

Working Paper No. 23-05

Using technology to get inside the black box of instructional coaching: a feasibility study

Sam Sims

University College London

Kate Forbes

Brunel University London

Josh Goodrich

Steplab

Instructional coaching has emerged as an effective form of teacher professional development. However, there is evidence of large variation in effectiveness between different coaches. What is it that differentiates more from less effective instructional coaching? Attempts to answer this question have been hampered by the difficulties of cost-effectively capturing variations in coaching practice. This paper reports on a pilot study using 360° (fisheye) video footage of teaching captured using classroom cameras, as well as audio recordings of coaching conversations uploaded via an online instructional coaching platform. The main aim of this research was to understand the feasibility of using such technology to get inside the black box of instructional coaching. We found that the camera technology could indeed capture meaningful variation in teachers' practice after a coaching session. Likewise, we found that the audio uploads (recorded via mobile phones) could capture content of the coaching conversation relevant to assessing leading hypotheses about what differentiates more and less effective coaching. Having said that, the project also surfaced several important challenges related to the way in which the cameras were used, which hampered our ability to consistently capture time-series data. The paper concludes with recommendations for researchers considering using this sort of technology in future projects.

VERSION: June 2023

Suggested citation: Sims, S., Forbes, K., & Goodrich, J. (2023). *Using technology to get inside the black box of instructional coaching: a feasibility study* (CEPEO Working Paper No. 23-05). Centre for Education Policy and Equalising Opportunities, UCL.

Disclaimer

Any opinions expressed here are those of the author(s) and not those of the UCL Institute of Education. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

CEPEO Working Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Highlights

- 360° (fisheye) video footage of teaching from classroom cameras can capture meaningful variation in teachers' practice after a coaching session
- Audio uploads (recorded via mobile phones) can capture content of the preceding coaching conversation relevant to assessing leading hypotheses about what differentiates more and less effective coaching
- Taken together, this technology has the potential to capture data relevant to getting inside the black box of effective instructional coaching
- However, there are important user-related (as opposed to technology-related) challenges involved in capturing the consistent time-series necessary to get inside the black box of instructional coaching

Why does this matter?

Without testing innovative new ways of collecting data, we are unlikely to gain an understanding of what differentiates more and less effective instructional coaching.

Using technology to get inside the black box of instructional coaching: a feasibility study

Sam Sims (UCL)

Kate Forbes (Brunel University London)

Josh Goodrich (Steplab)

Instructional coaching has emerged as an effective form of teacher professional development. However, there is evidence of large variation in effectiveness between different coaches. What is it that differentiates more from less effective instructional coaching? Attempts to answer this question have been hampered by the difficulties of cost-effectively capturing variations in coaching practice. This paper reports on a pilot study using 360° (fisheye) video footage of teaching captured using classroom cameras, as well as audio recordings of coaching conversations uploaded via an online instructional coaching platform. The main aim of this research was to understand the feasibility of using such technology to get inside the black box of instructional coaching. We found that the camera technology could indeed capture meaningful variation in teachers' practice after a coaching session. Likewise, we found that the audio uploads (recorded via mobile phones) could capture content of the coaching conversation relevant to assessing leading hypotheses about what differentiates more and less effective coaching. Having said that, the project also surfaced several important challenges related to the way in which the cameras were used, which hampered our ability to consistently capture time-series data. The paper concludes with recommendations for researchers considering using this sort of technology in future projects.

Key Words: teachers, professional development, instructional coaching

Acknowledgments: Thanks to ONVU Learning for providing the cloud platform, 360° camera & classroom audio technology, and to Steplab for providing the coaching audio capture technology, used in this paper. Thanks also to the staff and pupils at the two multi-academy trusts that participated in this research and the SCUDO project team for steering it.

Declaration of interests: Josh Goodrich is founder and CEO of Steplab. The cameras used in this project were generously lent to us by ONVU Learning. Use of the Steplab platform was partly supported by Steplab.

Instructional coaching has a positive effect on teacher practice and pupil achievement (Kraft et al., 2018). In an important recent paper, however, Blazar et al. (2021) have shown that instructional coaches vary considerably in their effectiveness, estimating that the variation in impact across coaches is close to the average impact of coaching PD programmes. Indeed, Blazar goes as far as to say that “coaches *are* the intervention, with a unique ability to directly influence teachers’ practice” (2020, p.122). This striking finding prompts the question: what it is that effective coaches do that makes them so much more influential in terms of improving teachers’ practice, relative to other coaches? Better understanding these, apparently strong, moderators of coach effectiveness is essential to improve selection and training of coaches.

A leading hypothesis is that the *relationship* between coach and coachee is important in explaining variations in effectiveness (Blazar et al., 2021; Boguslav, 2023; Woulfin & Jones, 2017). Indeed, guides for instructional coaches often emphasise the importance of coaches finding common ground with coachees, empathising with the challenges they face, and building trusting relationships with them (Knight, 2010). This is thought to be a pre-condition for coachees openly sharing their challenges with coaches and for coachees being receptive to formative feedback on their practice (Lowenhaupt et al., 2014; Woulfin et al., 2020; Yopp, 2019). These relational aspects of effective coaching are captured by the idea of rapport, by which we mean the feelings of interest, positivity and coordination experienced within a relationship with another (Cappella, 1990). However, while there is some suggestive and indirect empirical support for this claim (Blazar et al., 2021; Lowenhaupt et al., 2014; Woulfin & Jones, 2017), there is currently no strong evidence either way.

One important reason for the lack of evidence on the importance of rapport - or indeed any potential moderator of coach effectiveness - is the difficulty of cost-effectively yet systematically capturing data on the content of coaching conversations. In this pilot study, we set out to test the feasibility of using recently-developed audio and video capture technology to systematically capture variation in coaches’ methods and whether these relate to changes in coachees’ practice. We use fisheye classroom camera technology in an attempt to capture quantitative changes in coachees’ use of positive behaviour management (PBM) techniques in the weeks before and after receiving coaching on PBM. We then use data from audio uploads embedded within an online coaching platform in an attempt to capture variation in coaches’ approach to building rapport with their coachees. In both cases, our aim for this small pilot study is to test whether we can capture variation of the sort that would be necessary to test hypotheses about moderators of coaching effects. While our ultimate interest is in testing the hypothesis about rapport – and other potential moderators of instructional coaching – that would require a much larger sample.

Substantive Focus: positive behaviour management (PBM)

We chose to study instructional coaching focused on behaviour management. There are three reasons for this. First, behaviour management is frequently cited by early-career teachers as their top priority for professional development (Jones, 2013). This is important because we want to avoid a situation in which teachers fail to change their practice to begin with because they do not see the relevance of the PD content. Second, the codifiable and observable nature of behaviour management methods allows us to precisely capture persistence and fadeout from one lesson to another. For example, it is straightforward for us to observe and count instances of a teacher praising pupils for good behaviour over the course of a lesson. This contrasts with less directly observable aspects of teaching, which would likely have to be captured through high-inference observation rubrics that typically require averaging scores across four or more lessons to obtain minimum reliability thresholds (Kane & Staiger, 2012; Meyer et al., 2011). Third, behaviour management is in an important part of all lessons, in that no teaching can occur without an orderly classroom. By contrast, if we had focused on assessment, some lessons may contain no observable instance of assessment – resulting in a very noisy measures of changes in practice.

More specifically, we study PD focused on the use of positive behaviour management (PBM) methods. PBM focuses on the proactive teaching and positive reinforcement of desirable behaviours (Hawkins & Weis, 1985; Horner et al., 1990). A distinctive advantage of PBM is that it can be used as an antecedent (upstream) influence on behaviour, whereas sanctions or reprimands can only meaningfully be used after misbehaviour has occurred. By contrast, PBM is intended to prevent the occurrence of misbehaviour and thereby reduce the need to use sanctions (Epstein et al., 2008; Horner et al., 1990). In doing so, PBM is theorised to reduce ‘secondary behaviour incidents’ in which pupils dispute that they were misbehaving or protest any sanctions imposed. This prevents further wasting of classroom time (Colvin & Scott, 2014). In cases where misbehaviour does occur, a PBM approach would involve (at least initially) framing the correction in terms of the desired behaviour, ideally without naming the misbehaving pupils. This is thought to further reinforce the norm around good behaviour, rather than drawing attention to and thereby normalising misbehaviour (Ellis & Tod, 2018; Rhodes et al., 2020).

The empirical literature on behaviour management methods remains quite small and evaluations largely focus on multifaceted programmes that reflect multiple theoretical approaches (Evertson, 1989; Korpershoek et al., 2016; Oliver et al., 2011). For example, there are a number of evaluations of the so-called School Wide Positive Behaviour Interventions and Support approach, which reflects certain aspects of PBM, and seems to have positive effects on behaviour (Bradshaw et al., 2010; Muscott et al., 2008; Ward & Gersten, 2013). However, this programme also involves

whole-school reforms and the close involvement of school leaders. The few teacher-level PBM evaluations in the existing literature are however also broadly supportive of the approach (Närhi et al., 2015; 2017; Sutherland et al., 2000). In any case, what is critical for our purposes is that teachers can be coached on the use of PBM and that we can test the feasibility of measuring changes in their use of PBM using our camera technology.

Methods

Setting and participants

This research was conducted in two large multi-academy trusts (MATs, or groups of schools) in England. One of the MATs is based in the north of England and the other is based in the south-west of England. Both MATs contain a mixture of primary schools (pupils age 4-11) and secondary schools (pupils age 11-16+). Both MATs also offer school-led initial teacher training/preparation. This involves pre-qualification/service teachers being trained on-the-job, while working in a school on a reduced timetable. The northern MAT provides this training in partnership with a local higher education institution, usually to around 20 teachers per year. The southern MAT provides this training independently, usually to around 50 teachers per year. We recruited the teacher participants in this research from the 2022/23 cohort of trainees across the school-based initial teacher training programmes in these two MATS. Once we had recruited the trainees, we approached their coaches and invited them to participate in the research. This left us with an initial sample of 35 consenting coach/coachee pairs. This research was given ethical clearance by the UCL Institute of Education Ethics Committee.

Procedure

All of the early-career teachers participating in our study received coaching on PBM, delivered via the Steplab platform. Steplab is a web-platform and content library designed to support instructional coaching in schools. The Steplab coaching sequence has two phases: observation and feedback. During the observation phase, coaches observe the teacher's lesson and identify an area in which the teacher could improve. The Steplab online coaching library supports coaches in selecting a relevant, granular 'action step' on which to work with the teacher. Each action step comes with a series of 'success criteria', which reflect observable aspects of high-quality use of the focal action step. To begin the feedback phase, the coach asks a series of probing questions about the observed lesson to help the teacher build better mental models of what is happening in their classroom. The coach then models the focal action step, taking care to exemplify all the success criteria. The teacher then engages in a rehearsal, with the coach providing feedback focused on the success criteria. Finally, the teacher works with the coach to plan how the action step will be put into practice in a

subsequent lesson. The coach observes the teacher using the action step in this lesson, and the coaching cycle repeats.

The positive behaviour management action step in Steplab is: “Ensure expectations are met and maintain a positive classroom culture by framing behavioural interventions positively.” The four success criteria for this action step are:

- 1) stating the desired positive behaviour from students e.g., “eyes on the board” rather than “nobody should be looking at their books”
- 2) verbalising recognising good behaviour e.g. “Thank you everybody for being silent”
- 3) acknowledging rather than praising good behaviour e.g. “John’s eyes are fixed on me” rather than “John is looking at me, fantastic stuff John”
- 4) anonymous corrections e.g., “we need one more people looking at the board” rather than “Alice, you need to look at the board”.

These four success criteria clearly reflect the PBM philosophy of using upstream reinforcement (criteria 1 and 3) of good behavioural norms or positively framed correction in a way that further reinforce norms and minimises secondary behaviour incidents (criteria 2 and 4). All participating teachers focused on the PBM action step for one week early in the first half term of the academic year. During the other weeks of the autumn term, coaches provided additional weekly coaching using the Steplab platform but were asked to avoid providing coaching on PBM, or other behaviour management action steps related to PBM.

Data and measures

We captured changes in teachers’ use of PBM using ONVU Learning. The ONVU Learning platform includes 360° cameras that are ‘always on’ during the lessons taught by participating teachers. To minimise data capture we only accessed videos for one video-recorded lesson (‘focal lesson’) for each participant in each week of the study. All videos of non-focal lessons were automatically deleted. Comparing the use of PBM methods in the weeks before and after the coaching session allowed us to capture change in the use of PBM. Using ONVU Learning to capture this data is substantially more cost effective than sending a human observer into particular lessons and limits any potential ‘Hawthorne effects’ in which teachers temporarily adapt their practice because they are aware of an observer at the back of the room.

For each focal lesson video, we coded the first 12 minutes of the lesson to capture our outcome metrics. Focusing on a fixed period within the lesson (12 minutes) allowed us to ensure that all teachers have the same opportunity to be observed using PBM methods. We chose to focus on the

first 12 minutes because this is a period in the lesson in which teachers are dealing with getting the class settled and then getting pupils started on the lesson, both of which are rich with opportunities to use PBM methods. For each of these twelve-minute videos, we chose to record the frequency with which two of four PBM success criteria occurred in coachees' lessons:

- 1) *Net positive instructions*: a count variable capturing the number of positively framed instructions given by a teacher, minus the number of negatively framed instructions given by the teacher. A positively framed instruction is any command that states what the teachers wants the students to do, while a negatively framed instruction is any command that states something the teacher doesn't want the pupils to do. Consistent with the theory behind PBM emphasising the reinforcement of norms of good behaviour in the classroom, we are interested in the *net* number of positively framed instructions. Also consistent with the theory behind PBM emphasising upstream avoidance of misbehaviour, we include all positively framed instructions, not just those that are a direct response to misbehaviour.
- 2) *Net positive recognition*: a count variable capturing the number of times that the teacher verbally recognises pupils' good behaviour, minus the number of times the teacher verbally recognises misbehaviour. This differs from a positive instruction in that it is a backward-looking recognition of something a pupil has already done, not a forward-looking request. Consistent with the theory behind PBM emphasising reinforcement of desirable norms, we are interested in the *net* number of positive recognitions.

Every time a participating coach logged into Steplab and selected the PBM action step, we prompted them to scan a QR code with their phone camera. Participants were then asked to capture the audio from their coaching conversation using their phone. On completion of the coaching conversation, the audio file was then automatically uploaded to the Steplab platform. The audio files were never saved on the users' phone. This allowed us to record the coaching conversations in order to explore the ways in which coaches built rapport with their coachees during the conversation. The data on the conversations was then linked back to the camera data for the same coachee using a pseudonymised identifier.

Results

Feasibility of using cameras to capture changes in teaching practice following coaching

Of our 35 consenting coach/coachee pairs, 23 classroom cameras returned a usable lesson recording in at least one week of the study period. Of the remaining 12, 4 were 'offline' (not available over the remote network). A further 4 of the 12 were online and capturing data but the teachers'

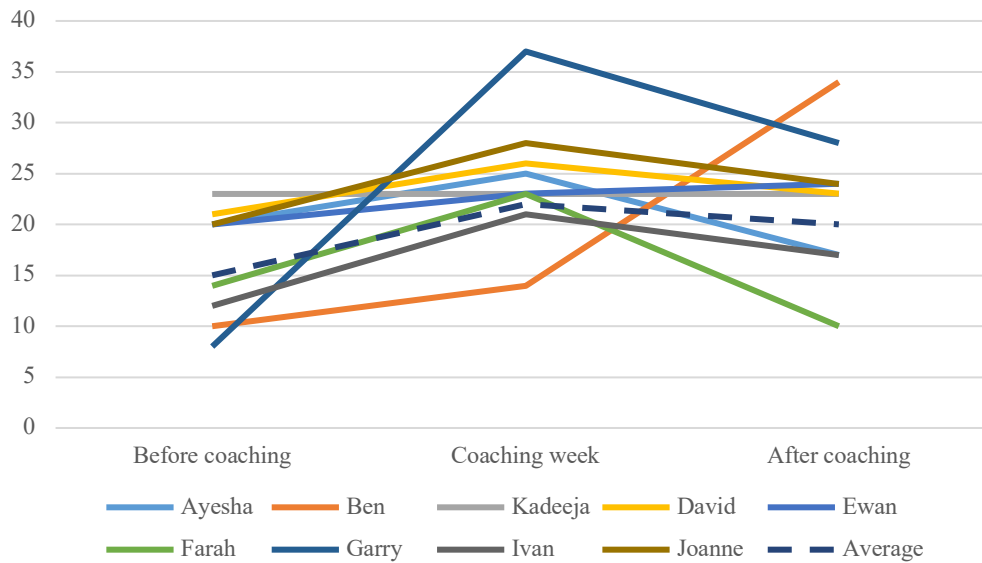
speech was inaudible, usually due to the microphone being installed close to other noisy technology such as a fan or air conditioning unit. A further 2 of the 12 did not have the schedule set up that allowed us to only access lessons conducted by the participating teacher, as opposed to other teachers using the same classroom. Finally, two cameras did not provide usable data because the coachee was moved into another classroom at short notice early in the term.

Recall that our goal here is to assess the feasibility and potential of capturing changes in coachees' PBM practice either side of a coaching conversation focused on PBM. This required us to have, at a minimum, a recording in the week prior to, during, and after the coaching conversation focused on PBM. Of the 23 classroom cameras that returned at least one usable recording, 10 provided the necessary string of before/during/after recordings that we needed to track changes in practice. The remaining 12 were missing footage in one or more of the relevant weeks. Despite the cameras being designed to be 'always on' (though not always recording), it appears that the cameras were sometimes being switched off. Our study coincided with a global spike in energy prices and, based on conversations with participating schools, it seems that staff had been instructed to switch off all electrical equipment when not being used.

For each of the 10 coachees with a consistent set of before/during/after recordings, we coded the videos to capture our two PBM metrics. For these 10 sets of recordings, we found that the video and audio data from the cameras was of high quality and allowed us to straightforwardly code up our two PBM metrics. Figures 1 and 2 show the time series for each of these two outcomes. Both charts include one (solid) line for each of the 10 teachers, as well as dashed line to show the average across the 10 teachers. The teachers have all been given pseudonyms to protect their anonymity. The lesson we coded in the 'Coaching week' may or not occur after the coaching itself. For example, 'Ben' received his coaching after the coded lesson, meaning that his use of PBM increased fastest between the 'Coaching week' and 'After coaching' week.

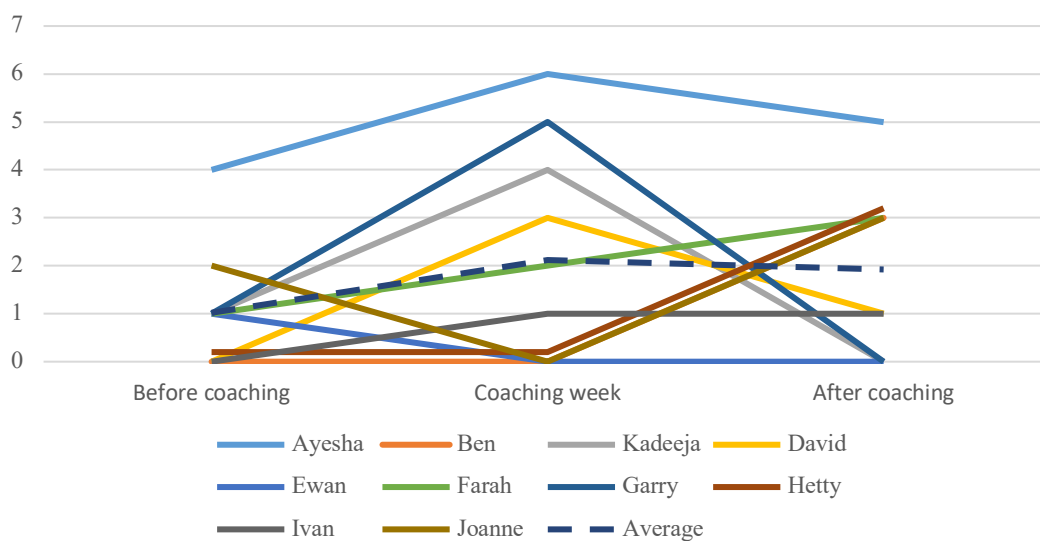
The first thing that stands out from the graphs is that, consistent with what we would expect based on the existing evidence about the efficacy of instructional coaching for changing practice (Kraft et al., 2018), we see an increase in both our PBM metrics. Net positive instructions increase from an average of 15 to an average of just over 20 in the week of the coaching and remains at 20 in the week following the coaching. Net positive recognition increases slightly, from an average of 1 to an average of 2. We interpret this as evidence that the 10 cameras which did provide the necessary string of data were sensitive enough to capture an increase in the use of PBM that we might expect after coaching focused on PBM.

Figure 1: Net positive instructions



The second result that stands out from these graphs is the large degree of heterogeneity in the trajectories across the teachers in both graphs. For example, some coaches show marked improvements in their use of net positive instructions (e.g., Gary and Ben), while others show very little change (e.g., Kadeeja and Ewan). This is consistent with what we would expect based on existing evidence about the substantial heterogeneity in coach effectiveness (Blazar et al., 2021). We now turn to analysing our coaching conversations data to explore the feasibility of using it to explain this between coach/coachee heterogeneity.

Figure 2: Positive recognition



Feasibility of using audio in explaining heterogeneity in changes in teaching practice following coaching

Of the 35 consenting coach-coachee pairs, 32 attempted to record at least one of their coaching conversations. Of these 32, 6 did not provide any usable recordings due to either the failure of the file(s) to upload properly or the uploaded recording(s) being inaudible. Of the 10 coach/coachee pairs that has a consistent set of before/during/after lesson videos, 5 uploaded usable coaching conversations (Ben, Kadeeja, David, Gary, Joanne). Due to space constraints, we focus our analysis here on two illustrative cases - Kadeeja and Gary - using all available recordings for each coachee. To reiterate, our aim is to test whether the audio recording of the coaching conversations can illuminate variation in rapport between coach-coachee pairs. This feasibility study is not concerned with systematically explaining the variation described above, which would require systematic measurement and a larger sample.

Gary and his coach, Amy

Amy often began coaching conversations by praising Gary's good practice: "Fantastic [...] I was really impressed." Although praise is a standard part in the Steplab cycle, we judged Amy's affirmations to feel authentic, rather than formulaic. Through most of the coaching conversation, Amy used a colloquial, conversational style ("when they were faffing around with whiteboard pens"), which also contributed to a sense of authenticity. Early in the coaching session, Gary raised a particularly difficult student in the class and Amy took care to listen to and empathised with Gary's frustrations: "It is something that is really, really difficult, especially with a class like that on a Monday afternoon."

Amy tended to begin the 'feedback' phases of the coaching cycles by inquiring about a particular part of Gary's lesson. For example, why in one lesson he had not sanctioned a student for some serious misbehaviour, which ran counter to the schools' wider behaviour policy: "You avoided going through the behaviour routine and taking the planner. Can you tell me why you did that?" Gary is willing to openly discuss his error here: "I think initial nerves. I felt it was a bit too confrontational but of course that is the expectations and so they should be followed through for consistency. I think I just shied away from being able to do that." This vulnerability suggests that Gary trusts Amy to put this information to good use in helping him improve his practice.

When discussing possible foci for the practice stage of the coaching, we noticed that, contrary to some of the other coaches in our data, Amy generally had not prejudged the answer. Instead, Amy was careful to provide choice at certain points in the conversation ("there's two ways we can do this"), which suggests that she is remaining attentive to Gary's input. We also noticed that Gary paid

close attention to what Amy was saying, at times rephrasing her sentences in his own words. This suggests that he also respects Amy's views. Following the rehearsal phase, Amy was again careful to acknowledge the challenges involved in using the PBM methods consistently well, while also reinforcing the value of the approach: "It's hard but it works really well."

Kadeeja and her coach, Tom

The conversation between Kadeeja and Tom tended to focus on teaching and learning and broadly stuck to the suggested coaching protocol. Having said that, it was notable that Tom used the protocol, and the notes he had made during the lesson observation, as a scaffold to support the conversation, rather than mechanically going through it, step-by-step. This lent a sense of authenticity and responsiveness to the conversations. Tom tended to start the conversation by thanking Kadeeja for allowing him to observe her lesson, which helped build a sense of reciprocity. In one conversation, he also apologised for the coaching session occurring at the end of the school day, rather than straight after the lesson observation, and acknowledging and empathising that "it's been a very long day".

Tom tended to begin the feedback phase of the conversation by making a more general observation about something that happened in the lesson. We noticed that this was often accompanied by a question about how Kadeeja felt about that part of the lesson, for example: "How did you feel when you had a go at that? How do you feel that it impacted on pupils' engagement?" This demonstrates Tom's regard for Kadeeja's experiences. This prompts Kadeeja to respond with an honest reflection "I think I really just had to sit with the discomfort of the pausing [...] I still don't feel thank that I have, like, complete control over the, you know like, silence" which in turn prompts Kadeeja to take ownership of this as a priority for improvement "I need to figure out how to nip that in the bud."

Discussion

Some instructional coaches are much more effective than others but research has yet to determine why (Blazar et al., 2021). One prominent hypothesis is that effective coaches invest in building rapport with their coachees, which creates a safe space for honest and constructive discussion about how to improve practice (Blazar et al., 2021; Woulfin & Jones, 2017). Attempts to test this hypothesis have, however, been hampered by the challenges of cost-effectively collecting granular data on the ways in which coaches build rapport and linking this to high-resolution data on changes in coachees' classroom practice. This study set out to test the feasibility of using recently-

developed audio-visual technology to get inside the black box of instructional coaching, with a view to testing rapport (and other potential moderators) of coach effectiveness.

We found that, by and large, the technology proved capable of capturing the necessary data. Of our 35 cameras, only four failed to provide productive data because the recording was not audible. In all four of these cases, we deemed it likely that simply installing the microphone in a different location (further away from other noises sources) would have fixed this problem. Reassuringly, for the 23 cameras that did provide usable lesson recordings, we found the quality of the audio was sufficient to support systematic coding of the focal teachers' practice, as reported in the charts above. With respect to the (phone-based) audio capture of coaching conversations, six of the 32 coachees that tried to upload an audio file failed for technical reasons, with the other 26 being successful. Although we are not entirely clear why some of the files failed to upload, it is plausible that a lack of Wi-Fi connection may be at fault. Again, the audio files that did upload were generally of high-quality. This allowed us to capture the nuanced ways in which coaches built rapport during conversations with their coachees, as illustrated in the case studies above.

Of course, the data we collected have some limitations. First, we chose to focus on two highly observable aspects of teachers classroom practice. Our findings should therefore be interpreted as proof that is *possible* to observe changes in classroom practice. There are many less frequent, less overt, yet nevertheless important aspects of teaching that would likely to be harder to capture and quantify. Second, the coaching conversations that we captured were loosely scaffolded by the Steplab platform. While Steplab does not provide scripting for the conversations, it does provide prompts around e.g., praising progress made by the coachee. While the case studies presented above do illustrate a variety of ways in which coaches build rapport with their coachees, it is important to keep in mind that coaches were responding to prompts from the coaching software.

The main challenges we encountered related more to the way in which the technology was used, rather than the technology itself. Two thirds of the cameras that failed to provide a usable lesson recording did so because they were offline, did not have the lesson schedules set up, or because the focal teachers moved to another classroom. Of the cameras that did provide a usable lesson recording, many did not capture the sequence of lesson necessary to observe change over time. Likewise, the average number of coaching conversations uploaded was three, out of possible eight opportunities. Since we are fundamentally interested in linking these two sources of data together, to understand how coach practice relates to changes in teaching practice, we ended up with only five coach/coachee pairs with the necessary *set* of data.

Implications for researchers

Future research considering using this approach should focus on minimising the number of ‘points of failure’ involved in the way that the technology is used. In particular, we would make four recommendations. First, time and resources should be set aside for rigorous testing of the cameras and microphones at the point of installation in order to avoid problems with audibility. Second, direct lines of contact should be established between those with oversight of the cameras and those with the ability to switch them on/off or established the recording schedules. Third, cameras should be clearly labelled to discourage people from switching them off and researchers should consider reimbursing schools upfront for the additional electricity costs. Fourth, researchers should consider sending personalised text message reminders to encourage consenting coaches to upload their coaching conversations.

Conclusion

Recently-developed technology is capable of capturing the type of data necessary to test potential moderators of instructional coaching. However, careful attention to the use and users of the technology is necessary to capture the *range* of data necessary to truly get inside the black box. Future research on this question should consider using this sort of technology while also paying careful attention to the human factors involved.

REFERENCES

- Blazar, D. (2020). Teacher Coaching to Improve Instruction at Scale: Opportunities and Challenges in Policy Contexts. *Teachers College Record*, 122(10), 1-9.
- Blazar, D., McNamara, D., & Blue, G. (2021). *Instructional coaching personnel and program scalability* (EdWorkingPaper No. 21-499). Annenberg Institute for School Reform at Brown University.
- Boguslav, A. (2023). Parsing coaching practice: a systematic framework for describing coaching discourse. https://education.virginia.edu/sites/default/files/2023-02/EPW_Parsing-Coaching-Practice_Working-Paper_02-02-2023.pdf
- Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions*, 12(3), 133-148.
- Cappella, J. N. (1990). On defining conversational coordination and rapport. *Psychological Inquiry*, 1(4), 303-305.
- Colvin, G., & Scott, T. M. (2014). *Managing the cycle of acting-out behavior in the classroom*. Corwin Press.
- Ellis, S., & Tod, J. (2018). *Behaviour for learning: Promoting positive relationships in the classroom*. Routledge.
- Evertson, C. M. (1989). Improving elementary classroom management: A school-based training program for beginning the year. *The Journal of Educational Research*, 83(2), 82-90.
- Hawkins, J. D., & Weis, J. G. (1985). The social development model: An integrated approach to delinquency prevention. *Journal of Primary Prevention*, 6(2), 73-97.
- Horner, R. H., Dunlap, G., Koegel, R. L., Carr, E. G., Sailor, W., Anderson, J., ... & O'Neill, R. E. (1990). Toward a technology of "nonaversive" behavioral support. *Journal of the Association for Persons with Severe handicaps*, 15(3), 125-132.
- Jones, V. (2013). How do teachers learn to be effective classroom managers?. In *Handbook of classroom management* (pp. 897-918). Routledge.
- Kane, T. J., & Staiger, D. O. (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Research Paper. MET Project. *Bill & Melinda Gates Foundation*.
- Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research*, 86(3), 643-680.
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547-588.
- Lowenhaupt, R., McKinney, S., & Reeves, T. (2014). Coaching in context: The role of relationships in the work of three literacy coaches. *Professional Development in Education*, 40(5), 740-757.
- Meyer, J. P., Cash, A. H., & Mashburn, A. (2011). Occasions and the reliability of classroom observations: Alternative conceptualizations and methods of analysis. *Educational Assessment*, 16(4), 227-243.
- Muscott, H. S., Mann, E. L., & LeBrun, M. R. (2008). Positive behavioral interventions and supports in New Hampshire: Effects of large-scale implementation of schoolwide positive behavior support on student discipline and academic achievement. *Journal of Positive Behavior Interventions*, 10(3), 190-205.
- Oliver, R. M., Wehby, J. H., & Reschly, D. J. (2011). Teacher classroom management practices: Effects on disruptive or aggressive student behavior. *Campbell Systematic Reviews*, 7(1), 1-55.

- Rhodes, N., Shulman, H. C., & McClaran, N. (2020). Changing norms: A meta-analytic integration of research on social norms appeals. *Human Communication Research*, 46(2-3), 161-191.
- 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.
- Ward, B., & Gersten, R. (2013). A randomized evaluation of the safe and civil schools model for positive behavioral interventions and supports at elementary schools in a large urban school district. *School Psychology Review*, 42(3), 317-333.
- Woulfin, S. L., & Jones, B. (2017). Rooted in relationships: An analysis of dimensions of social capital enabling instructional coaching. *Journal of Professional Capital and Community*.
- Yopp, D. A., Burroughs, E. A., Sutton, J. T., & Greenwood, M. C. (2019). Variations in coaching knowledge and practice that explain elementary and middle school mathematics teacher change. *Journal of Mathematics Teacher Education*, 22(1), 5-36.
- Woulfin, S. L. (2020). Crystallizing Coaching: an examination of the institutionalization of instructional coaching in three educational systems. *Teachers College Record*, 122(10), 1-32.

 @cepeo_ucl

ucl.ac.uk/ioe/cepeo