



The Disaggregated Effects of Visual Performance Feedback on Teachers' Use of Behavior Specific-Praise and Reprimands

Ashley Rila¹ · Allison L. Bruhn¹ · Lanqi Wang¹

Accepted: 1 June 2022 / Published online: 28 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Behavior-specific praise (BSP) is one of the simplest classroom management strategies to implement and considered an evidence-based practice. Unfortunately, teachers underuse BSP and deliver more reprimands to students in their classrooms. Secondary students receive the highest rates of reprimands and exclusionary discipline (i.e., office discipline referral [ODR], suspension, expulsion) with students of color receiving disproportionate rates compared to their White peers. Performance feedback is a commonly used strategy to change teacher practices however, little is known about the impact of performance feedback on the equitable delivery of BSP and reprimands to students by race and sex. The purpose of this multiple baseline design study was to examine the effects of a visual performance feedback (VPF) intervention with secondary teachers on their equitable delivery of BSP and reprimands and the collateral impacts on student outcomes. In the first phase of intervention, teachers received VPF on their total BSP and reprimands. In the second phase, teachers received disaggregated VPF on their rates of BSP and reprimands delivered to students by race and sex. Results indicate a functional relation between VPF and total BSP and an overall reduction in total reprimands. Mixed results were found between VPF and the equitable delivery of BSP and reprimands rates delivered to students by race and sex. Student outcomes indicated an increase in average class-wide academic engagement and no impact on ODRs as no teacher delivered a single ODR. Key findings, limitations, and future research are discussed.

Keywords Behavior-specific praise · Reprimands · Secondary · Visual performance feedback · Equity · Single case design

✉ Ashley Rila
Ashley-rila@uiowa.edu

Extended author information available on the last page of the article

Introduction

Exclusionary discipline practices occur when students are sent out of the classroom due to unwanted behavior. When students are removed from the classroom, they lack access to critical instruction. This lack of instructional access has been linked to reduced achievement, grade retention, school avoidance, increased disruptive behavior, drop out, drug abuse, and involvement in juvenile justice systems (Fabelo, et al., 2011; Fredricks et al., 2004). Research has indicated exclusionary discipline is disproportionately experienced by students of color in comparison to White students in K-12 schools (Gage et al., 2020; Krezmien et al., 2006; Lloyd et al., 2019; Losen et al., 2015). In secondary grades (6–12), discipline rates increase compared to elementary grades where students of color receive substantially more office discipline referrals (ODR) or suspensions compared to White students (Gion et al., 2018; Losen et al., 2015). Though not as extreme, other students of color (e.g., First Nations, Hispanic) also experience greater disciplinary actions than White students (Gage et al., 2020; Krezmien et al., 2006), thus indicating punitive discipline practices may be applied disproportionately by race (Lloyd et al., 2019). Exacerbating this issue, Black students receive harsher punishments than White students for similar violations (Anderson & Ritter, 2020; Skiba et al., 2011) and for more subjective behaviors (e.g., defiance, disrespect; Girvan et al., 2016). Some research indicates Black females receive the highest rates of discipline compared to all students (Blake et al., 2011; Wun, 2016) and female First Nations (U.S. Department of Education, 2014) and female Hispanics (Lehmann et al., 2021) are at higher risk than White males of receiving a disciplinary infraction.

Numerous factors may lead to these disproportionate practices. First, many teachers struggle to effectively manage their classrooms (Freeman et al., 2014), report being unprepared to handle student misbehavior (Chesley & Jordan, 2012), and have higher stress and burnout rates than teachers who can effectively control a classroom (Smith & Smith, 2006). This may be the result of pre-service teachers receiving little or no classroom management training (Freeman et al., 2014). Professional development may be used to improve classroom practices. However, typical professional development is generally ineffective in changing practices when it does not include hands-on practice, coaching, or performance feedback (Chappuis et al., 2009). Teachers are more likely to benefit from training that includes opportunities to apply their learning within a classroom context, set goals, and reflect on practices (Bruce et al., 2010).

Though inadequate training in classroom management may contribute to teachers resorting to ineffective practices like delivering high rates of reprimands, implicit biases—that is, peoples' unconscious thoughts or attitudes toward others—also may play a role (Girvan et al., 2016; Riddle & Sinclair, 2019). Implicit bias occurs when a decision is made automatically or without conscious thought (Greenwald & Banaji, 1995), such as a snap decision, and may be incompatible with a persons' actual beliefs (Greenwald & Krieger, 2006). Implicit bias can take many forms and can be applied across multiple domains such as race, sex, sexual

orientation, age, and disability. Implicit biases may occur in a single domain (e.g., race alone), or they may intersect (e.g., race and sex). Implicit biases have the potential to impact classroom decisions that result in disproportionate outcomes.

Some evidence suggests that biased decision-making is more prominent with subjective student behaviors (e.g., disrespect, defiance) that require a teacher to make a judgment call on a behavioral violation (Skiba et al., 2002). When individuals are mentally or physically exhausted, implicit biases are more likely to impact a person's decisions (Kouchaki & Smith, 2014). As teachers are increasingly required to do more with less while ensuring student success (Madigan & Kim, 2021), teachers' mental and physical exhaustion is prone to increase. Consequently, it becomes more probable that teachers will make biased decisions under these conditions. Holding individuals accountable for their biases has been ineffective in reducing disproportionate practices (Girvan et al., 2015). Rather, emerging evidence suggests that providing individuals with data and guidance in decision-making may result in equitable practices (Girvan et al., 2016). Providing teachers with data on their practices (e.g., praise), may be one way to help close the discipline gap and reduce the number of students who are unfairly removed from classrooms.

In 2014, the federal government outlined recommendations for reducing disproportionate discipline such as, using evidence-based practices to create a positive and safe school environment, creating clear and consistent expectations and consequences, and using positive interventions instead of exclusionary discipline (U.S. Department of Justice, 2014). Several evidence-based classroom management practices exist that could address these disparities and also align with the federal recommendations. They include (a) maximize classroom structure; (b) teach and reinforce positive expectations; (c) engage students in observable ways; (d) use strategies to acknowledge appropriate behaviors; and (e) use strategies to correct inappropriate behaviors (Simonsen et al., 2010). When these evidence-based strategies are implemented with fidelity, disruptive behavior decreases (Oliver et al., 2011), academic engagement increases (Sutherland et al., 2000), and teacher well-being improves (Ross et al., 2012).

One effective way for teachers to acknowledge appropriate student behavior is delivering student behavior-specific praise (BSP). BSP can be described as providing a student with positive acknowledgment about a specific behavior the student displayed (e.g., "Nice job applying your strategies to solve that problem;" Allday et al., 2012). BSP has been found to improve academic and social behaviors across grade levels (Downs et al., 2019; Sutherland et al., 2000), is considered an evidence-based practice by the Council of Exceptional Children (Royer et al., 2019), and described as one of the simplest strategies to implement (Gable et al., 2009). Some scholars recommend BSP should be delivered a minimum of six times every 15 min (Sutherland et al., 2000). Although, research on BSP is lacking at the secondary level, scholars agree that BSP should exceed reprimands (Spilt et al., 2016). Data suggest, however, that teachers underuse BSP (Gage et al., 2018) and secondary students receive twice as many reprimands than praise (Hirn & Scott, 2014).

To this end, performance feedback is a frequently used strategy to change teacher behaviors that involves an outside observer monitoring specific teacher behaviors and providing them with feedback (Noell et al., 2005). The feedback

can be delayed (after an observation) or immediate (during an observation) and may be delivered orally (e.g., in-person, video conference, bug-in-ear technology), through writing (e.g., note, email), or through visual displays (e.g., graphs of data; Author, in review). Performance feedback has been used effectively in a variety of educational settings from pre-k through high school (Solomon et al., 2012); general and special education (Fallon et al., 2015); and for numerous behaviors such as BSP (Allday et al., 2012), opportunities to respond (Simonsen et al., 2010), treatment integrity (Solomon et al., 2012), and reprimands (Pisacreta et al., 2011).

A recent literature review in secondary settings found that performance feedback has been implemented as a stand-alone intervention or in conjunction with additional components such as self-monitoring, goal setting, visual prompting, or modeling (Author, in review). Although performance feedback is touted as an evidence-based practice in some literature reviews (e.g., Cavanaugh, 2013; Fallon et al., 2015), these reviews included a range of grade levels with most studies located in primary settings and relatively little research in secondary settings. Thus, more rigorous research examining stand-alone performance feedback interventions (e.g., VPF alone) is needed to establish performance feedback as an evidence-based practice for increasing BSP and decreasing reprimands in secondary settings (Author, in review). Furthermore, the review revealed that none of the included studies provided the teachers with data on their rates of praise or reprimands disaggregated by student race or sex. Because exclusionary discipline is higher in secondary settings (Hirn & Scott, 2014) and race and sex may play a role in the interaction between students and teachers, research examining the effects of performance feedback targeting equitable treatment of students in classrooms is needed.

Surprisingly, little research has been conducted to examine which students are receiving praise and reprimands in classroom settings where (a) disciplinary actions are likely to occur and (b) students spend the majority of their day. Further, given the plausible link between teacher reprimands and exclusionary discipline (e.g., ODRs), additional research is warranted. With students of color being more at risk for exclusionary punishments, it is logical to hypothesize that students of color receive more reprimands and less praise than their peers, as documented in a recent study (Gion et al., 2020). Specifically, Gion et al. (2020) implemented a multicomponent classroom intervention with teachers in a K-8 school who delivered higher rates of reprimands to African American students in their classrooms compared to all other students. Teachers examined their classroom expectations from a culturally responsive lens, surveyed their students on how they preferred to be praised, then researchers provided teachers with coaching and visual performance feedback (VPF) on their praise-to-reprimand ratio. VPF was provided to the teachers via email that included a graph of their praise-to-reprimand ratio delivered to African American students and to all other students. Results indicated teachers increased their rates of praise-to-reprimand ratio for their African American students. As this was one of the first studies targeting disproportionate treatment of students in classrooms, further exploration of the (a) effects of VPF on BSP and reprimand delivery by race, (b) collateral effects on student behaviors, and (c) distal effects on exclusionary discipline is necessary.

In light of evidence indicating disproportionate discipline practices, the need for effective strategies to increase teachers' use of BSP and reduce reprimands, research suggesting that providing data-based performance feedback can change teacher behavior, and the lack of rigorous research on performance feedback at the secondary level, the purpose of this study was to examine the effects of VPF on secondary teachers' use of praise and reprimands and the associated rates by race and sex. Research questions included: (a) What are the effects of VPF on BSP and reprimands delivered by secondary teachers?; (b) How do rates of BSP and reprimands differ by race and sex of students?; (c) Are there collateral effects of changes in teachers' rates of BSP and reprimands on students' ODRs and academic engagement?; and (d) To what extent did teachers find VPF to be socially valid?

Methods

Participants and Setting

We conducted this study in a comprehensive Midwestern high school serving approximately 1600 students grades 9 through 12. The student body was 57% White, 18% Black, 17% Hispanic, 4% multi-racial, and 3% Asian; 40% qualified for free and reduced lunch. Following institutional review board (IRB) approval and permission from the principal, the first author emailed the building teachers ($n=50$). Nine teachers expressed interest in the study and six consented to participate. Teachers were eligible if they were a licensed teacher, taught academic or skills instruction (e.g., English, math, special education), and delivered low rates of praise or high rates of reprimands. Specifically, teachers were eligible if they delivered an average of four or fewer praise statements or their reprimand average was equal or greater than their praise statement average (Myers et al., 2011). We conducted three 15-min screening observations to tally the frequency of praise and reprimands the teacher delivered to all students. We averaged their frequency for each statement over the three observations to calculate their average rates. Two teachers were dropped from the study due to high rates of praise. Thus, four teachers participated, and they were compensated \$100 for completing the study.

Once a teacher was deemed eligible, we screened the class for eligibility. Eligible classes had to include (a) students of color and White students with each group comprising no less than 25% and no more than 75% of the total population (Gion et al., 2020), and (b) 5 or more students. These criteria were used to ensure the class had a diverse population with more than one student being identified as a student of color or White such that the teacher had multiple students to deliver statements to as opposed to just one student for each demographic group. Following consent, each teacher selected a class that met these criteria. During the consent meeting, teachers counted the number of students on their class roster that they identified as a student of color or White. During each screening observation, we tracked the total number of students in attendance and their associated race to confirm the class was eligible. Consistent with a similar study, we opted for visual identification and teacher report on aggregate student demographics (Gion et al., 2020), as we did not have IRB or

administration permission, to collect individual student demographics. The same procedures were used to identify student sex. However, student sex percentages were not used to determine class eligibility.

We conducted the study in four general education classes. Mr. Brown taught Science to grades 9 and 10. His class averaged 23 students (range=14–30) and was 52% students of color, 48% White students, 30% female, and 70% male. Ms. Ball also taught Science to grades 9 and 10 and averaged 23 students (range=20–26) and was 57% students of color, 43% White students, 61% female, and 39% male. Mr. Cox taught Advanced Placement Psychology to grades 11 and 12 and averaged 24 students (range=14–28). His class was 46% students of color, 54% White students, 71% female, and 29% male. Finally, Mrs. Cobb taught English to 9th grade and averaged 29 students (range=26–32). Her class was 27% students of color, 73% White students, and 50% for both female and male students.

Measures

The primary dependent variables were BSP and reprimands and were used to answer Research Questions 1 and 2 relative to the proximal effects of intervention.

Behavior-Specific Praise

BSP was defined as any positive approval statement by the teacher that included a precise behavior (academic or social) *and* was delivered specifically to a student or group of students (Knochel et al., 2022). Examples included statements like “Good job using your strategies to solve that problem,” “Class, thank you for working quietly,” or “Sam, I like how you invited Bob into your group.” In the instance the teacher delivered a praise statement that included the precise behavior but lacked a student name, the statement was recorded as BSP only if the teacher was looking directly at the student or in response to a student participating in class and it was clear to the observers that the statement was delivered to a specific student. BSP was recorded when the teacher concluded the statement. If BSP was a long, detailed statement, it was only recorded as one instance unless at least 3 s passed between one statement and the beginning of the next or the content of the BSP changed. In which case, a second BSP was recorded. BSP non-examples included any non-specific praise statement, gesture indicating approval or disapproval, or a reprimand. For example, statements such as “Mary, good job,” “That’s right,” “Stop that,” “Devon, you’re so smart,” or giving a thumbs up or down, high five, or a head shake.

Reprimands

Reprimands were any oral statement or gesture aimed to redirect or correct an individual student’s or group of student’s social behavior (Gion et al., 2020). Examples included statements like “Stop that,” “Joe, eyes on your own paper,” or a head shake indicating disapproval. A reprimand was recorded when the teacher finished the statement or gesture. If a reprimand was long, it was only recorded as one instance

unless at least 3 s passed between one statement and the beginning of the next or the content changed. In which case, a second reprimand was recorded. Non-examples included a statement or gesture indicating approval. For example, stating “Good job,” “Excellent work turning in your project,” a high five, or a head nod indicating approval. If the teacher corrected an academic error (e.g., “No, that’s incorrect”) the statement was not counted.

BSP and reprimands were collected using the Multi-Option Observation System for Experimental Studies (MOOSES) software which allows for time sequenced data to be collected for multiple variables simultaneously (Tapp et al., 1995) that ran on a Microsoft Surface Pro tablet. When a teacher statement was delivered, we first determined the statement type (BSP, reprimand). Second, we determined if the statement was delivered to an individual or a group of students. Last, we determined to whom the statement was delivered by student demographics of race (student of color, White) and sex (female, male) to record the appropriate category. If a statement was delivered to a group of students that included both students of color and White, we recorded the group as mixed race. Similarly, if the statement was delivered to a group that included both females and males, we recorded the group as mixed sex.

To answer Research Question 1, we used total frequency counts as every observation during the study was 15 min in length (Cooper et al., 2020). To answer Research Question 2, we converted statement frequencies into averages per student group to allow rates to be comparable as the groups were not equal in size. We totaled the frequency counts for each statement type for each student demographic and calculated an average by dividing this number by the number of students present in the observation in specific demographic group (Gion et al., 2020).

The secondary dependent variables were class-wide academic engagement and ODRs. These were used to answer Research Question 3 about the collateral effects of intervention.

Academic Engagement

Academic engagement was defined as a student appropriately attending to the assigned or approved activity (Myers et al., 2011). A student was considered academically engaged if they answered a teacher question directed at them or volunteered on-topic information to the lesson, made appropriate motor responses (i.e., writing, talking, following rules of activity, using materials appropriately) for the assigned activity, read silently with signs of scanning or page turning, engaged in on-topic conversations when approved with peers, quietly listened to the teacher or students talking to the class or group, looked at or attended to the assigned task, or waited appropriately for the teacher to begin or continue instruction. Students were considered not academically engaged if they shared off-topic information, engaged in behavior that disrupted their own or their peers’ engagement, slept, worked on other classwork, were on their cell phone when not approved, or were on a website on their computer not related to class.

We measured class-wide academic engagement via MOOSES using 15-s momentary time sampling. Prior to each observation, we divided the class into four

approximately equal groups. Every 15 s, MOOSES prompted observers, via an auditory cue in a single headphone, to assess if a group was academically engaged. If every student in that group was engaged, we recorded the group as academically engaged and thus, the group was counted as engaged for the entire 15-s interval. If one or more students within that group was not engaged, we recorded the entire group as not engaged and thus, the group was counted as not engaged for the entire 15-s interval. We assessed each group using round robin procedures. We assessed group 1 after the first 15-s interval, then assessed group 2 after the next 15-s interval and continued until all four groups were assessed. Then, we went back to group 1 and continued to assess each group in the same order throughout the observation. Momentary time sampling provides an estimate of behavior and we opted for this form of measure as it allowed observers to focus their attention on measuring the primary dependent variables (Cooper et al., 2020). We calculated class-wide academic engagement by summing the number of seconds each group was recorded as engaged divided by 900 s (15 min \times 60 s) and multiplying by 100 for each observation. We report average class-wide academic engagement by totaling the average engagement across observations for each phase divided by the number of observations per phase for each teacher.

Office Discipline Referrals

An ODR was a documented incident in which the teacher referred a student to the office for inappropriate behavior or violating a school rule (Garvin et al., 2016). The participating school collected and analyzed ODR data regularly, and thus, the data were readily available. Generally, ODRs capture low-frequency, high-intensity behaviors and may be used as a distal outcome measure (McIntosh et al., 2017) and are a moderately valid (disruptive behavior $r=0.38$) and reliable ($r=0.57$) measure (Pas et al., 2011).

Each teacher provided the researcher with the total number of ODRs they delivered in their target class pre-intervention and during intervention (only for the time that they were in an intervention phase). We analyzed these data through a pre-post comparison by converting ODRs into a daily rate. That is, we divided the total number of ODRs delivered pre-intervention by the total number of days school was in session prior to the teacher entering intervention. Then, we compared that rate to the daily rate of ODRs while the teacher was in intervention and divided the total number of ODRs by the number of days the teacher was in intervention.

Interobserver Agreement (IOA) The first and third authors, doctoral students studying special education, conducted observations. The first author was a White female who served as the primary observer and researcher. The third author was an Asian female who served as the second observer. The first author trained the third author in using MOOSES. First, the third author studied the codebook of operational definitions and asked clarifying questions. Second, a researcher-generated quiz on definitions was given until 100% accuracy was achieved. Third, classroom videos were watched until 90% reliability was reached on all dependent variables. Lastly, 15-min

direct observations were conducted in non-participating secondary classes, until 90% reliability was reached on dependent variables for three consecutive observations.

We collected IOA independently and simultaneously on an average of 28% (range=20–33%) of observations across each phase and participant. We calculated IOA via MOOSES by comparing our electronic files. Within a 5-s window around each dependent measure selected, one agreement was scored if a match was found between the two files (i.e., each observer selected the same measure within the 5-s window). All unmatched selections were scored as disagreements (Tapp et al., 1995). MOOSES calculated IOA by dividing the number of agreements by the total number of agreements plus disagreements then multiplied by 100. IOA for Ms. Ball averaged 97% (range=90–100%) for BSP, 100% for reprimands, and 92% (range=85–98%) for academic engagement. For Mr. Brown, IOA averaged 93% (range=86–100%) for BSP, 100% for reprimands, and 94% (range=92–97%) for academic engagement. For Mr. Cox, IOA averaged 89% (range=67–100%) for BSP, 100% for reprimands, and 94% (range=92–97%) for academic engagement. For Mrs. Cobb, IOA averaged 100% for BSP, 100% for reprimands, and 92% (range=86–94%) for academic engagement.

Descriptive Measures

Fidelity

The third author assessed procedural fidelity using a checklist to score if intervention components were present in emails (i.e., email and observation date match, sent before 4 p.m., greeting, BSP to teacher, graph, statement offering to answer questions, prompt for teacher reply). We used an online random number generator to determine which emails were assessed. We calculated fidelity by dividing the total present components by the total present plus absent components and multiplied by 100. A total of 25% of Mr. Brown's, 29% of Ms. Ball's, 25% of Mr. Cox's, and 20% of Mrs. Cobb's emails were checked and fidelity averaged 97% (range=86–100%). The only fidelity errors were related to timeliness. Two emails were sent later than 4:00 p.m., 4:25 p.m. and 6:05 p.m., respectively, the same day as the observation.

Social Validity

Teachers completed a modified version of the Intervention Rating Profile-15 (IRP-15; Witt & Elliott, 1985) by indicating their level of agreement to statements using a 6-point Likert-type scale. We averaged the ratings across all items and teachers to calculate social validity; higher scores indicated higher acceptability of intervention. The IRP-15 has evidence of reliability (Cronbach's $\alpha=0.98$) and validity ($r=-0.86$) for assessing the social validity of behavioral interventions (Martens et al., 1985). Teachers also completed a researcher-created, open-ended, post-intervention questionnaire to assess the acceptability of the VPF intervention.

Experimental Design and Procedures

We used a concurrent multiple baseline across teachers design, wherein, several staggered ABC designs were used to detect experimental effects within and between participants. We used visual analysis to make phase change decisions based on BSP and to determine experimental effects for dependent measures (Kratochwill et al., 2013). All observations for all conditions were 15 min and occurred when the teacher delivered instruction to the whole class. Teacher-directed instruction was selected to maximize the opportunities to observe teacher interactions with students. Data were not collected if the teacher administered an assessment, showed a film, or had a guest speaker that limited their interactions with students. For all observations, we tracked the number of students in attendance and their race and sex. If the population fell outside of the racial inclusion criteria listed in the participants and setting section above (e.g., students were absent, thus inclusion ratios were not maintained), data were not collected. This rarely occurred.

Baseline

All teachers entered baseline simultaneously and they continued their normal teaching procedures. The only change was outside observers entered the class to conduct direct observations. Two observers were only present during IOA observations, all other observations only had one observer. Teachers did not receive VPF during this phase.

Training

Prior to starting the study and unbeknownst to the teachers, we used an online random number generator to determine the tier at which each teacher would enter intervention. The first teacher began training following a minimum of five days in baseline and total BSP indicated a need for intervention as determined through visual analysis of the level, trend, and stability of BSP. That is, once there was a low and stable pattern or a flat or declining trend in total BSP, the first teacher received training while the remaining teachers stayed in baseline. The next teacher received training following a minimum of three data points indicating a positive change in BSP for the first teacher in intervention and analysis of the second teacher's baseline BSP indicating a need for intervention. This process continued until all teachers entered intervention.

During training, the first author met one-on-one with the teacher at a convenient time and in a private location within the school as identified by the teacher. Training sessions lasted 30 min and teachers were provided with a (a) definition of BSP and reprimands, (b) handout with BSP examples, (c) rationale for increased BSP and decreased reprimands, (d) worksheet to develop BSP statements, (e) line graph of their BSP and reprimand baseline data, (f) explanation of the intervention, and (g) opportunities to ask questions. The graphed data included frequencies of BSP and reprimands, the data were not disaggregated by race or sex in this phase. The first author explained how to interpret the graph and checked for teacher understanding.

After discussing their data, the teacher completed the BSP worksheet to develop their own examples of BSP that they could see themselves delivering to specific students. Then, the dyad discussed each example and the researcher provided feedback.

VPF

Following training, teachers immediately entered intervention. After every observation, the researcher sent teachers an email that included a (a) greeting, (b) BSP statement about their data (e.g., “Nice job today increasing BSP.”), (c) line graph of their BSP and reprimand frequencies, (d) statement offering to address questions (Rathel et al., 2014), and (e) prompt to reply they reviewed their graph. If teachers did not reply, the teacher was emailed again to remind them to review their data. No teacher needed this second email. Teachers were only required to review their graphs and data over time; they were not required to indicate in their reply any details of their review of the data. The graph included both BSP and reprimand frequencies observed for each phase so the teacher could monitor their progress overtime.

VPF+ Disaggregation Training

Following a minimum of five VPF sessions, if needed, teachers received a second training and intervention phase (VPF+DIS). Teachers only entered VPF+DIS if there were inequitable or unstable data patterns in BSP by race *or* sex as determined through visual analysis. Specifically, teachers entered VPF+DIS if (a) there was a clear separation in data paths between rates of BSP delivered to student of color versus White *or* females versus males (e.g., females received higher rates of BSP compared to males *or* White students received higher rates of BSP compared to students of color), (b) BSP was delivered at nonequivalent, variable rates between student race *or* sex (i.e., large gaps between BSP data points for White vs. students of color *or* for males vs. females), or (c) there were opposite trends in BSP by student demographic (e.g., sharp increasing trend for White vs. sharp decreasing trend for students of color). Given that the rates of BSP were dependent upon their total BSP, the variability in the separate data paths were not used to determine if the teacher required VPF+DIS. Rather, we examined the variability between data points to analyze for disparities. If the teacher’s data points for BSP rates for race *or* sex were consistently close or established a pattern of closeness, we determined the rates of BSP to be relatively equivalent and therefore not requiring VPF+DIS. If there was variability in the closeness of the BSP data points (i.e., large gaps between BSP data points for 3 sessions and small gaps for 2 sessions by student race *or* sex), then we determined BSP rates to be disproportionate and thus requiring VPF+DIS.

During the second training, we followed nearly identical training procedures. In this session, however, teachers were provided with (a) line graphs with their BSP and reprimand rates disaggregated by sex and race; (b) instruction and a check for understanding on reading the disaggregated graphs; (c) a rationale for increased BSP and decreased reprimands for all students, (d) examples of BSP and reprimands; (e) a worksheet to develop BSP focused on students not receiving BSP; and (f) an opportunity to ask questions.

Fig. 1 Total behavior-specific praise and reprimands delivered by teacher and average class academic engagement across baseline, visual performance feedback (VPF), disaggregated visual performance feedback (VPF+DIS), and maintenance (Maint.) phases. * Is trimester change, wherein class rosters changed slightly but all classes remained within inclusion criteria

VPF + Disaggregation (VPF + DIS)

Following VPF+DIS training, teachers immediately entered VPF+DIS. Identical VPF procedures were used except the email included the teacher's disaggregated data. Each email included a line graph of their total BSP and reprimands plus additional graphs for BSP and reprimands disaggregated by race and sex. Each statement had two separate graphs, one for race (student of color vs. White) and one for sex (female vs. male). At a minimum, teachers received three graphs, one with their total BSP and reprimands and two with their disaggregated BSP. If teachers also had a detectable separation in reprimands or if they wanted to receive all their data, then all five graphs were emailed daily. Mr. Cox and Mrs. Cobb received five graphs daily.

Maintenance

Following a minimum of five sessions with the last teacher in intervention, teachers entered maintenance during which they spent 1 week with no intervention. After 1 week, teachers returned to baseline procedures for maintenance data collection.

Results

Due to the unprecedented COVID-19 pandemic, this study was terminated prematurely. Also, the school closed for four days due to inclement weather and all teachers missed one to three days for various reasons. Consequently, Ms. Ball was the only teacher who completed the maintenance phase and Mrs. Cobb had a truncated VPF+DIS phase by at least one day.

Class-Wide Dependent Measures

Figure 1 displays the total frequency of BSP (closed data path, scaled to left y-axis) for teachers across phases. In baseline, all teachers had low and stable rates of BSP. After entering VPF, all teachers demonstrated large increases in their BSP that were relatively stable, establishing a functional relation between baseline and VPF. Rates of BSP demonstrated in VPF maintained at similar levels when teachers entered the subsequent phase, VPF+DIS (Mr. Brown, Mr. Cox, Mrs. Cobb) or maintenance (Ms. Ball).

The total frequency of reprimands (open data path, scaled to left y-axis) is also displayed in Fig. 1. During baseline, all teachers consistently delivered more

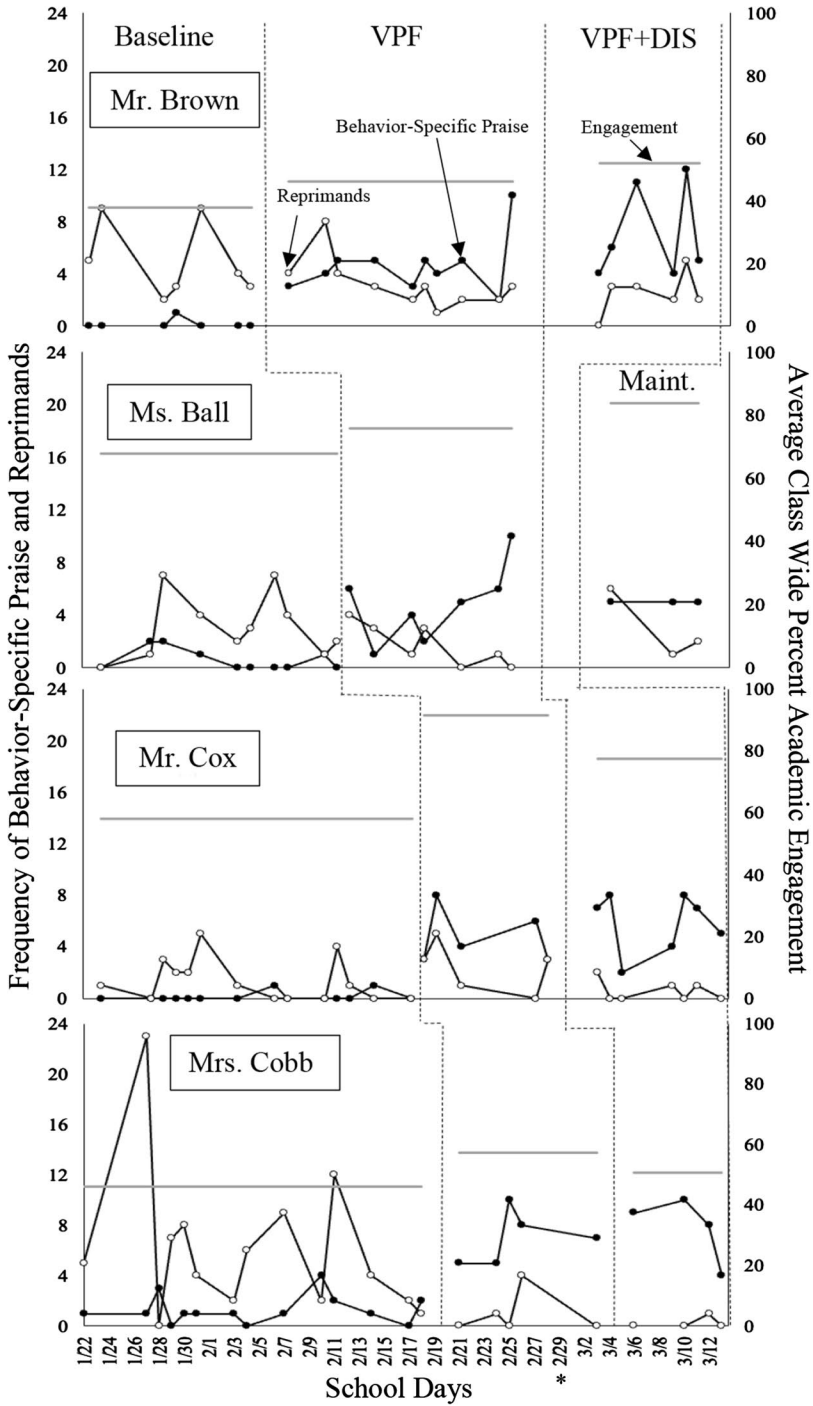


Fig. 2 Mr. Brown's disaggregated data across phases. BSP is behavior-specific praise; SOC is student of color; VPF is visual performance feedback; VPD+DIS is disaggregated visual performance feedback. Y-axis scales shift across graphs

reprimands than BSP. Although all teachers had a decreasing trend toward the end of baseline, overall, their rates of reprimands exceeded or were equal to their rates of BSP. During VPF, three teachers had a decreasing trend (Mr. Brown, Ms. Ball, Mr. Cox). With the exception of one session (2/26) for Mrs. Cobb, all teachers had a reduction in their reprimand level in VPF compared to baseline. In the final phase, VPF+DIS or maintenance depending upon the teacher, similar patterns of reprimands continued from VPF. All reprimand data in VPF, VPF+DIS, and maintenance phases overlapped with baseline data for all teachers.

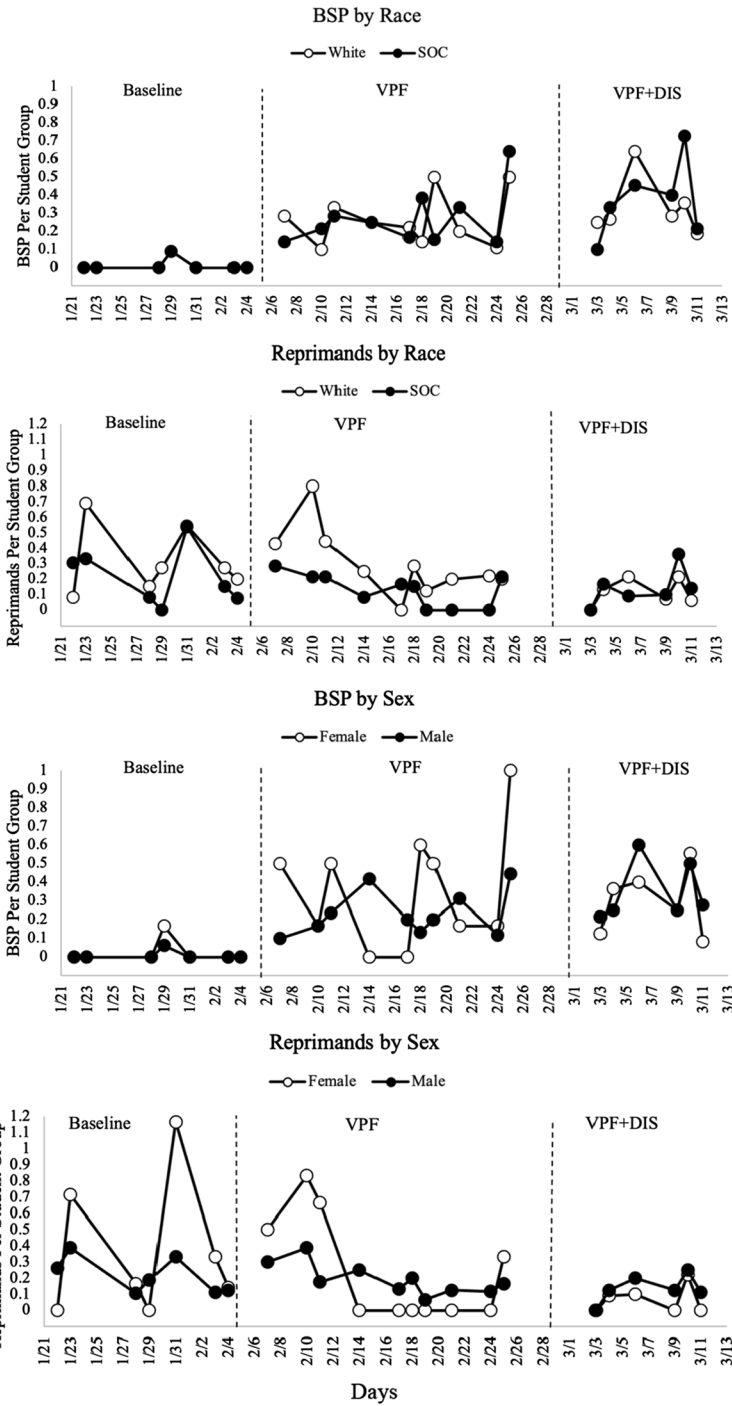
Lastly, average class-wide academic engagement (solid gray line, scaled to right y-axis) is depicted in Fig. 1 for each teacher across phases. During baseline, three teachers had low average rates of class engagement (Mr. Brown=38%, Ms. Ball=65%, Mr. Cox=58%, Mrs. Cobb=46%). After entering VPF, average engagement increased for every class by varying levels, with a large increase for Mr. Cox (Mr. Brown=46%, Ms. Ball=76%, Mr. Cox=92%, Mrs. Cobb=57%). During VPF+DIS, Mr. Brown's class was the only class to increase engagement and Mrs. Cobb's class engagement decreased to near baseline levels (Mr. Brown=52%, Mr. Cox=78%, Mrs. Cobb=51%). During maintenance, average engagement in Ms. Ball's class increased to 84%.

Disaggregated Dependent Measures

Mr. Brown

Figure 2 depicts Mr. Brown's disaggregated data by student group. The first graph shows BSP rates by race (White students, open data path; Students of color, closed data path). During baseline, he delivered nearly zero BSP. During VPF, BSP rates were relatively stable and similar (i.e., data points consistently close). This pattern remained during VPF+DIS. In the second graph, reprimands by race, Mr. Brown delivered relatively comparable rates, with the exception of two days where he delivered higher rates of reprimand to White students. During VPF, initially, there were higher rates of reprimands for White students but, the gaps between the two groups merged; White students consistently received more reprimands during VPF. During VPF+DIS, rates became more aligned.

The third graph shows BSP by sex (female students, closed data path; male students, open data path). During baseline, BSP was comparable by race as Mr. Brown delivered nearly zero BSP. During VPF, he delivered variable rates of BSP to females, hence the reason he required VPF+DIS. Overall, during VPF+DIS, BSP rates by sex became more similar as data points and trends were closer together compared to VPF. The fourth graph depicts reprimands by sex. During baseline, females received unstable and variably high rates of reprimands that continued initially into



VPF but stabilized in parity with males as the phase progressed. Relatively comparable rates of reprimands occurred in VPF + DIS.

Ms. Ball

Figure 3 depicts disaggregated data for Ms. Ball. The first graph displays BSP by race (White students, open data path; Students of color, closed data path). During baseline, she delivered low, comparable rates of BSP by race. Overall, during VPF, she established relatively equivalent rates. Initially, there was a discrepancy where White students received more BSP than students of color. However, after one session, the data merged and remained relatively close or equal to one another. Given this merged and stable data for BSP by race (and sex, see below), Ms. Ball did not require VPF + DIS. During maintenance, students of color received slightly higher rates of BSP. Graph two displays reprimands by race. With the exception of one session during baseline, 1/31, she delivered comparable rates of reprimands by race across phases.

Graph three shows Ms. Ball's rates of BSP by sex (female students, closed data path; male students, open data path). During baseline, her BSP for both sexes were near zero. During VPF, she delivered relatively similar rates as the gaps between data points were consistently close together. During maintenance, inequities emerged as males received more BSP. Graph four shows rates of reprimands by sex. She delivered relatively equivalent rates by sex. Again, during maintenance, she delivered higher rates to males.

Mr. Cox

Figure 4 depicts Mr. Cox's disaggregated data. Graph one displays BSP by race (White students, open data path; Students of color, closed data path). During baseline, Mr. Cox delivered nearly zero BSP. During VPF, three sessions were nearly equal and trends by race were opposite and one reason why Mr. Cox required VPF + DIS. With the exception of the first and last day, he delivered relatively similar rates of BSP by race during VPF + DIS. Graph two shows reprimands by race. In all phases, Mr. Cox displayed a pattern of comparable reprimand delivery by race.

Graph 3 displays BSP by sex (female students, closed data path; male students, open data path). Similar to other teachers, BSP was near zero and equivalent. However, during VPF, females and males had sharp, opposite trends and thus the second reason he required VPF + DIS. In VPF + DIS, he delivered more comparable rates of BSP by sex. Graph 4 depicts reprimands by sex. During baseline, there were patterns of disparity where males received more reprimands. In VPF, there continued to be obvious discrepancies initially but was eliminated in the last two days. Then during VPF + DIS, Mr. Cox delivered equivalent rates of reprimand by sex.

Mrs. Cobb

Figure 5 displays Mrs. Cobb's disaggregated data. Graph one depicts BSP by race (White students, open data path; Students of color, closed data path). During

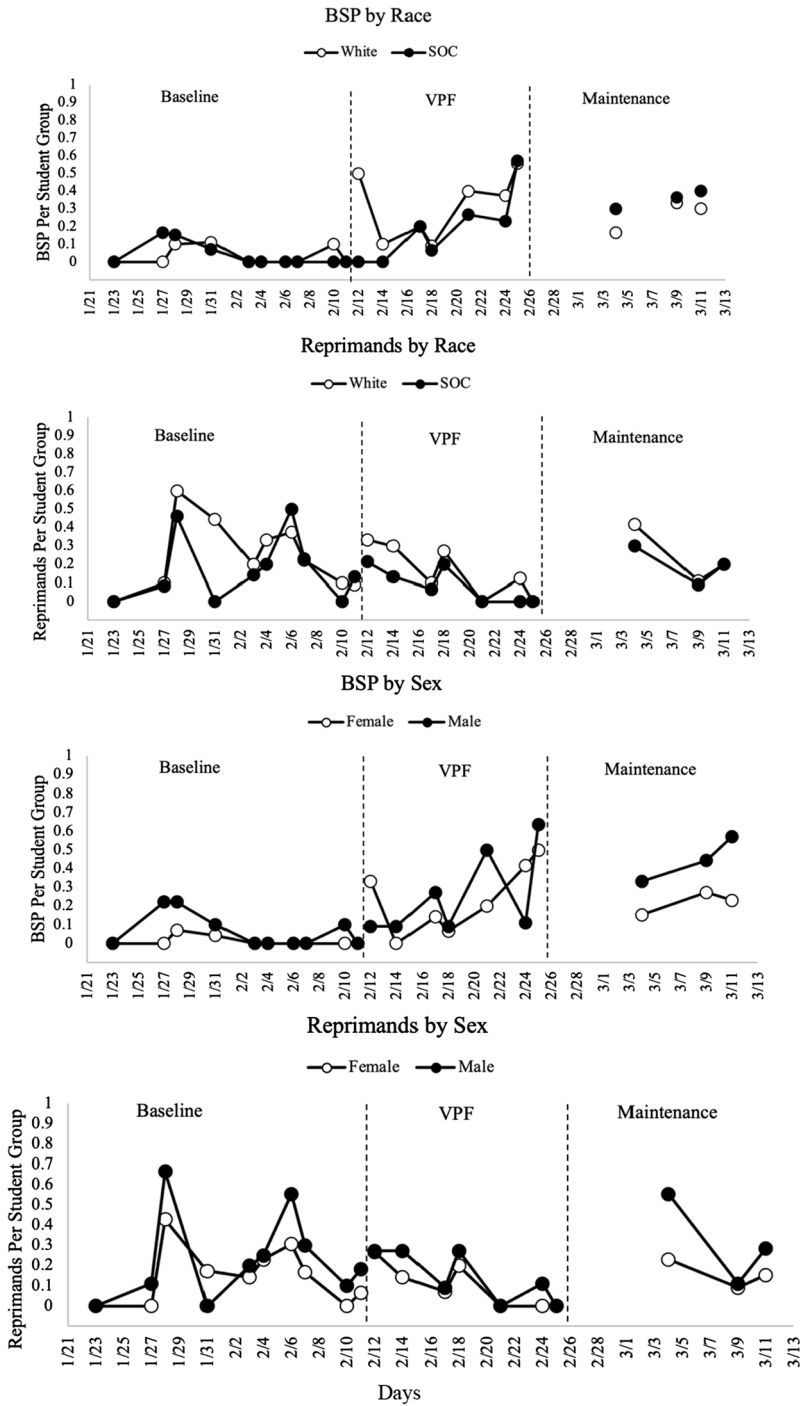


Fig. 3 Ms. Ball’s disaggregated data across phases. BSP is behavior-specific praise; SOC is student of color; VPF is visual performance feedback

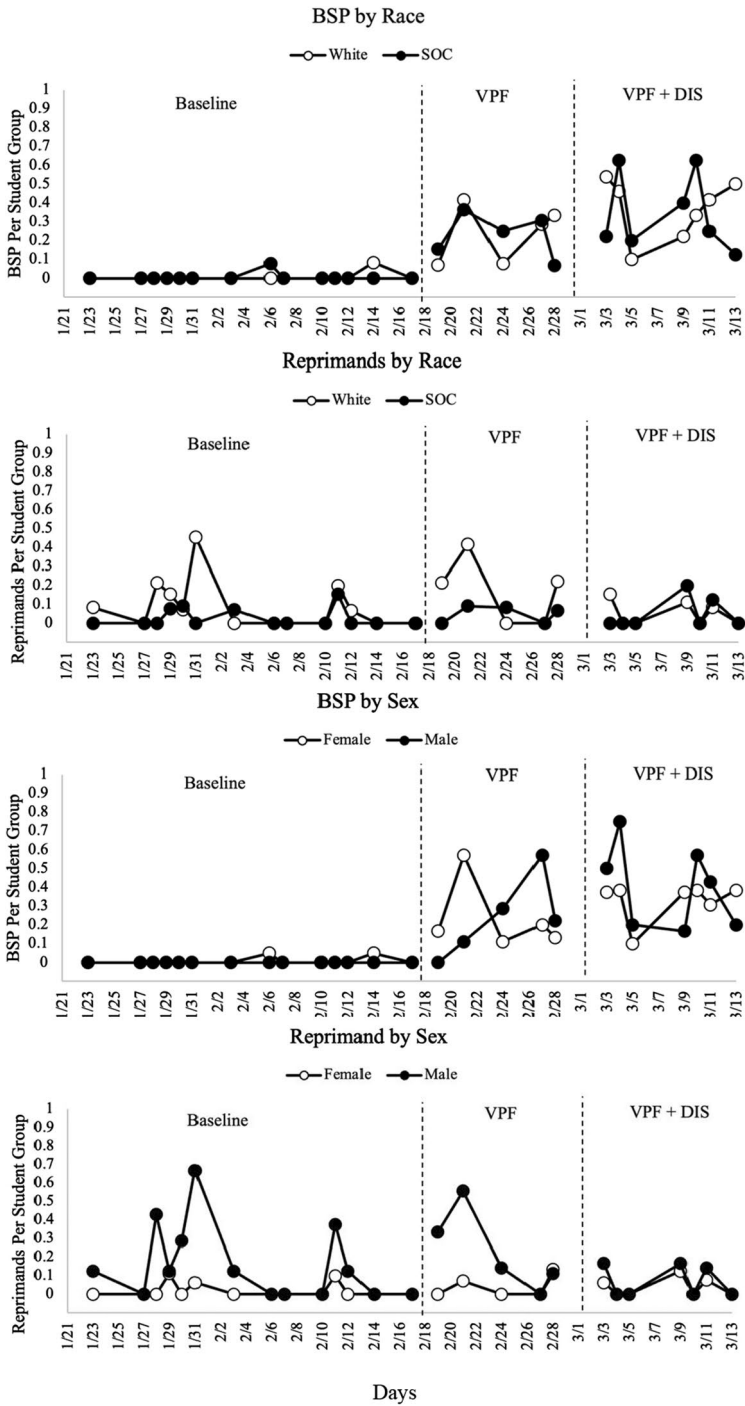


Fig. 4 Mr. Cox’s disaggregated data across phases. BSP is behavior-specific praise; SOC is student of color; VPF is visual performance feedback; VPD + DIS is disaggregated visual performance feedback

baseline, she delivered low and relatively comparable rates of BSP. After entering VPF, a separation in data paths for BSP by race was evident; she delivered more BSP to students of color and thus one reason she required VPF + DIS. During VPF + DIS, she consistently delivered similar rates as the data points for each race remained close. The second graph shows reprimands by race. During baseline, she delivered higher rates of reprimands to students of color. In VPF, these rates converged to low levels for both race paths; this pattern continued into VPF + DIS.

Graph three shows Mrs. Cobb's BSP by sex (female students, closed data path; male students, open data path). During baseline, she delivered low and comparable rates of BSP. During VPF, her trends for BSP by sex were in opposite direction and so she required VPF + DIS. During VPF + DIS, BSP trends became comparable and relatively equivalent in the last three days. The final graph shows reprimands by sex. Through all phases, Mrs. Cobb delivered comparable rates of reprimands by sex.

ODR

No teacher delivered an ODR in target class. Thus, no pre-post analysis was necessary.

Social Validity

The average rating was 5.28 (range = 3–6; $SD = 0.69$) on the modified IRP-15. Ms. Ball was the only teacher who marked a three (slightly disagree) indicating that the intervention was not consistent with others she had used previously. Two teachers completed all open-ended questions, but due to COVID-19 school closures, a follow-up to complete the questionnaire with the other two teachers did not occur. Ms. Ball and Mr. Brown both wrote (a) the intervention was simple to engage in; (b) it was easy to review their data; and (c) they would recommend the intervention to other colleagues. Although Mr. Brown entered VPF + DIS due to inequities by sex, he indicated the intervention did not help him with racial discrepancies, though, his data indicated there were no racial discrepancies requiring intervention.

Discussion

The goal of this study was to examine the effects of VPF on teacher's use of BSP and reprimands across students' race and sex. A functional relation was demonstrated between VPF and increased BSP, as well as VPF and decreases in reprimands. Increases in BSP following implementation of teacher performance feedback interventions is consistent with previous findings (Allday et al., 2012; Knochel et al., 2022). Additionally, high rates of reprimands were observed during baseline and were similar to previous research (Gage et al., 2018; Hirn & Scott, 2014). Upon entering intervention, three teachers demonstrated lower rates of reprimands compared to baseline. Thus, to answer Research Question 1, we conclude that VPF was an effective intervention to decrease the frequency of reprimands and increase BSP for participating teachers. Overall, these results contribute to the literature base on

Fig. 5 Mrs. Cobb's disaggregated data across phases. BSP is behavior-specific praise; SOC is student of color; VPF is visual performance feedback; VPD+DIS is disaggregated visual performance feedback. Y-axis scales shift across graphs

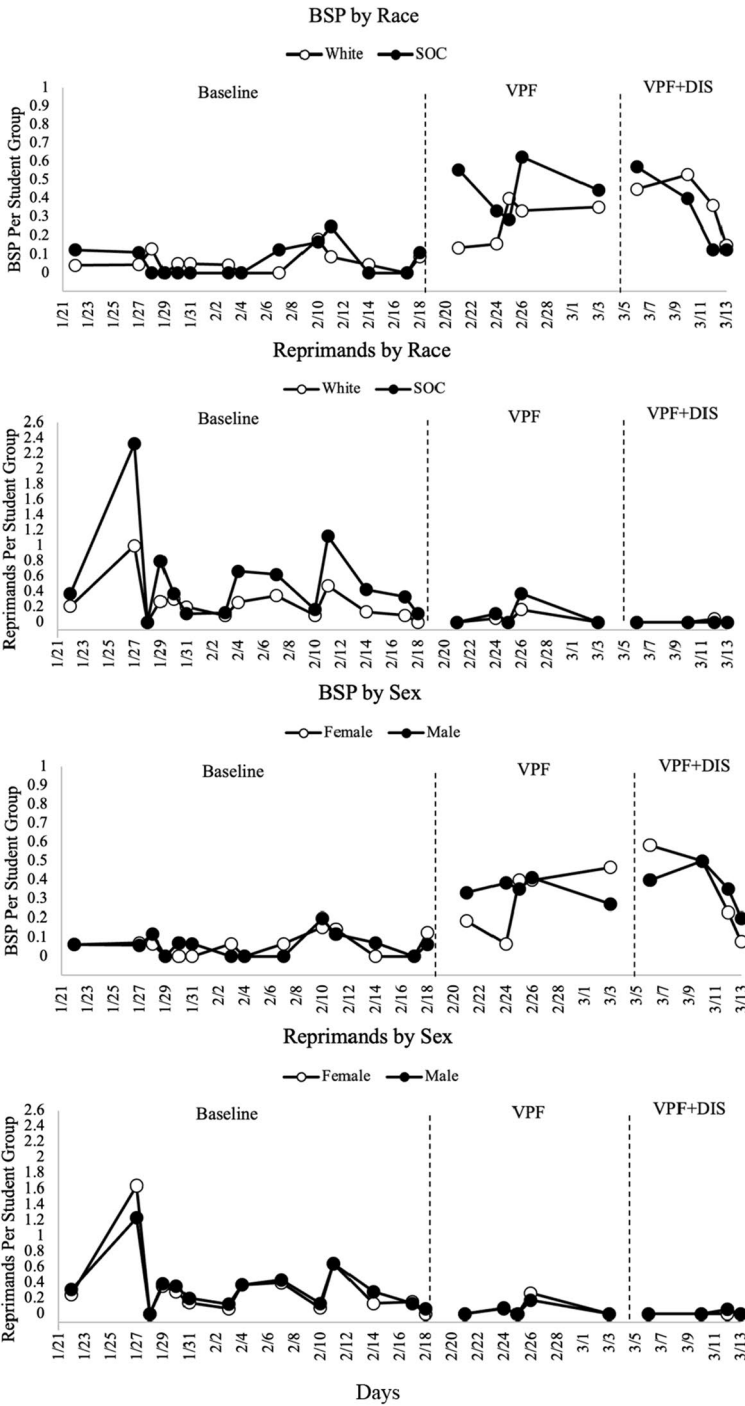
the effects of performance feedback in secondary settings, though further research is needed to establish it as an evidence-based practice in secondary settings (Author, in review).

Disaggregated Outcomes

We found mixed results for the teachers who required VPF+DIS. For instance, Mr. Brown and Mr. Cox trendlines for BSP by sex flipped directions and some patterns of data began to diverge despite being comparable in VPF (i.e., Mr. Brown's BSP by race). It may be that teachers in VPF+DIS focused their effort on delivering BSP by the characteristic that warranted entry into this phase and did not focus their attention on BSP by the other demographics. In similar studies targeting the equitable treatment of students in classroom, researchers provided teachers with performance feedback that included (a) only their praise-to-reprimand ratios of African American students compared to all other students (Gion et al., 2020), or (b) their total number of BSP and a list of students who received the most BSP and fewest corrections and who received the most corrections and fewest BSP (Knochel et al., 2022). It is plausible that a more parsimonious performance feedback intervention that incorporates less data or focuses on one characteristic may be more effective.

Another reason for mixed results may be related to student attendance fluctuating daily. During VPF+DIS, rates of statements per student group were discussed in terms of the average proportions of students in the class. However, teachers were not made aware of daily proportions of students in attendance. Given this daily fluctuation in attendance, results were likely impacted as individual sessions were subject to the number of students present in each demographic. Teachers may have had different results if they were made aware of the daily ratio. Relatedly, attendance likely impacts students' class behavior, as the presence or absence of students can impact the entire class. In conjunction with attendance data, future researchers could collect data on student problem behavior to analyze differences in teacher responding. Analyzing student problem behaviors in relation to how a teacher responds (i.e., reprimand, ODR) may aid in providing teachers with alternate strategies on how to intervene (Allday et al., 2021).

Second, VPF+DIS may not have provided enough support. Teachers may have benefited from individualized coaching to deliver equitable rates of BSP. This could include informing teachers that their BSP was developing signs of inequity, informing teachers about specific instances when they missed opportunities in delivering BSP to the underserved student, and providing examples on how to deliver BSP in multiple ways. This coaching could be delivered via a handwritten note (Knochel et al., 2022), email (Allday et al., 2012), an in-person conversation (Gion et al., 2020), or through real-time bug-in-ear technology (Schaefer & Ottley, 2018). Additionally, although the researcher checked for understanding in interpreting graphs during training, teachers



may have benefited from a discussion about their graphs in an in-person conversation as opposed to relaying only on email.

Because students of color receive higher rates of exclusionary discipline (Skiba et al., 2011), we hypothesized that students of color would receive fewer BSP and more reprimands. Overall, our hypothesis was not observed in this study. Although encouraging, this may have occurred for several reasons. First, the analytic method used may not have accounted for the appropriate trends. Data were aggregated by statements delivered to the entire class, small student groups, and to individual students for analysis. Analyzing statements delivered only to individual students (Knochel et al., 2022) may have yielded different trends. Second, no teacher sent any student to the office (ODR) in the target class. We can reasonably infer, then, these teachers were not experiencing considerable student disruptions, which may also have resulted in fewer reprimands. Finally, teacher inclusion criteria did not specify patterns of disproportionate treatment of students. Future researchers should implement procedures to examine data (screening observations, ODR) for bias patterns to determine a participant eligibility. Then, providing those teachers with data specific to which type of students they may have biases toward may reduce discrepancies. Researchers are cautioned in specifically pointing out the identified biases as this may further perpetuate their biases. Instead, providing individualized data is generally the preferred method for reducing bias (Girvan et al., 2016), though further research is needed to establish the most effective methods for reduction. Despite these mixed results, this study adds to the literature in important ways as few studies have examined the differential treatment of students in classrooms (Gion et al., 2020; Knochel et al., 2022) and these results can be used to inform future work.

Student Outcomes

In relation to Research Question 3, this study provides some evidence that when teacher behavior changes via VPF by increasing BSP and reducing reprimands, this may result in improved student outcomes as an increase in average class-wide academic engagement was evident. However, this was not the primary dependent variable and phase changes were not made based on student behavior. It is important to note that the engagement measure we used may not accurately reflect student engagement (Cooper et al., 2020). That is, when one student in a group of five displayed off-task behavior, the entire group was marked as not engaged even if the other four students were engaged. Thus, this measure may have underrepresented the actual class-wide engagement. Additionally, the measure did not capture student demographics, therefore we cannot comment on the differential effects of a change in teacher behaviors on the different student groups. Future researchers should consider simultaneously measuring student demographics with class-wide engagement to analyze the differential impacts on student groups.

Social Validity

All teachers rated the intervention with high social validity on the modified IRP-15. Anecdotally, all teachers made comments to the researcher about their preferences toward the intervention. Three teachers inquired about how this intervention could be implemented in colleagues' classrooms or built into the school's support system as an option for teams to consider for teachers needing support with BSP, reprimands, and/or equity. Ms. Ball mentioned that she had generalized effects in her athletic coaching as several players made comments on how much more positive she was at practice while she was in intervention. Collectively, these statements indicated the teachers' strong support for the intervention's social validity.

Limitations and Future Research

Several limitations should be considered. First, Mr. Brown and Ms. Ball, informed the first author after the study concluded that they had been members of the school's equity team. This team educated themselves about disproportionate practices and discussed avenues for improvement. Additionally, Mr. Brown reported the entire school received implicit bias training the previous year. Combined, these experiences may have impacted how these teachers interacted with their students. That said, all teachers screened into the study and despite these trainings and experiences, three teachers demonstrated a need for VPF+DIS. This may indicate their previous equity-related trainings were insufficient. When interventions target inequities, future researchers should survey participants on their training in this domain (e.g., course work, professional development) and consider analyzing teacher ODR trends to determine the extent to which biases exist prior to participation. We also encourage future researchers to add inclusionary criteria that measures disproportionate delivery of teacher statements into screening procedures. It is plausible that the two teachers excluded from the study at the onset may have delivered disproportionate praise and may have benefited from intervention.

We tracked statements delivered to individual students and to students in a group and these data were combined for phase change decisions. When a teacher delivered BSP to the entire class, this statement was recorded as delivered to every student. It is possible for a teacher to never deliver a BSP to an individual of a specific demographic yet the group gets marked as receiving a BSP when it was delivered to the entire class. Therefore, future researchers should analyze teacher statements delivered to individual students (Knochel et al., 2022) and exclude statements delivered to students in a group. Analyzing statements delivered to individual students may paint a more accurate picture of disproportionate practices.

Relatedly, researchers should outline clearer decision and analysis rules for disaggregated data. The triangulation of single-case visual analysis (i.e., level, trend, stability) for two data paths for different demographics and dependent measures made analysis and phase change decisions complex. This complexity was elevated when taking into account that, ideally, the two separate paths should be close together, yet they bounced

when the teacher's total frequency of statements fluctuated. Thus, examining the stability between individual data points, as opposed to overall pattern in the data paths, may be the most appropriate way to analyze disaggregated data visually, but clearer rules are warranted. Additionally, adding statistical analysis, such as Hedge's g , may aid in analysis (Gion et al., 2020).

Further, student demographics were not provided by the student or their families, as we did not have IRB or district approval to collect this information. Instead, we relied on our perceptions and the teachers' perceptions of student race and sex (Gion et al., 2020) and these perceptions may not have aligned with actual student identity. To advance equity work within schools, future researchers should obtain individual student demographic data from the students themselves or their families when targeting teacher treatment of students. It may be interesting to compare actual student demographics to teacher perceptions for large discrepancies and then examine how teachers' perceptions impact their interactions with students.

Another limitation is the inability to conclude that VPF alone resulted in a change in teacher behavior, as multiple components were included in the design. For instance, immediately following training, teachers entered intervention. It is possible that the training itself resulted in a change in behavior. Second, the researcher included BSP with the graphic feedback in every email making it difficult to examine the unique effect of graphic feedback alone. Thus, future researchers are encouraged to add in one additional component at a time and collect data following each added component or outline withdrawal procedures to test the effects of removing a single component. This would aid in determining which components are necessary and sufficient to result in a change in behavior.

Lastly, a doctoral student trained in school-based research, direct observations, and data analysis, served as interventionist. Although the intervention was designed to be parsimonious, a practitioner may not be able to implement such an intervention without training and support. To enhance the scalability, future researchers should include practitioners as interventionists. This would require professional development on how to collect reliable data and make data-informed decisions about phase changes, as teacher preparation programs generally do not provide practical experience in data-based decision-making (Majeika et al., in review). This type of training could be done in a series of sessions where teachers get hands-on experience collecting data, implementing intervention, and making data-based decisions alongside experts who provide on-going support (Bruhn et al., 2019). Providing this level of training could help practitioners as they grow in their knowledge, experience, and self-efficacy with data, and eventually, move this intervention from efficacious to effective (Hoagwood et al., 1995). Further, an effectiveness trial in which intervention delivery was not so tightly controlled and was implemented by practitioners would help determine usability, feasibility, and sustainability.

Conclusion

Teacher delivery of high rates of BSP is an evidence-based practice associated with positive student outcomes; yet, teachers tend to deliver less than optimal rates (Gage et al., 2018). A commonly used strategy to improve teaching practices is providing teachers with performance feedback. Little research has been conducted on the differential rates of BSP and reprimands by student race and sex. Results of the present study add to the body of research in important ways. First, results indicated that providing these high school teachers with VPF increased BSP, establishing a functional relation, and also reduced reprimands. Second, results provide some evidence that VPF may result in more equitable delivery of BSP and reprimands by race or sex although more research is needed. Third, increasing BSP and reducing reprimands resulted in improvements in average class academic engagement for majority of the teachers. Finally, the teachers found the intervention to be socially valid. Though limitations of this study exist, the results of this study are promising and portend well for future research and practice.

Declarations

Conflict of interest The authors report no conflicts of interest.

Ethical Approval This study followed ethical guidelines. Prior to conducting this study, the University of Iowa's Institutional Review Board approved the study, including the participant informed consent procedures and we obtained permission from the school district to conduct this study.

References

- Allday, R. A., Burt, J. L., & Haggard, K. N. (2021). Considering a function-base approach for reducing disproportional disciplinary practices. *Preventing School Failure: Alternative Education for Children and Youth*, 65(4), 332–342.
- Allday, R., Hinkson-Lee, K., Hudson, T., Neilsen-Gatti, S., Kleinke, A., & Russel, C. S. (2012). Training general educators to increase behavior-specific praise: Effects on students with EBD. *Behavioral Disorders*, 37(2), 87–98. <https://doi.org/10.1177/019874291203700203>
- Anderson, K., & Ritter, G. (2020). Do school discipline policies treat students fairly? *Evidence from Arkansas. Educational Policy*. <https://doi.org/10.1177/0895904818802085>
- Blake, J., Butler, B., Lewis, C., & Darensbourg, A. (2011). Unmasking the inequitable discipline experiences of urban Black girls: Implications for urban educational stakeholders. *The Urban Review*, 43(1), 90–106. <https://doi.org/10.1007/s11256-009-0148-8>
- Bruce, C. D., Esmonde, I., Ross, J., Dookie, L., & Beatty, R. (2010). The effects of sustained classroom-embedded teacher professional learning on teacher efficacy and related student achievement. *Teaching and Teacher Education*, 26, 1598–1608.
- Bruhn, A., Estrapala, S., Mahatmya, D., Rila, A., & Vogelgesang, K. (2019). Professional development on data-based individualization: A mixed research study. *Beyond Behavior*. <https://doi.org/10.1177/0198742919876656>
- Cavanaugh, B. (2013). Performance feedback and teachers' use of praise and opportunities to respond: A review of the literature. *Education and Treatment of Children*, 36, 111–136. <https://doi.org/10.1353/etc.2013.0001>

- Chappuis, S., Chappuis, J., & Stiggins, R. (2009). Supporting teachers. *Educational Leadership*, 66(5), 56–60.
- Chesley, G., & Jordan, J. (2012). What's missing from teacher prep. *Educational Leadership*, 69(8), 41–45.
- Cooper, J., Heron, T., & Heward, W. (2020). *Applied behavior analysis* (3rd ed.). Pearson.
- Downs, K., Caldarella, P., Larsen, R., Charlton, C., Wills, H., Kamps, D., & Webby, J. (2019). Teacher praise and reprimands: The differential response of students at risk of emotional and behavioral disorders. *Journal of Positive Behavior Interventions*, 21(3), 135–147. <https://doi.org/10.1177/1098300718800824>
- Fabelo, T., Thompson, M., Plotkin, M., Carmichael, D., Marchbanks III, M., & Booth, E. (2011). Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement. *Council of State Governments Justice Center*. http://knowledgecenter.csg.org/drupal/system/files/Breaking_School_Rules.pdf
- Fallon, L. M., Collier-Meek, M. A., Maggin, D. M., Sanetti, L. M., & Johnson, A. H. (2015). Is performance feedback for educators an evidence-based practice? A systematic review and evaluation based on single-case research. *Exceptional Children*, 81(2), 227–246. <https://doi.org/10.1177/0014402914551738>
- Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Freeman, J., Simonsen, B., Briere, D. E., & MacSuga-Gage, A. S. (2014). Pre-service teacher training in classroom management: A review of state accreditation policy and teacher preparation programs. *Teacher Education and Special Education*, 37(2), 106–120. <https://doi.org/10.1177/0888406413507002>
- Gable, R. A., Hester, P. H., Rock, M. L., & Hughes, K. G. (2009). Back to basics: Rules, praise, ignoring, and reprimands revisited. *Intervention in School and Clinic*, 44, 195–205. <https://doi.org/10.1177/1053451208328831>
- Gage, N., Katsiyannis, A., Carrero, K., Miller, R., & Pico, D. (2020). Exploring disproportionate discipline for Latinx students with and without disabilities: A national analysis. *Behavioral Disorders*. <https://doi.org/10.1177/0198742920961356>
- Gage, N., Scott, T., Hirn, R., & MacSuga-Gage, A. S. (2018). The relationship between teachers' implementation of classroom management practices and student behavior in elementary school. *Behavioral Disorders*, 43, 302–315. <https://doi.org/10.1177/0198742917714809>
- Gion, C., McIntosh, K., & Falcon, S. (2020). The effects of a multifaceted classroom intervention on racial disproportionality. *School Psychology Review*. <https://doi.org/10.1080/2372966X.2020.1788906>
- Gion, C., McIntosh, K., & Smolkowski, K. (2018). Examination of American Indian/Alaska Native school discipline disproportionately using the vulnerable decision points approach. *Behavioral Disorders*, 44(1), 40–52. <https://doi.org/10.1177/0198742918773438>
- Girvan, E. J., Deason, G., & Borgida, E. (2015). The generalizability of gender bias: Testing the effects of contextual, explicit, and implicit sexism on labor arbitration decisions. *Law and Human Behavior*, 39(5), 525. <https://doi.org/10.2139/ssrn.2595796>
- Girvan, E. J., Gion, C., McIntosh, K., & Smolkowski, K. (2016). The relative contribution of subjective office referrals to racial disproportionality in school discipline. *School Psychology Quarterly*, 32(3), 392. <https://doi.org/10.1037/spq0000178>
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4. <https://doi.org/10.1037/0033-295x.102.1.4>
- Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, 94(4), 945–967. <https://doi.org/10.2307/20439056>
- Haydon, T., & Musti-Rao, S. (2011). Effective use of behavior-specific praise: A middle school case study. *Beyond Behavior*, 20(2), 2564.
- Hirn, R., & Scott, T. (2014). Descriptive analysis of teacher instructional practices and student engagement among adolescents with and without challenging behavior. *Education and Treatment of Children*, 37(4), 589–610. <https://doi.org/10.1353/etc.2014.0037>
- Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to the special section: Efficacy and effectiveness in studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology*, 63(5), 683.

- Knochel, A. E., Blair, K., Kincaide, D., & Randazzo, A. (2022). Promoting equity in teachers' use of behavior-specific praise with self-monitoring and performance feedback. *Journal of Positive Behavior Interventions*, 24, 17–31. <https://doi.org/10.1177/1098300720951939>
- Kouchaki, M., & Smith, I. (2014). The morning morality effect: The influence of time of day on unethical behavior. *Psychological Science*, 25, 95–102.
- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, 34(1), 26–38. <https://doi.org/10.1177/0741932512452794>
- Krezmien, M. P., Leone, P. E., & Achilles, G. M. (2006). Suspension, race, and disability: Analysis of statewide practices and reporting. *Journal of Emotional and Behavioral Disorders*, 14, 217–226. <https://doi.org/10.1177/10634266060140040501>
- Lehmann, P. S., Azimi, A. M., Fortney, K., & Alaniz, K. (2021). Racial and ethnic disparities in school discipline: The interactive effects of gender and parental educational attainment. *Crime and Delinquency*. <https://doi.org/10.1177/00111287211029879>
- Lloyd, B. P., Bruhn, A. L., Sutherland, K. S., & Bradshaw, C. P. (2019). Progress and priorities in research to improve outcomes for students with or at risk for emotional and behavioral disorders. *Behavioral Disorders*, 44, 85–96. <https://doi.org/10.1177/0198742918808485>
- Losen, D. J., Hodson, C., Keith, M. A., Morrison, K., & Belway, S. (2015). Are we closing the school discipline gap? Center for Civil Rights Remedies at The Civil Rights Project at UCLA.
- Madigan, D. J., & Kim, L. E. (2021). Towards and understanding of teacher attrition: A meta-analysis of burnout, job satisfaction, and teachers' intentions to quit. *Teaching and Teacher Education*. <https://doi.org/10.1016/j.tate.2021.103425>
- Majeika, C. E., Wehby, J. H., & Wilkinson, S. (in review). How are we training preservice special education teachers to use data to monitor the progress of behavioral interventions?
- Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. X. (1985). Teacher judgements concerning the acceptability of school-based interventions. *Professional Psychology: Research and Practice*, 16(2), 191–198. <https://doi.org/10.1037/0735-7028.16.2.191>
- McIntosh, K., Campbell, A. L., Carter, D. R., & Zumbo, B. D. (2017). Concurrent validity of office discipline referrals and cut points used in schoolwide positive behavior support. *Behavioral Disorders*, 34(2), 100–113. <https://doi.org/10.1177/019874290903400204>
- Myers, D. M., Simonsen, B., & Sugai, G. (2011). Increasing teachers' use of praise with a response-to-intervention approach. *Education and Treatment of Children*, 34(1), 35–59. <https://doi.org/10.1353/etc.2011.0004>
- Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., & Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*, 34(1), 87–106. <https://doi.org/10.1080/02796015.2005.12086277>
- Oliver, R., Wehby, J., & Daniel, J. (2011). Teacher classroom management practices: Effects on disruptive or aggressive student behavior. *Campbell Systematic Reviews*. <https://doi.org/10.4073/csr.2011.4>
- Pas, E. T., Bradshaw, C. P., & Mitchell, M. M. (2011). Examining the validity of office discipline referrals as an indicator of student behavior problems. *Psychology in the Schools*, 48(6), 541–555. <https://doi.org/10.1002/pits.20577>
- Pisacreta, J., Tincani, M., Connell, J. E., & Axelrod, S. (2011). Increasing teachers' use of a 1: 1 praise-to-behavior correction ratio to decrease student disruption in general education classrooms. *Behavioral Interventions*, 26(4), 243–260. <https://doi.org/10.1002/bin.341>
- Rathel, J., Drasgow, E., Brown, W., & Marshall, K. (2014). Increasing induction-level teachers' positive-to-negative communication ratio and use of behavior-specific praise through e-mailed performance feedback and its effect on students' task engagement. *Journal of Positive Behavior Interventions*, 16, 219–233. <https://doi.org/10.1177/1098300713492856>
- Riddle, T., & Sinclair, S. (2019). Racial disparities in school-based disciplinary actions are associated with county-level rates of racial bias. *Psychological and Cognitive Sciences*. <https://doi.org/10.1073/pnas.1808307116>
- Ross, S. W., Romer, N., & Horner, R. H. (2012). Teacher well-being and the implementation of school-wide positive behavior interventions and supports. *Journal of Positive Behavior Interventions*, 14(2), 118–128. <https://doi.org/10.1177/1098300711413820>

- Royer, D. J., Lane, K. L., Dunlap, K. D., & Ennis, R. P. (2019). A systematic review of teacher-delivered behavior-specific praise on K–12 student performance. *Remedial and Special Education, 40*(2), 112–128. <https://doi.org/10.1177/0741932517751054>
- Schaefer, J., & Ottley, J. R. (2018). Evaluating immediate feedback via bug-in-ear as an evidence-based practice for professional development. *Journal of Special Education Technology, 33*(4), 247–259. <https://doi.org/10.1177/0162643418766870>
- Simonsen, B., Myers, D., & DeLuca, C. (2010). Teaching teachers to use prompts, opportunities to respond, and specific praise. *Teacher Education and Special Education, 33*(4), 300–318. <https://doi.org/10.1177/0888406409359905>
- Skiba, R., Horner, R., Chung, C., Rausch, M., May, S., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review, 40*, 256.
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review, 34*, 317–342. <https://doi.org/10.1023/A:1021320817372>
- Smith, D., & Smith, B. (2006). Perceptions of violence: The views of teachers who left urban schools. *The High School Journal, 89*, 34–42. <https://doi.org/10.1353/hsj.2006.0004>
- Solomon, B. G., Klein, S. A., & Politylo, B. C. (2012). The effect of performance feedback on teachers' treatment integrity: A meta-analysis of the single-case literature. *School Psychology Review, 41*, 160–175. <https://doi.org/10.1080/02796015.2012.12087518>
- Spilt, J. L., Leflot, G., Onghena, P., & Colpin, H. (2016). Use of praise and reprimands as critical ingredients of teacher behavior management: Effects on children's development in the context of a teacher-mediated classroom intervention. *Prevention Science, 17*, 732–742. <https://doi.org/10.1007/s11121-016-0667-y>
- Sutherland, K. S., Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *Journal of Emotional and Behavioral Disorders, 8*(1), 2–8. <https://doi.org/10.1177/106342660000800101>
- Tapp, J., Wehby, J., & Ellis, D. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, and Computers, 27*, 25–31.
- U.S. Department of Education Office for Civil Rights. (2014). *Civil rights data collection: Data snapshot*.
- U.S. Department of Justice & U.S. Department of Education. (2014). *Dear colleague letter on the nondiscriminatory administration of school discipline*.
- Van Acker, R., Grant, S. H., & Henry, D. (1996). Teacher and student behavior as a function of risk for aggression. *Education and Treatment of Children, 52*, 316–334.
- Witt, J. C., & Elliott, S. N. (1985). *Acceptability of classroom intervention strategies* (Vol. 4, pp. 251–288). T. R. Kratochwill (Ed.), *Advances in school psychology*.
- Wun, C. (2016). Unaccounted foundations: Black girls, anti-black racism, and punishment in schools. *Critical Sociology, 42*, 737–750. <https://doi.org/10.1177/0896920514560444>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Ashley Rila¹  · Allison L. Bruhn¹ · Lanqi Wang¹

Allison L. Bruhn
Allison-bruhn@uiowa.edu

Lanqi Wang
Lanqi-wang@uiowa.edu

¹ N230 Lindquist Center, University of Iowa, Iowa City, IA 52242, USA