al*., 2014). Experts noted high observer load, with one observer observing teacher and student behaviors simultaneously. Classrooms are complicated social microcosms where countless variables interact with each other. Choosing which variables to observe and which to omit can be challenging. Similarly, the absence of an observable variable (e.g. student responding to a prompt) could mean the presence of another variable (e.g. active listening) that is difficult to observe yet representative of the construct of interest. Taken together, the findings from our study provide importance guidance on direct observation of Tier 1 RP practices: Experts agreed on what observable teacher behaviors constitute a restorative classroom. Experts' feedback on what student behaviors constitute a restorative classroom points towards more research on how to observe student agency in and ownership of restorative practices in a classroom setting. It might mean that a series of direct observations might be necessary to observe trends and incremental changes in student (and teacher) behavior to capture the gradual emergence of a restorative classroom.

Limitations

Results should be interpreted in the context of the following limitations. First, we recruited a larger number of experts than the recommended maximal ten (Yusoff, 2019). Because we wanted to capture experts with various roles in implementing school-wide initiatives (teachers, administrators, student support specialists) as well as researchers in the field of RP, our panel comprised 14 experts. This larger number of experts made consensus more difficult,

but yielded rich qualitative feedback and a variety of perspectives. Second, our tool was developed as part of the SWPRD training. Although the tool was derived from this specific training, it should generalize to classrooms whose teachers received other RP training. The teacher behaviors to be observed clearly resonated with school personnel and experts familiar with the core RP practices.

Implications for Research and Practice

To build the evidence-base supporting RP implementation, the field needs validated measures that can yield data of use to researchers as well as practitioners. Experts noted the challenge and potential utility of designing a tool to help teachers and observers alike hone in on behaviors that are central to a restorative classroom. For teachers, these behaviors aligned with communication and relationship building skills associated with RP training. While our training focused exclusively on teachers, experts pointed out the importance of student voice in promoting a restorative classroom. Student behavior that stems from and reinforces a restorative classroom is importantly broader. Rather than focusing on adherence to circle rules, experts advised us to consider the general types of interactions and communication that appear marked by qualities of inclusion, respect, authenticity, accountability, and even vulnerability. Rather than privileging student outcomes, we were cautioned to attend to processes, however messy, that might represent instances of student voice, agency, and even empowerment. Finally, reviewers' recommendations for revision seemed designed to make the fidelity monitoring process as restorative as possible through providing teachers opportunities to co-articulate with the observer what is restorative about circles and classrooms and what is needed to support teachers in their implementation of RP. Thus, teachers (and students) stand to gain the most from this tool as they work towards creating classrooms where authority is shared by teachers and students, students

feel comfortable speaking up, and students and teachers respect each other's differences and vulnerabilities. Following the recommendations of the reviewers, we will revise the measure to highlight student voice and ownership and measure process as well as outcomes. Additional psychometric testing will be conducted to create a measure that truly captures what a restorative classroom looks and sounds like.

References

Acosta, J., Chinman, M., Ebener, P., Malone, P. S., Phillips, A., and Wilks, A. (2019). Evaluation of a whole-school change intervention: Findings from a two-year cluster-randomized trial of the restorative practices intervention. *Journal of Youth and Adolescence*, Vol. 48 No. 5, pp. 876-890.

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., and Pianta, R. (2013). Observations of effective teacher–student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system—secondary. *School psychology review*, Vol. 42, No. 1, pp. 76-98.

Amstutz, L. S. (2015). *The little book of restorative discipline for schools: Teaching responsibility; creating caring climates*. Simon and Schuster, New York, NY.

Amuseghan, S. A. (2009). Language and communication in conflict resolution. *Journal of Law and Conflict Resolution,* Vol. 1, No.1, pp. 001-009.

Augustine, C.H., Engberg, J., Grimm, G.E., Lee, E., Wang, E.L., Christianson, K. and Joseph, A.A. (2018). *Can restorative practices improve school climate and curb suspensions? An evaluation of the impact of restorative practices in a mid-sized urban school district*. RAND Corporation, Santa Monica, CA.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage, Thousand Oaks, CA.

Corbin, J. M., and Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, Vol. 13, No. (1), pp. 3-21.

Davison, M., Penner, A. M., & Penner, E. K. (2022). Restorative for all? Racial disproportionality and school discipline under restorative justice. *American Educational Research Journal, 59*(4), 687-718. doi: 10.3102/00028312211062613.

Evanovich, L. L., Martinez, S., Kern, L., and Haynes Jr, R. D. (2020). Proactive circles: A practical guide to the implementation of a restorative practice. *Preventing School Failure: Alternative Education for Children and Youth,* Vol. 64, No. 1, pp. 28-36.

Fisher, C. (2020). *Assessing the Effects of Restorative Practices on Teacher Practices in Elementary Classrooms*. Graduate Theses, Dissertations, and Capstones. 90. https://scholarworks.bellarmine.edu/tdc/90Gregory, A., and Evans, K. R. (2020). *The Starts and Stumbles of Restorative Justice in Education: Where do we go from here?* University of Colorado, Boulder, National Education Policy Center. Boulder, CO.

Gregory, A., Gerewitz, J., Clawson, K., Davis, A., and Korth, J. (2014). *Excerpt from RP-Observe Manual*. Rutgers University, Piscataway, NJ.

Gregory, A., Huang, F. L., Anyon, Y., Greer, E., and Downing, B. (2018). An examination of restorative interventions and racial equity in out-of-school suspensions. *School Psychology Review*, Vol. 47, No. 2, pp. 167-182.

Guckenburg, S., Hurley, N., Persson, H., Fronius, T., and Petrosino, A. (2016). *Restorative justice in US Schools: Practitioners' perspectives*. West Ed Justice & Prevention Research Center, San Francisco, CA.

Huang, F. L., Gregory, A., & Ward-Seidel, A. R. (2023). The impact of restorative practices on

the use of out-of-school suspensions: Results from a cluster randomized controlled trial.

Prevention Science, 1-12.

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (Vol. 3).

Alexandria, VA: Association for Supervision and Curriculum Development.

Lewis, T. J., Scott, T. M., Wehby, J. H., and Wills, H. P. (2014). Direct observation of teacher

and student behavior in school settings: Trends, issues and future directions. *Behavioral

Disorders*, *Vol. 39,* No. 4, pp. 190-200.

Lodi, E., Perrella, L., Lepri, G. L., Scarpa, M. L., and Patrizi, P. (2021). Use of restorative justice

and restorative practices at school: A systematic literature review. *International Journal

of Environmental Research and Public Health*, Vol *19, No.* 1, Available at

https://doi.org/10.3390/ijerph19010096 (Accessed on 1 October 2022).

Oluwatayo, J. A. (2012). Validity and reliability issues in educational research. *Journal of

Educational and Social Research*, Vol. *2*, No. 2, pp. 391-391.

Ortega, L., Lyubansky, M., Nettles, S., and Espelage, D. L. (2016). Outcomes of a Restorative

Circles Program in a High School Setting. *Psychology of Violence*, Vol 6, No 3, pp 459-

468.

Rideout, G., Karen, R., Salinitri, G., & Marc, F. (2010). Measuring the impact of restorative

justice practices: Outcomes and contexts. *Journal of Educational Administration and

Foundations, 21*(2), 35.

Sanetti, L. M. H., and Collier-Meek, M. A. (2014). Increasing the rigor of procedural fidelity

assessment: An empirical comparison of direct observation and permanent product

review methods. *Journal of Behavioral Education,* Vol, 23, No 1, pp. 60-88.

Vincent, C., Inglish, J., Girvan, E., Sprague, J. & McCabe, T. (2016). Integrating School-wide

      Positive Behavior Interventions and Supports (SWPBIS) and Restorative Discipline

      (RD). In Skiba, R., Mediratta, K., & Rausch, M.K. (Eds.). *Inequality in school discipline:*

      Research and practice to reduce disparities. (pp. 115-134). New York: Palgrave

      MacMillan.

Vincent, C.G., Inglish, J., Girvan, E., Van Ryzin, M., Svanks, R., Springer, S., & Ivey, A.

      (2021). Introducing restorative practices into high schools' multi-tiered systems of

      support: Successes and challenges. *Contemporary Justice Review: Issues, in Criminal.*

      Social, and Restorative Justice. doi.org/10.1080/10282580.2021.1969522

Vincent, C.G., McClure, H., Marquez, B., & Goodrich, D. (2021). Designing professional

      development in restorative practices: Assessing high school personnel's, students', and

      parents' perceptions of discipline practices. *National Association of Secondary School*

      Principals Bulletin 105(4), 250-275. doi.org/10.1177/01926365211045461.

Yusoff, M. S. B. (2019). ABC of content validation and content validity index

      calculation. *Resource*, Vol, 11, No. 2, pp. 49-54.

Zehr, H. (2015). *The little book of restorative justice: Revised and updated*. New York: Simon

      and Schuster.

Table I. *Section 1, Structure of the Tool: I-CVIs and S-CVI/Ave*

Domain	Item	I-CVI	S-CVI/Ave
Structure of the Tool	Are the sections of the measure logically sequenced?	1.00	
	Are the tool's three parts necessary?	.86	
	Is the tool's overall lay out conducive to easy data collection?	.69	
	Are the identifiers of the observation session (teacher name, observer name, date, class period, grade level, number of students, subject taught) adequate?	.93	
			.87

(Table by Vincent, McClure, Svanks, Girvan, Inglish, Reiley & Smith, 2023)

Table II. *Section 2, Pre-observation Meeting: I-CVIs and S-CVI/Ave*

Domain	Item	I-CVI	S-CVI/Ave
Pre-observation meeting	Please review the Circle Planning Tool. Are the items assessing the teacher's use of the circle planning tool adequate?	.77	
	Are the items to observe the overall classroom environment adequate?	.79	
	Is the time allocated for the pre-observation meeting adequate?	.79	
			.78

(Table by Vincent, McClure, Svanks, Girvan, Inglish, Reiley & Smith, 2023)

Table III. *Section 3, Observation: Proactive Circle Teacher Behavior: I-CVIs and S-CVI/Ave*

		I-CVI			S-CVI/Ave
		Relevance	**Necessity**	**Usefulness**	
Observation: Proactive Circle Teacher Behavior	Provided clear instructions for students to set up circle.	.79	.77	.77	
	Clearly stated purpose of circle.	.93	.92	.92	
	Clearly stated circle guidelines.	1.00	.92	1.00	
	Adhered to circle guidelines.	1.00	1.00	1.00	
	Clearly introduced the talking piece.	.86	.92	.92	
	Clearly stated prompts.	.92	.92	1.00	
	Intervened restoratively when circle guidelines were violated.	1.00	1.00	1.00	
	. Provided a closing statement	.86	.85	85	
					Relevance: .92
					Necessity: .91
					Usefulness: .93
	Is the scoring scale for the		.75		

proactive circle teacher
behaviors appropriate to yield
actionable data?

(Table by Vincent, McClure, Svanks, Girvan, Inglish, Reiley & Smith, 2023)

Table IV. *Section 3, Observation: Proactive Circle Student Behavior: I-CVIs and S-CVI/Ave*

		I-CVI		S-CVI/Ave
	Is it appropriate to observe student behavior in the aggregate?	.77		
	Is tallying an appropriate metric to observe student behavior?	.79		
				.78
		Relevance	**Necessity**	**Usefulness**
Observation: Proactive Circle Student Behavior	Time it took to set up the circle.	.36	.46	.46
	Number of students who responded to the prompt (not to each other).	.79	.77	.85
	Number of students who spoke when they did NOT have the talking piece?	.64	.62	.69
	Number of students who passed when they had the talking piece?	.64	.62	.77
	Number of students who referenced own needs and/or impacts.	.71	.23	.77
	Number of students who	.71	.77	.77

referenced others' needs and/or impacts				
Number of students who references an identification with the class community.	.64	.69	.69	
Number of students who made behavioral corrections in response to teacher restating circle rules.	.86	.85	.85	
Time it took to transition to next classroom activity.	.43	.38	.62	
				Relevance: .64
				Necessity: .60
				Usefulness:.72

(Table by Vincent, McClure, Svanks, Girvan, Inglish, Reiley & Smith, 2023)

Table V. *Section 3, Observation: Remaining Class Period: I-CVIs and S-CVI/Ave*

		I-CVI	**S-CVI/Ave**
Observation: **Remaining**	Are the listed behaviors conducive to a restorative classroom?	.93	
Class Period: **Teacher** **Behaviors**	Is the scoring scale ("did not display," "displayed partially," " displayed thoroughly") appropriate to yield actionable data?	.93	
Observation: **Remaining** **Class Period:**	Are the listed behaviors conducive to a restorative classroom?	.93	
Student **Behaviors**	Is tallying the number of students exhibiting the behavior an appropriate metric to yield actionable data?	.71	
			.86

(Table by Vincent, McClure, Svanks, Girvan, Inglish, Reiley & Smith, 2023)

Table VI. *Section 4, Post-Observation Meeting: I-CVIs and S-CVI/Ave*

		I-CVI	S-CVI/Ave
Post-Observation Meeting	Are these questions relevant for a debrief meeting?	1.00	
	Is the time allotted to the post-observation meeting adequate?	.57	
			.79

(Table by Vincent, McClure, Svanks, Girvan, Inglish, Reiley & Smith, 2023)

Table VII. *Overall Assessment: I-CVIs and S-CVI/Ave*

		I-CVI	S-CVI/Ave
Overall Assessment	How likely are classroom teachers to participate in direct observation guided by this tool?	.79	
			.79

(Table by Vincent, McClure, Svanks, Girvan, Inglish, Reiley & Smith, 2023)