Ambition Institute

Ark Teacher Training



Why do some teachers improve more than others? Exploring initial teacher trainees' feedback orientation

Briony Banks, Jennifer Curran, Stefanie Meliss,
Sam Sims, Nazlin Chowdhury, Havva Gorkem Altunbas,
Nikoletta Alexandri, Leila MacTavish and Isabel Instone

2024

Contents

Contributions **3**

Acknowledgements **4**

Foreword **5**

Abstract **6**

Introduction **7**

What is feedback orientation? 7

How is feedback orientation relevant to teachers' professional development? 8

The present study 9

Methods **10**

Participants 10

Study design 10

Measures 13

Statistical analysis 15

Results **16**

Hypothesis 1: Feedback orientation will predict trainees' performance in a simulator training exercise 18

Hypothesis 2: Feedback orientation will predict trainees' improvement in classroom observation ratings between autumn and spring terms 20

Hypothesis 3: Trainees' improvements in the simulator training exercise will predict their improvements in classroom observation ratings between autumn and spring terms 22

Hypothesis 4: Trainees' feedback orientation will improve over their first term of training 23

Discussion **25**

Implications 27

Limitations 27

Conclusion 28

References **29**

Appendix **32**

Contributions

Conceptualisation

Sam Sims
Jennifer Curran
Briony Banks
Stefanie Meliss

Methodology

Sam Sims
Jennifer Curran
Briony Banks
Stefanie Meliss

Software

Connect Training

Validation

Briony Banks
Stefanie Meliss
Jennifer Curran

Formal analysis

Briony Banks
Jennifer Curran
Havva Gorkem Altunbas
Nikoletta Alexandri
Grace Carmichael
Stefanie Meliss

Investigation

Jennifer Curran
Nazlin Chowdhury
Jacob Brierley
Briony Banks
Stefanie Meliss

Resources

Sam Sims
Jennifer Curran
Briony Banks
Stefanie Meliss

Data curation

Briony Banks
Jennifer Curran
Stefanie Meliss

Writing

Briony Banks
Jennifer Curran
Stefanie Meliss

Editing

Sam Sims
Briony Banks
Jennifer Curran
Stefanie Meliss
Hannah Cheetham-Joshi

Visualisation

Stefanie Meliss
Jennifer Curran
Briony Banks

Supervision

Hannah Cheetham-Joshi
Marie Hamer
Sam Sims

Project administration

Briony Banks
Stefanie Meliss
Jennifer Curran
Abigail Brown
Aisha Nazeer
Sarah James

Funding acquisition

Hilary Spencer
Marie Hamer

Recruitment

Sam Sims
Jennifer Curran
Briony Banks
Stefanie Meliss
Leila MacTavish
Isabel Instone

Acknowledgements

Thanks to Leila MacTavish and Isabel Instone, as well as the rest of the staff and trainee teachers at Ark Teacher Training. Thanks to Susan Day, Sarah James, Jacob Brierley and the rest of the staff at Connect Training. Thanks to Amarbeer Singh Gill, Jess Tratt, Nick Pointer, Michelle Minton, Emma Lindley-Thompson, Laura Wilkins, Emma Yates, Emma Tate, Amber Harcourt and John Jackson at Ambition Institute for their engagement in piloting. Thanks to Nicola Bretscher, Piers Saunders, Gorana Henry and the rest of the staff and trainee teachers at University College London; Gary Craggs and the trainee teachers at Durham SCITT; Karen Hudson, Sophie Meller and the trainee teachers at Northumbria University and North Tyneside SCITT; Jane Gargan and the trainee teachers at Liverpool Hope University; Naomi Robinson and the trainee teachers at Oxford Brookes University; the trainee teachers at the University of Buckingham and Sheffield SCITT. The project was funded by Ambition Institute and funding for this project was acquired by Hilary Spencer and Marie Hamer. Finally, thanks to Dr Dominic Kelly at UCL for reviewing this paper.

Foreword

There is a wealth of evidence which asserts that teachers are the greatest lever to impact the outcomes, future pursuits and life chances of young people in schools. The role of teaching is complex and cognitively demanding, and it is critical that we acknowledge the nuance behind what it means to teach well. Teaching well is highly dependent upon our combined knowledge of subject curriculum, assessment and pedagogical strategies. This expert synthesis is how we ensure students learn.

Why is this difficult? Because teaching exists for the most part in the live practice of our classrooms in real time. We therefore need to rely on a bank of tools to develop a shared understanding of what we mean by high quality teaching. Feedback is one such dialogical tool that helps develop this collective model. Beyond this, feedback holds a multi-faceted role in developing professional relationships that leave colleagues feeling supported and empowered as a result of the clarity that feedback offers.

We know that feedback is also a dynamic process between colleagues. While the research suggests that this tool should feature heavily in our repertoire for supporting colleagues, we should also exercise caution in viewing feedback as a solitary tool through which we can determine and improve teacher quality.

For these reasons, this new research from Ambition Institute is an important contribution to our understanding of how school leaders can add value to a new teacher's professional induction. While this paper focuses on feedback in the Initial Teacher Training sphere, it has broader applications. It will be of great interest to anyone involved in coaching, mentoring or supporting colleagues to develop expertise in the classroom, alongside a wider repertoire of tools.

I am delighted to introduce this research from Ambition Institute. It is an important addition to increasing our collective understanding of how to support teachers through an ongoing dialogue around practice. It will be of real value to our work at Windsor Academy Trust, as well as colleagues across the sector who work to induct and develop teachers so that we can ensure our students thrive. Above all, this research is a demonstration of what it means to be a teacher, a role that is hugely valuable and deeply rewarding: they continue to be our most vital resource.

Kat Howard
Executive School Improvement & Curriculum and Assessment Lead,
Windsor Academy Trust

Abstract

Novice teachers receive a large amount of feedback during their training, whether from mentors, training providers or colleagues. Teachers' propensity to respond positively to feedback (their feedback orientation) could therefore affect how much they learn and improve during training. We tested whether novice teachers' feedback orientation predicts 1) how much they improve in a classroom simulator training exercise where they receive feedback on their teaching skills; and 2) classroom observations of their teaching practice over the first two terms of their school placements. We also examined whether teachers' feedback orientation improved during the first academic term of training.

Feedback orientation did not predict how much novice teachers improved during the training exercise, or how much they improved between their first and second terms of training. However, teachers' social awareness (for example whether they use feedback to make a good impression) predicted their classroom observation ratings in the first term of training. Trainees' feedback orientation also improved during this time, particularly their self-efficacy towards receiving feedback. Greater improvements in feedback orientation were related to better quality feedback from their coach or mentor, suggesting that providing high quality feedback to novice teachers may be important in developing their confidence in dealing with feedback. Overall, the findings suggest that feedback orientation is not a good indicator of how much novice teachers are likely to improve during their training, and further research is needed to understand why some teachers improve more than others.

Key words: teachers, professional development, teacher educators, feedback

Introduction

Receiving feedback[1] is an integral part of learning to teach. Trainee teachers receive a large amount of diverse feedback during their initial training, whether formally from their training providers, mentor or coach, or less formally from peers or other teaching staff. Feedback also plays a key part in teachers' professional development (PD) throughout their careers, for example, receiving peer feedback following lesson observations (Guskey & Link, 2022). However, anecdotally, some teachers are better able to process and act on this feedback than others. Empirically, we know that teachers respond to and perceive feedback differently (Cherasaro et al., 2016). These different responses to feedback could arguably affect the quality of their teaching and particularly their ability to develop and improve.

What drives these different responses to receiving feedback? Multiple factors are likely to play a role, such as how credible an evaluator is or how useful feedback is perceived to be (Cherasaro et al., 2016). One potential candidate is an individual's feedback orientation (FO), or their propensity to value and act on feedback (Linderbaum & Levy, 2010; London & Smither, 2002). Developed from performance management research, this individual characteristic can account for differences in how individuals react to, interpret and use feedback in the workplace (Katz et al., 2023; London & Smither, 2002). Differences in teachers' FO could thus impact how effectively teachers process and act upon the feedback they receive and thereby help to explain differences in teaching quality. The present study investigated the role of FO in novice teachers in three ways: whether it predicted their improvement during an online PD exercise (Hypothesis 1); whether it predicted improvements in their teaching skills during their first school placements (Hypotheses 2 and 3); and whether their FO improved over the first term of their training (Hypothesis 4).

What is feedback orientation?

In a theoretical paper, London and Smither (2002, pp. 82–83) proposed the concept of FO as "an individual's overall receptivity to feedback and the extent to which the individual welcomes guidance and coaching". They theorised it to be a multi-dimensional construct based on whether an individual likes and values receiving feedback; their propensity to seek feedback, process it carefully, and act on it; and how sensitive an individual is to others' views of them. Linderbaum and Levy (2010) developed and validated the popular Feedback Orientation Scale (FOS), which built on the theory of London and Smither (2002) as well as a previous Feedback Propensity Scale (Herold et al., 1996). Linderbaum and Levy's (2010) scale comprises four dimensions. These dimensions were initially proposed by London and Smither (2002), but have each been shown to predict how individuals respond to and use feedback: 1) utility, which refers to an individual's tendency to view feedback as useful for achieving their goals (for example, Brett & Atwater, 2001); 2) accountability, an individual's sense of obligation to follow and act on feedback (for example, London et al., 1997); 3) social awareness, an individual's sensitivity to how others' view them (for example, Levy et al., 1995); and 4) feedback self-efficacy, an individual's competence toward processing and responding to feedback (Renn & Fedor, 2001). These dimensions each represent a unique component of FO and have distinct relationships with other constructs such as the perceived benefits of self-development, intention to use feedback, self-monitoring and participation in developmental activities (Linderbaum & Levy, 2010).

London and Smither (2002) predicted that individuals high in FO would be more receptive to coaching and would be more likely to act on and improve their performance following feedback. Indeed, FO has been found to correlate with a variety of performance-related measures, characteristics and behaviours in the workplace that support these predictions. A recent meta-analysis (Katz et al., 2023) found that individuals with high FO are consistently more goal-oriented, more

1. There is a distinction in the literature between feedback on results (for example, scores) and feedback on behaviour (for example, "next time, try waiting longer after posing a question"). Throughout this paper feedback refers to the latter type of feedback.

likely to seek feedback regarding their job performance (see also Anseel et al., 2015; Cavanaugh, 2016; Dahling et al., 2012; Linderbaum & Levy, 2010), and perform better in their role. Individuals with high FO have greater satisfaction with feedback (Rasheed et al., 2015), greater career insight, and are more likely to participate in self-development (Cavanaugh, 2016). Employees with high FO are also more likely to be coachable (Weiss et al., 2023) and have better quality coaching relationships (Gregory & Levy, 2012; Katz et al., 2023). Interestingly, FO is also related to an individual's feedback environment (Katz et al., 2021); that is, when employees experience day-to-day feedback in a more positive context (for example receiving high quality feedback from a credible supervisor), individuals are more likely to be higher in FO, suggesting that FO may be influenced by the specific feedback culture in a workplace.

FO is theorised to be a somewhat malleable construct that can develop and improve over time, over multiple cycles of receiving, processing and using feedback (Linderbaum & Levy, 2010; London & Smither, 2002). Particularly, London and Smither (2002) predicted that it would develop in the long term (after approximately six to 12 months) as individuals experience feedback and are influenced by an organisation's feedback culture. To our knowledge, this prediction has not been empirically tested, although limited research suggests a certain stability. For example, two studies found similar levels of FO in different age groups (undergraduates and older full-time employees, Elsner Twesme et al., 2021; medical students and medical residents, Mills et al., 2023), suggesting that FO doesn't increase with experience – although critically, this was not based on longitudinal data. Whether FO changes over time is therefore an unanswered question, but one which may be relevant to training providers and employers, particularly if FO is important to individuals' PD.

How is feedback orientation relevant to teachers' professional development?

Research has demonstrated that novice teachers improve at different rates (Bardelli et al., 2023; Kraft et al., 2020; Kraft & Papay, 2014), however it is not fully understood why. For example, different training providers only account for a small proportion of the variance in teachers' initial effectiveness (Boyd et al., 2009; Goldhaber, 2019; Ronfeldt & Campbell, 2016) and rates of improvement (Bardelli et al., 2023). The evidence summarised earlier suggests that FO might be a good candidate for predicting how much novice teachers are likely to improve. Given the amount of feedback teachers receive during their training, and the consistent relationship between FO and job performance (Katz et al., 2023), differences in teachers' FO may go some way to explain their different rates of improvement. However, FO has not been researched much within the teaching profession.

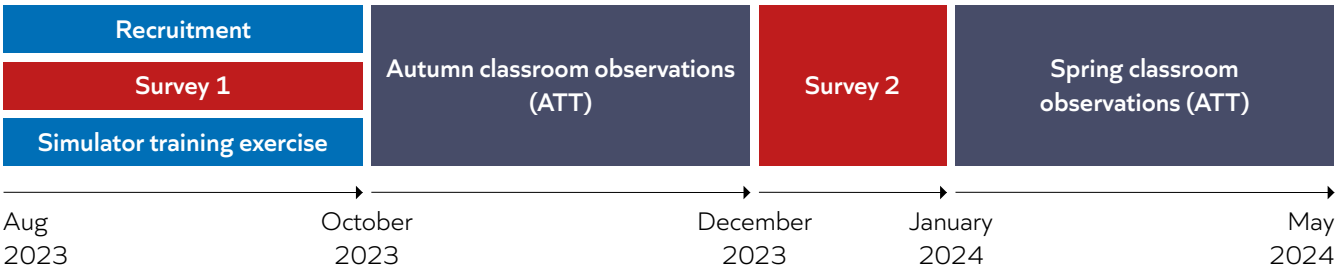
In the few studies that have investigated teachers' FO, the dimensions of utility and self-efficacy appear to be particularly relevant to the profession. Quintelier et al. (2020) investigated teachers' responses to inspection feedback in 80 Flemish primary schools. Teachers with higher feedback utility were more willing to use the inspection feedback, gained more awareness from it, and accepted it more readily. Feedback self-efficacy also influenced teachers' perceptions of the inspection feedback – teachers with higher self-efficacy viewed it as fairer and more relevant. Tuytens and Devos (2014) observed a small positive relationship between teachers' feedback utility and their engagement in professional learning activities following evaluation feedback, suggesting a relationship between teachers' FO and motivation to engage in PD. Finally, in a survey of 69 US teachers working in kindergarten to Grade 12 (ages 5-18), feedback self-efficacy and utility were found to predict teachers' job satisfaction and organisational commitment (Hemler, 2020). This small group of studies suggests that FO, and particularly feedback utility and self-efficacy, could play a role in multiple aspects of teachers' PD and experience. However, to our knowledge, FO has not been studied in relation to teachers' actual performance or improvement during PD (that is, how much they improve during and after their training).

To summarise, there are several reasons that FO may be an important characteristic for schools and teacher training providers to consider. First, understanding why some teachers' respond to and act on feedback better than others may help schools and training providers to use feedback more effectively (Linderbaum & Levy, 2010). Second, high FO may be a desirable quality in novice teachers, particularly if it is related to greater improvement in teaching skills. Indeed, principals often prefer new teachers who are coachable because they can adapt to the school's culture (Giersch & Dong, 2018; Harris et al., 2018). Finally, if FO can indeed improve over time, this could be an important characteristic to nurture in novice teachers. That is, fostering a positive and receptive approach to feedback in trainees (and indeed more experienced teachers) could in turn help teachers to improve and develop throughout their careers, and could also improve teachers' self-efficacy and satisfaction with their role.

The present study

We investigated the role of FO in novice teachers' development of teaching skills during the first two terms of their training (see Figure 1). Over this period, we conducted two surveys to measure novice teachers' FO as well as an online simulator training exercise. We also obtained classrooms observation ratings for participants training with Ark Teacher Training (ATT) from their first and second terms of training.

Figure 1. Timeline of the study



We tested the following hypotheses:

Hypothesis 1 (H1): FO (measured using the FOS; Linderbaum & Levy, 2010) would positively predict novice teachers' improvement in a simulator training exercise using a classroom simulator, over and above other individual measures which may also predict improvement (conscientiousness, working memory, task mastery and goal orientation (H1A). We also predicted that the utility and self-efficacy sub-dimensions of the FOS would be stronger predictors of participants' improvement in the training, compared to the awareness and accountability dimensions (H1B). We hypothesised that teachers' perceived utility of feedback would likely affect how carefully they attend to and process the feedback, and feedback self-efficacy would affect how well they are able to focus on learning from constructively critical feedback. By contrast, the awareness and accountability dimensions would be less important because participants would be given very explicit and direct feedback by a teacher educator in a targeted training exercise. Accountability would be created by the teacher educator choosing to target a specific area for feedback.

Hypothesis 2 (H2): FO measured at the start of the academic year will positively predict novice teachers' classroom observation scores from the end of trainees' first term (autumn 2023), and improvements in scores between the autumn 2023 and spring 2024 terms[2], over and above conscientiousness, working memory, and goal orientation (H2A). We predicted that the accountability dimension of the FOS would be a stronger predictor of improvements in classroom observation ratings, compared to the utility, awareness and efficacy dimensions (H2B). Our reasoning was that, unless participants hold themselves to account for acting on the feedback they receive, the feedback would not result in changes in practice, regardless of whether they become aware of it, consider it useful, or consider themselves able to use it.

Hypothesis 3 (H3): A concern with self-report measures is that trainees will report their FO to be higher than it really is. Improvement in a training exercise (which is highly dependent on processing and acting on feedback) may reflect a more objective measure of trainees' receptivity to feedback and propensity for self-development. We tested whether the simulator training exercise could provide credible information about which teachers would improve faster, based on their responses to feedback, even before they had started training. We hypothesised that novice teachers' improvements in the simulator training exercise would positively predict improvements in their classroom observation scores between the autumn 2023 and spring 2024 academic terms.

Hypothesis 4 (H4): FO measured at the end of trainees' first term (December 2023 – January 2024) will be higher than at the start of their academic year[3] (August – October 2023).

2. For H2 and H3, our pre-registration states that we would analyse improvements in classroom observation scores over the course of the autumn term. This should have stated improvements over the course of the autumn and spring terms as this is when observations were conducted.
3. Our pre-registration states that we would measure FO before the start of the academic year (August 2023), and again in December 2023. Due to slower recruitment than we expected, our baseline measure of FO was taken between August – October 2023, and the second measure between mid-December 2023 – mid-January 2024.

Methods

We pre-registered the study with As Predicted on 9 August 2023 (Registry ID: 140545, https://aspredicted.org/h994x.pdf). Pre-registrations, as well as all results and materials, are freely and openly available at https://osf.io/ct73e/. Ethical approval was granted from UCL IOE Research Ethics Committee on 28 June 2023, with registration number Z6364106/2023/06/54.

Participants

All participants were recruited from postgraduate initial teacher training programmes between August and October 2023, at the start of their first year of training. We initially recruited 103 trainees from ATT, with whom we partnered for this research. Due to lower participant numbers than expected, we then recruited another 103 participants from University College London (UCL) and seven other training providers (see Table 1). Participants gave informed consent as part of Survey 1. We gave them a £25 Amazon voucher on completion of both the survey and simulator training between 10 August and 31 October 2024. We gave a second £25 Amazon voucher to participants who completed Survey 2 between December 2023 and January 2024.

Following recruitment, 31 participants withdrew from the study and did not take part in the simulator training exercise. A further eight participants did not fully complete Survey 1, and three were excluded as outliers on at least one of the survey measures (see statistical analysis for details). This gave us an available sample of 164 novice teachers, however the final sample size varied for each analysis (see Table 1). Of the 164 trainees, 148 took part in the simulator training exercise (H1), where a further four were excluded due to incomplete data[4]. Classroom observation data (used for H2 and H3) was provided by ATT and so was only available for the 89 participants training with this provider[5], 74 of whom also took part in the simulator training (H3). 135 trainees fully completed both Survey 1 and Survey 2 and so were analysed for H4.

Study design

Overview

The study comprised several parts (see Figure 1): 1) two surveys to measure FO plus other psychological and feedback-related variables; 2) a simulator training exercise in behaviour management using a classroom simulator; and 3) analysis of classroom observations obtained as secondary data from ATT, from the trainees' school placements. The first survey and the training exercise formed part of a different experimental study; the methods are summarised below but for a detailed explanation, including for the training materials, see Banks et al. (2024) and https://osf.io/dwkb6.

Surveys

Survey 1 was conducted before participants took part in the simulator training exercise. The survey included our baseline measure of FO, as well as four psychological variables (see Measures). We aimed to collect the baseline measure of FO before trainees started their school placements. However, due to recruitment issues this was not possible, and so the timing of Survey 1 varied per participant. Dates ranged from 9 – 17 August 2023 for ATT participants (the survey was distributed during a recruitment event) and 14 September – 31 October 2023 for all other trainees[6].

4. One due to technical issues with the software, and three because the participants did not complete their simulator training session.
5. Note that we had intended to only recruit participants from ATT, who had agreed to share classroom observation data from their trainees. However, due to lower-than-expected recruitment we opened up the study to other providers. Trainees from these providers could participate in the simulator training and complete Surveys 1 and 2, but we could not obtain comparable classroom observation data from them due to different assessment procedures.
6. This discrepancy is addressed in the sensitivity analysis for H4.

We emailed Survey 2 to all participants who had completed Survey 1 on 14 December 2023, at the end of their first term of training. Completion dates ranged from 15 December 2023 – 8 January 2024, when the survey was closed. Survey 2 included the same FO questionnaire as well as three of the psychological measures. We also added the feedback environment (Steelman et al., 2004) and feedback seeking behaviour (VandeWalle et al., 2000) scales to explore relationships with feedback orientation and classroom observation ratings (Katz et al., 2021). We were aware that the amount of teaching experience our participants received during their placements would vary, even for participants training with the same providers. To capture this variance, we also asked trainees about the number of weeks of school placement they had completed, and what percentage of the time they had been teaching, since their course start date and since the date of their simulator training exercise[7].

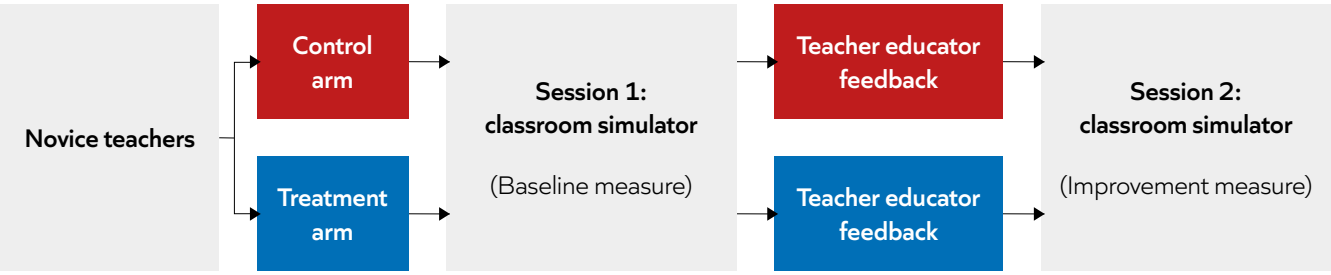
Table 1. Demographic characteristics of analytical sample for each hypothesis

	H1	H2	H3	H4
No. of participants	144	89†	74†	135
Gender				
Female	108	67	58	103
Male	36	22	16	32
Ethnicity				
Asian	26	14	13	22
Black	16	13	10	14
Mixed	18	8	8	17
White	84	54	43	82
Education level				
Undergraduate	90	61	49	84
Postgraduate	54	28	25	51
Experience as a teaching assistant				
Yes	61	50	36	59
No	83	39	38	76
School phase				
Primary	35	22	19	36
Secondary	108	67	55	98
Unknown	1	0	0	1
Training provider				
ATT	75	89	74	70
UCL	54	0	0	50
Other	15	0	0	15
Mean age in years (*SD*)	29.97 (8.99)	30.10 (9.08)	30.95 (9.51)	29.38 (8.42)

Note. Each sample analysed is a subset of the total 164 participants who took part in the study, that is, they are not independent groups. H = Hypothesis. ATT = Ark Teacher Training. UCL = University College London. † Sample sizes for H2 and H3 analyses of spring observation data were *N*=88 and *N*=73 respectively due to one participant withdrawing from the ATT training programme.

7. These questions were included for the purposes of a sensitivity analysis, but due to very poor and inconsistent data quality (likely due to participants misunderstanding the questions) we were not able to analyse responses accurately. We therefore could not control for the different amounts of teaching our participants experienced (see Limitations).

Figure 2. High-level overview of the online simulator training exercise



Simulator training exercise

The training exercise involved participants interacting with avatar pupils in a classroom simulator (Banks et al., 2024; Cohen et al., 2020; Sims et al., 2023). The avatar pupils were operated by a simulator specialist (who was not visible) and could interact in real time with the participants – for example, answering questions or responding to instructions (see https://osf.io/ct73e/ for a demo). The training was implemented within a Zoom online video call using Mursion software (Cohen et al., 2020; Ferguson & Sutphin, 2022). The training focused on positive behaviour management. Participants were first provided with instructions and a school behaviour policy, and were then asked to manage pupils' behaviour in the classroom simulator while transitioning from one activity to another. Participants took part in short classroom simulator sessions in which pupils misbehaved in a variety of ways, and they were required to minimise pupil time off task by managing this behaviour. After the first simulator session, participants received feedback from a teacher educator which focused on three specific behaviour management practices. Participants were given the opportunity to apply these three practices in the second simulator session. If the participants used the appropriate practice at the appropriate time, the pupils would comply. If they did not use the appropriate practice, or used it at the incorrect time, then the pupil would either ignore the teacher or argue back. The first and second simulator sessions presented participants with the same classroom scenario, as this allowed us to directly measure how well participants implemented the feedback. The full training session lasted approximately 30 minutes, including three simulator sessions and two sets of feedback. Figure 2 shows a simplified overview of the simulator and feedback sessions that were analysed in the present study (for the full training design see Banks et al., 2024).

One of two teacher educators provided feedback to participants after the first and second simulator sessions. Each set of feedback took approximately five minutes (see Banks et al., 2024 for full details). Participants were first given feedback about whether they used the three behaviour management practices correctly in the preceding classroom simulator. The teacher educator then explained how to use them correctly and modelled them to the participant within the simulator. Finally, the teacher educator gave the participants a chance to practice for themselves, also within the simulator. All participants experienced the same amount of feedback. However, the exact way in which the teacher educator delivered this differed depending on the experimental arm to which the participant had been allocated – this was related to a separate experimental study (Banks et al., 2024). Although we controlled for this in our analysis[8], it will not be discussed further in this paper

8. We controlled for any effects of the experimental arms in our analyses by calculating residual variables – see Statistical Analysis.

Measures

Survey measures

Feedback orientation. We used the FOS (Linderbaum & Levy, 2010) to measure feedback orientation. The FOS has been used in the majority of research into FO, and has been shown to be a reliable measure that generalises to different populations, for example predicting full and part-time employees' job performance (Dahling et al., 2012; Elsner Twesme et al., 2021, 2021; Rasheed et al., 2015) as well as university students' academic achievement (Abaci, 2014; Gjerde et al., 2022). The FOS consists of 20 items across four dimensions: utility, accountability, social awareness, and feedback self-efficacy. Participants responded to each statement using a five-point Likert scale (1 = "disagree strongly", 5 = "agree strongly"). We calculated participants' FO (and their score on each dimension) using a simple sum score of all items. This is how the scale has been used most often in the literature (as a predictor of job performance, for example) and would allow our results to be directly comparable to other professions.

For H4, which compared FO measured at Survey 1 and 2 as the dependent variable, we additionally calculated a weighted version [9] of FO using confirmatory factor analysis. This provides a principled way of weighting the contribution of each item to the overall score. It may thus enable the detection of subtle changes in participants' FO, and further allowed us to check for measurement invariance between the two timepoints. We conducted confirmatory factor analyses in R Studio using the lavaan package (version 0.6.17; Rosseel, 2012). We specified a two-factor model with each timepoint (Survey 1 and Survey 2) as latent factors, allowing for covariance between them and specifying fixed loadings for each item (for example, the loading for item 1 was the same for each survey). To check for measurement invariance, we compared this model to one with free-loading estimates between the timepoints; there was no significant difference between the models ($\chi^2 = 14.62$, $df = 19$, $p = .746$). We then multiplied each item score with its corresponding item factor loading to obtain the weighted measures of FO at each timepoint. To calculate weighted measures for each FO subscale, we followed the same procedure but the specified model included latent variables for each subscale, at each timepoint: all other parameters were identical.

Feedback environment. We included the Feedback Environment Scale (Steelman et al., 2004) in Survey 2, after participants' first school placement during the autumn term. The scale consists of seven subdimensions (credibility, quality, delivery, favourable feedback, unfavourable feedback, availability and promoting feedback seeking), and participants respond to questions in relation to feedback from their supervisors (32 items), and in relation to their coworkers (31 items)[10]. Supervisor in our context referred to a trainee's "in-school mentor, tutor or classroom teacher", while coworkers were defined as "other colleagues you work closely with in school, such as your department or phase"[11]. Participants responded using a seven-point Likert scale (1 = "strongly disagree", 7 = "strongly agree"). We calculated participants' overall feedback environment by summing responses to all questions. We also summed responses to the supervisor and coworker items separately to examine differences between these aspects of feedback environment.

Feedback seeking behaviour. We included a measure of feedback seeking behaviour in Survey 2, adapted from VandeWalle et al. (2000). Participants were asked "how frequently have you sought feedback from your supervisor?" regarding several aspects of their job role and performance. Responses were measured on a seven-point Likert scale (1 = "never", 7 = "daily"). As with feedback environment, we asked participants to respond to the same questions in relation to their supervisor and coworkers. We summed all participant responses to calculate overall feedback seeking behaviour.

Psychological measures. We included four psychological measures in Survey 1 and three in Survey 2. Conscientiousness (a person's propensity to be self-controlled, responsible, hardworking and orderly; Roberts et al., 2014) was measured using the Big Five Inventory-2 (BFI-2; Soto & John, 2017) on a five-point Likert scale and summed across all 12 items (1 = "strongly disagree", 5 = "strongly agree"). Goal orientation (an individual's preference for developing and demonstrating ability in achievement contexts; Dweck, 1986; Vandewalle, 1997) was measured using the revised Achievement Goal Questionnaire (AGQ-R; Elliot & Murayama, 2008).

9. Note that this additional measure was specified in the pre-registration.
10. One item from the scale is only relevant to supervisor feedback and so is not included for coworkers.
11. Phase refers to a key stage in a primary setting, such as Early Years or Key Stage 1.

Each of the four achievement goals (mastery approach, mastery avoid, performance approach and performance avoid) was measured on a five-point Likert scale (1 = "strongly disagree", 5 = "strongly agree"), and the three items for each subdomain were averaged. Task mastery (an individual's motivation to attempt and persist with challenging tasks) was measured using the Dimensions of Adult Mastery Motivation Questionnaire (DAMMQ; Doherty-Bigara & Gilmore, 2015). Four of the dimensions measuring task persistence, absorption, pleasure, and preference for challenge were measured using a five-point Likert scale (1 = "not at all typical", 5 = "very typical"), and all 24 items were summed. Working memory (short-term memory storage and manipulation; Baddeley & Hitch, 1974) was measured using an adapted computerised version of an n-back task from PsyToolkit (Stoet, 2010, 2017). Participants viewed a sequence of letters presented one-by-one on screen and were asked to decide whether they saw the same letter two trials ago (that is, $n = 2$ back). We calculated individual working memory as accuracy rate (hit rate – false alarm rate) according to standard signal detection theory. Full details of the measurements we used can be found in Appendix A.

Measures from the simulator training exercise

Behaviour management. First, we captured participants' skill in the three techniques they received feedback on, including whether they used the right technique at the right time in each simulator session. This gave a score of between zero and 12 for each session. Second, we estimated the total time that the avatar pupils spent off-task in each simulator session, in seconds. For both measures, we used an observation rubric applied to video recordings of participants completing the simulator training. Videos were independently coded by four researchers. After a process of feedback and correction, agreement between the four coders was excellent for both simulator measures and all behaviour management practices (intraclass correlation range: .93 – .98, all ps < .001). We measured the two behaviour management variables during the first (baseline) and second simulator sessions as this allowed us to analyse trainees' improvements following feedback.

As our training exercise involved an experimental manipulation, we used residual measures of participant skill and pupil time off task that were comparable between the two randomised groups. To calculate these, we conducted regression models predicting participants' performance based on the group to which they were randomly assigned, and then extracted the residual (unexplained) variance. These measures captured participants' performance excluding any influence of which group they were assigned to, therefore controlling for the experimental arm participants were assigned to.

Feedback satisfaction. Research suggests that initial reactions to feedback may influence whether recipients respond positively to it (Smither et al., 2005). We therefore captured participant's subjective reaction to the feedback they received following the first classroom simulator exercise. We used a modified version of the satisfaction with feedback questionnaire (Rasheed et al., 2015; Tonidandel et al., 2002) consisting of four items (for example, "I think my feedback exactly represents my output"). Participants rated their level of agreement on a five-point Likert scale (1 = "strongly disagree", 5 = "strongly agree"), and responses to all four items were summed to calculate overall feedback satisfaction score.

Classroom observation ratings. We obtained classroom observation data from 89 participants training with ATT. Observations were collected by trained observers based on Ark's Great Teacher Rubric, which covers five critical domains of teaching practice (climate for learning, planning and preparation, teaching and learning, assessing and responding, and professionalism). These are further divided into subdomains and key observation criteria (for example, a subdomain for climate for learning is managing behaviour, which is based on observations of setting high expectations, being positive and proactive, and using systems and techniques). For each observation, trainees are rated based on three possible outcomes: not yet attempting, attempting or foundational. Most trainees are expected to reach foundational level by the end of their first year. Observations were collected at the end of the autumn 2023 and spring 2024 terms.

Following Kraft et al. (2020) we combined the scores given across the different areas of the rubric using a graded response model. This allowed for differences in difficulty and interval scale across teaching domains, as well as differences in how well items differentiate between teachers with similar skill levels. This measure results in more variation in scores than averaging across domains [12].

12. A limitation of this method is that it doesn't account for measurement invariance between time points. However, as the spring data in particular had limited variance (see Limitations), it allowed us to conduct the planned analyses using observation ratings as a continuous variable.

To calculate graded measures for the autumn and spring ratings, we fitted a graded response model to each set of ratings using the mirt package in R Studio (version 1.41; Chalmers, 2012). We then extracted the theta score for each participant, which estimates their position along the latent variable captured by all items and domains. We included all subdomains and observation criteria except ratings for "phonics and early maths" (part of the "teaching and learning" domain) as the majority of trainees had not yet been assessed on this criterion at the time of autumn data collection.

Classroom climate binary measure. In addition to the main classroom observation ratings above, we analysed a binary "yes"/"no" measure of how well trainees managed their classroom climate. This was obtained through classroom observation carried out by ATT at the end of the first half of the autumn term, based on whether trainees were able to gain and maintain 100% pupil compliance for the first 15 minutes of their lesson. This data was also used to inform ATT providers whether to adopt an intervention (support plan) for trainees who had not embedded key aspects of training into their teaching practice.

Statistical analysis

Data preparation

We collated and processed all data using R (R Core Team, 2022) and JASP (JASP Team, 2023). Outliers for all variables were assessed using the interquartile range (IQR), using a threshold of 1.5 IQRs above/below the third/first quartile. Applying this criterion, we excluded three participants based on their responses to at least one of the survey measures. We identified one participant who had misinterpreted the response scale of the FOS – we asked this participant to redo the scale rather than excluding them. Several participants identified as outliers for the working memory task hadn't made a single response or had an accuracy of almost zero. As the task had been conducted online, we assumed that these participants had misunderstood the instructions. We provided extra guidance and asked them to redo the task. One participant's accuracy of 0.09 (from their second attempt) was subsequently excluded as an outlier, but all others were included.

Statistical models

We tested H1 (whether FO would predict trainees' performance in the simulator training exercise) using linear regression. For H1, FO was the main independent variable, and residual participant skill and pupil time off task from the second session of the simulator training were the dependent variables. This meant that we effectively controlled for the experimental arm that participants were assigned to, and did not need to include this in the regression models. We included participant skill and pupil time off task from the first classroom simulator session (baseline measures) as covariates. For H1A we added the four psychological measures (conscientiousness, goal orientation, task mastery and working memory) as covariates. To test H1B, we carried out separate regression analyses for each of the four sub-dimensions of the FOS (accountability, social awareness, utility and self-efficacy) and compared standardised coefficients and 95% confidence intervals (CI). We also plotted a binned scatterplot to check for a linear relationship between FO and each dependent variable (see Figure B1).

We tested H2 (whether FO would predict trainees' improvement in classroom observations) using linear regression. The dependent variable was the graded measure of participants' classroom observation ratings from the autumn 2023 or spring 2024 terms, and the independent variable was FO. We first analysed whether FO predicted classroom observations from the autumn term (with the autumn ratings as the dependent variable). Second, we analysed whether FO predicted improvements in ratings between autumn 2023 and spring 2024 by using the spring ratings as the dependent variable, and adding the autumn ratings as a covariate[13]. For H2A we added the four psychological measures as covariates, and for H2B we carried out separate regression analyses for each of the four sub-dimensions of the FOS and compared standardised coefficients and 95% CI.

We analysed H3 (whether improvements in the training exercise would predict improvements in classroom observations) using linear regression. The independent variables were the difference in participant skill and pupil time off task between the first and second simulator training sessions, and the dependent variables were classroom observation ratings from the autumn 2023 and spring 2024 terms. As in H2, we analysed whether FO predicted classroom observations from the autumn term, and whether FO predicted improvements in ratings between autumn 2023

13. The pre-registration stated in error that we would analyse improvements in data over the autumn term, however this should have stated autumn to spring as this is when the classroom observations were conducted.

and spring 2024 by using the spring ratings as the dependent variable, and adding the autumn ratings as a covariate.

In all analyses for H2 and H3, we included an additional binary covariate indicating whether teachers had received an intervention (support plan) from ATT (1 = yes, 0 = no). Support plans are implemented by ATT if the data at key points in the year suggest that a trainee has not embedded key aspects of training into their teaching practice, which may limit them making progress. A support plan runs for four weeks and gives trainees one or two action steps to embed in their teaching. It is supported by additional coaching, live feedback, and co-planning where necessary. Adding this covariate was a necessary deviation from the pre-registration as 42% of ATT participants in our study received an intervention during our measurement period, which would undoubtedly have an impact on how much they improved. To test our hypotheses, we therefore needed to account for the variance explained by the interventions[14] and examine whether FO explained additional variance over and above this additional support.

We tested H4 (whether FO improved over time) using two-sided, paired-sample *t*-tests to compare FO measured at Survey 1 and Survey 2. We completed a separate *t*-test for the global variable of FO and then for each subscale of the FOS. We also conducted separate analyses for the summed and weighted scores of FO and its subscales. We calculated Cohen's *d* to indicate effect sizes.

Results

Overview

Table 2 provides descriptive statistics for all measures. Trainees tended to rate themselves as relatively high in FO, with scores ranging from 60 – 100 (max. possible = 100), and fairly similarly in each of the four subdomains – that is, there wasn't one subdomain rated substantially higher than the others. Trainees rated the feedback environment of their first placement relatively positively, ranging from 259 – 438 (max. possible = 441). However, trainees' feedback seeking behaviour varied widely, ranging from 23 – 63 (max. possible = 70).

Participants in the simulator training exercise improved over the course of the training, increasing their behaviour management skills and decreasing pupil time off task following the first set of feedback, in line with the training being effective (see Banks et al., 2024 for full results and discussion). However, participants' satisfaction with the feedback they received in the training varied greatly. A bimodal distribution indicated that trainees were split between low and high satisfaction; for example, 28% participants rated their satisfaction as ≤ 7 out of a possible 20 (indicating low satisfaction), while 31% participants rated their satisfaction as ≥ 18 (indicating high satisfaction).

To further understand the psychometric properties of the scales used in Survey 1 and Survey 2, we computed the reliability of the scores as their internal consistency (see Table 2) using Cronbach's α (Cronbach, 1951). With the exception of FO accountability and FO social awareness, all survey measures had internal consistencies above the often-used threshold of Cronbach's α = .70 (Taber, 2018). In Survey 1, the FO subscales ranged from Cronbach's α = .62, 95% CI [.50; .71] for FO accountability, to .82, 95% CI [.77; .86] for FO self-efficacy. In Survey 2, the FO subscales ranged from Cronbach's α = .63, 95% CI [.52; .72] for FO accountability to .80, 95% CI [.74; .85] for FO self-efficacy. For the survey measures that were taken twice, we additionally computed the test-retest reliabilities. The correlation of scale scores over time was $r = .55$ for the FO scale and ranged from $.40 < r < .59$ for the subscales. In comparison, the test-retest reliability for task mastery was $r = .61$ and between $.36 < r < .47$ for the goal orientation subscales.

14. All results were the same when the intervention data was not included as a covariate, except that none of the FO subscales predicted classroom observations. All results with and without the intervention variable are freely available here: https://osf.io/ct73e/.

Table 2. Descriptive statistics for all measures

Measure	Min. possible	Max. possible	*M* (*SD*)	*N*	Cronbach's α [95% CI]
FO (Survey 1)	20	100	81.48 (7.24)	164	.80 [.75; .85]
FO (Survey 2)	20	100	82.49 (7.69)	135	.86 [.82; .89]
Feedback environment	63	441	359.70 (41.14)	135	.95 [.94; .96]
Feedback environment – supervisor	32	224	188.61 (25.60)	135	.94 [.93; .96]
Feedback environment – coworkers	31	217	171.09 (23.28)	135	.93 [.91; .95]
Feedback seeking behaviour	10	70	44.27 (10.03)	135	.87 [.84; .90]
Feedback seeking behaviour – supervisor	5	35	23.19 (4.96)	135	.78 [.71; .83]
Feedback seeking behaviour – coworker	5	35	21.07 (6.31)	135	.86 [.83; .90]
Simulator training measures					
Participant skill 1 (baseline)	0	12	2.24 (1.73)	144	N/A
Participant skill 2	0	12	6.53 (2.22)	144	N/A
Pupil time off task 1 (baseline)	N/A	N/A	81.23 (21.32)	144	N/A
Pupil time off task 2	N/A	N/A	53.95 (20.81)	144	N/A
Feedback satisfaction	4	20	12.55 (5.75)	143	.76 [.68; .82]
Psychological measures					
Conscientiousness	12	60	48.15 (6.94)	164	.82 [.78; .86]
GO – mastery approach (Survey 1)	3	15	13.29 (1.86)	164	.75 [.67; .82]
GO – mastery approach (Survey 2)	3	15	12.74 (1.94)	135	.69 [.59; .77]
GO – mastery avoidance (Survey 1)	3	15	11.20 (2.90)	164	.75 [.67; .81]
GO – mastery avoidance (Survey 2)	3	15	10.76 (2.55)	135	.68 [.57; .76]
GO – performance approach (Survey 1)	3	15	10.29 (2.62)	164	.82 [.77; .87]
GO – performance approach (Survey 2)	3	15	9.91 (2.77)	135	.88 [.83; .91]
GO – performance avoidance (Survey 1)	3	15	10.26 (3.11)	164	.90 [.86; .92]
GO – performance avoidance (Survey 2)	3	15	9.47 (3.03)	135	.90 [.87; .93]
Task mastery (Survey 1)	20	120	79.56 (9.42)	164	.88 [.85; .91]
Task mastery (Survey 2)	20	120	78.85 (10.07)	135	.90 [.88; .93]
Working memory	0	1	0.78 (0.17)	161	N/A
Classroom observation ratings					
Autumn ratings	N/A	N/A	-0.001 (0.94)	89	N/A
Spring ratings	N/A	N/A	-0.053 (0.94)	88	N/A

Note. To compute Cronbach's α, only data from $n = 135$ participants analysed for H4 were used to allow comparability between Survey 1 and Survey 2. Classroom observation ratings refer to the graded response variable rather than raw ratings (see Measures for details). FO = feedback orientation; GO = goal orientation. CI = confidence interval.

Hypothesis 1: Feedback orientation will predict trainees' performance in a simulator training exercise

Confirmatory analysis

FO did not significantly predict improvements in participant skill or pupil time off task in the simulator training exercise (H1; see Table 3), even when controlling for psychological covariates (H1A). We predicted that the utility and self-efficacy subdimensions would be stronger predictors than social awareness and accountability (H1B). Against our hypothesis, the 95% CI of the utility and self-efficacy subdimensions overlapped with those of the social awareness and accountability subdimensions for participant skill (Figure 3, top panel) and pupil time off task (Figure 3, bottom panel). Based on standardised coefficients, self-efficacy was the strongest predictor of participant skill, and social awareness was the strongest predictor of time off task. However, *none* of the subdimensions significantly predicted improvements in participant skill, based on a Bonferroni-corrected alpha level of $p < .05/4$ analyses = .0125 for each dependent variable.

Table 3. H1: Regression analyses of improvements in participant skill and pupil time off task between the first and second sessions of the simulator training exercise

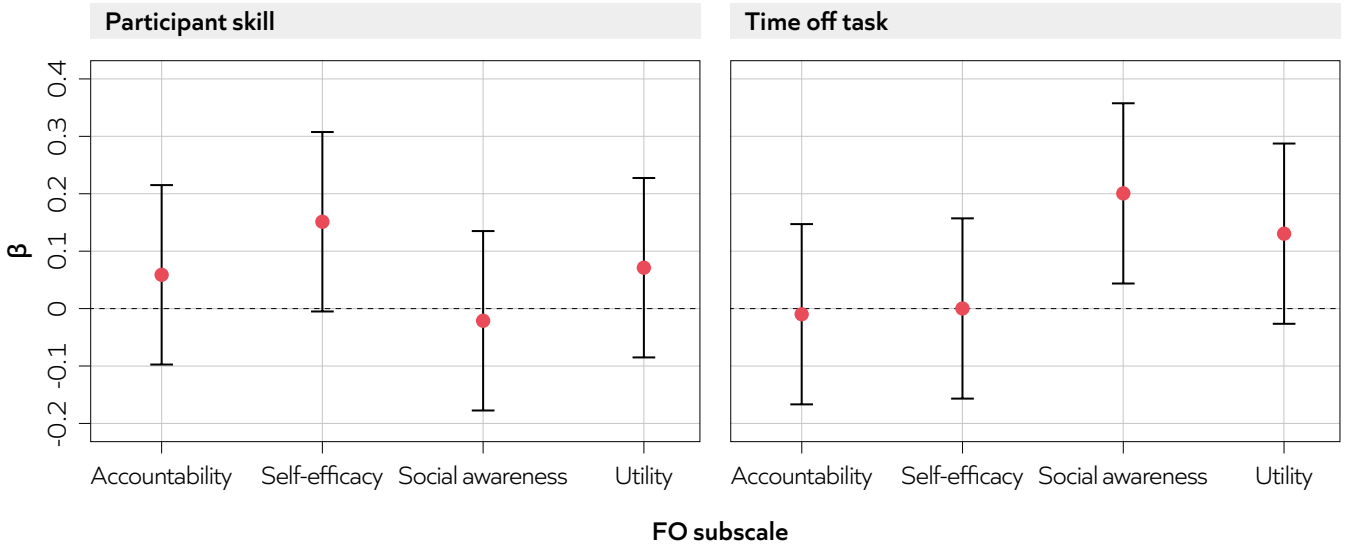
	Participant skill		Pupil time off task	
	β (*SE*)	*p*	β (*SE*)	*p*
H1 model	**R^2 = .15**		**R^2 = .12**	
FO	0.10 (0.08)	.207	0.12 (0.08)	.153
H1A model	**R^2 = .18**		**R^2 = .11**	
FO	0.15 (0.09)	.103	0.08 (0.09)	.390
Working memory	0.16 (0.08)	.066	-0.07 (0.09)	.459
Conscientiousness	0.13 (0.09)	.147	-0.15 (0.09)	.103
Task mastery	-0.10 (0.10)	.299	0.14 (0.10)	.140
GO – mastery approach	0.05 (0.09)	.611	-0.18 (0.09)	.065
GO – mastery avoidance	0.05 (0.09)	.612	0.03 (0.10)	.762
GO – performance approach	-0.19 (0.11)	.106	0.07 (0.12)	.540
GO – performance avoidance	0.08 (0.12)	.544	-0.04 (0.13)	.756
H1B models				
FO utility	0.07 (0.08)	.395	0.13 (0.08)	.109
FO accountability	0.07 (0.08)	.424	-0.01 (0.09)	.882
FO social awareness	-0.02 (0.08)	.785	0.20 (0.08)	.015
FO self-efficacy	0.16 (0.08)	.052	0.00 (0.09)	.993

Note. Each model refers to a separate regression analysis. All models include participants' baseline measure from the first simulator training session as a covariate. Participant skill and pupil time off task measures have been residualised to regress out the effects of an experimental manipulation (see Banks et al., 2024).

R^2 for H1B models of participant skill: FO utility =.11, accountability = .11, awareness = .11, self-efficacy = .13. R^2 for H1B models of time off task: FO utility =.05, accountability = .04, awareness = .08, self-efficacy = .04.

β (*SE*) indicates standardised coefficients and standard errors. FO = feedback orientation; GO = goal orientation.

Figure 3. H1B: Comparison of standardised regression coefficients of the FO subscales on participant skill and pupil time off task in the simulator training exercise



Note. Error bars represent 95% CIs. FO = feedback orientation. β = standardised coefficient.

Exploratory analysis

We explored whether participants' satisfaction with their training feedback mediated the relationship between FO and how much they improved. Initial reactions to feedback may influence whether individuals respond to it (Smither et al., 2005), and individuals are more likely to accept feedback if it is positive and matches their self-concept (Ilgen et al., 1979; Lechermeier & Fassnacht, 2018). Thus, trainees who were more satisfied with the feedback they received may have subsequently performed better, regardless of their FO. We conducted two path analyses in R Studio using the lavaan package (version 0.6.17; Rosseel, 2012). We tested whether feedback satisfaction (measured immediately after the first round of feedback received by trainees) mediated the effect of FO on 1) participant skill and 2) pupil time off task during the simulator training. We used the same regression models and variables as in H1, except that we included feedback satisfaction as a mediator between FO and the dependent variable (participant skill or time off task). Feedback satisfaction and FO were specified as latent variables. One participant did not complete the feedback satisfaction questionnaire due to a technical error, so the sample size for both analyses was $N = 143$. The full models and results are available at https://osf.io/ct73e/. Feedback satisfaction did not mediate the relationship between FO and participant skill ($\beta = -0.01$, $p = .816$) or pupil time off task ($\beta = -0.09$, $p = .108$). However, participants' FO predicted their feedback satisfaction (path model of participant skill: $\beta = 0.44$, $p = .002$; path model of time off task: $\beta = 0.43$, $p = .002$). Participants with higher FO were more satisfied with the feedback they received during the training.

Hypothesis 2: Feedback orientation will predict trainees' improvement in classroom observation ratings between autumn and spring terms

Confirmatory analysis

FO did not significantly predict ratings from either term (H2), even when controlling for psychological covariates (H2A; see Table 4). We predicted that the feedback accountability subdimension would be a stronger predictor of improvement in classroom observation ratings compared to the utility, awareness and efficacy subdimensions (H2B). Against our hypothesis, social awareness had the largest coefficient for the spring ratings. The coefficient for accountability was close to zero and its 95% CI overlapped with all other subdimensions for the spring term (Figure 4, bottom panel). Based on a Bonferroni-corrected alpha level ($p < .05/4$ analyses $= .0125$ for each dependent variable), only the social awareness subscale of FO predicted trainees' ratings, but only in the autumn term – it did not predict improvements between then and the spring term.

Table 4. H2: Regression analyses of trainees' classroom observation ratings from autumn and spring terms

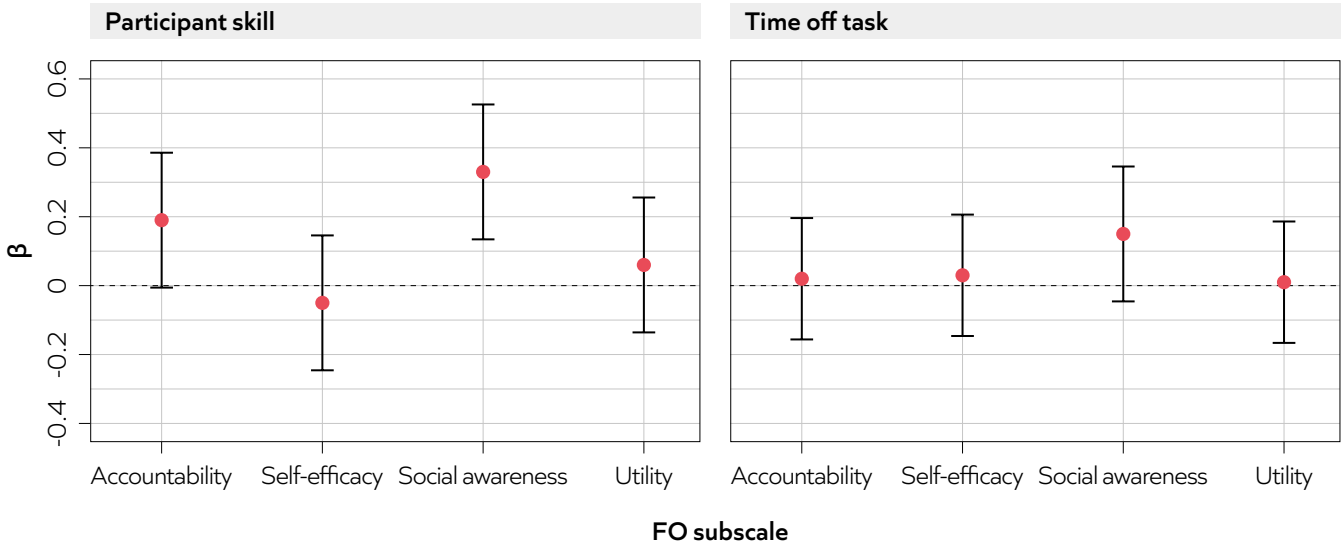
	Autumn classroom observation ratings		Spring classroom observation ratings	
	β (*SE*)	*p*	β (*SE*)	*p*
H2 model	**R^2 = .18**		**R^2 = .34**	
FO	0.18 (0.10)	.068	0.07 (0.09)	.435
H2A model	**R^2 = .25**		**R^2 = .38**	
FO	0.21 (0.11)	.090	0.06 (0.11)	.571
Working memory	0.19 (0.10)	.062	0.16 (0.09)	.090
Conscientiousness	-0.14 (0.12)	.244	0.03 (0.11)	.787
Task mastery	0.08 (0.12)	.500	-0.05 (0.11)	.626
GO – mastery approach	0.07 (0.18)	.674	0.13 (0.18)	.448
GO – mastery avoidance	-0.12 (0.12)	.320	0.02 (0.11)	.889
GO – performance approach	0.08 (0.17)	.644	0.08 (0.15)	.609
GO – performance avoidance	0.05 (0.18)	.784	-0.08 (0.16)	.606
H2B models				
FO utility	0.06 (0.09)	.918	0.04 (0.08)	.674
FO accountability	0.20 (0.10)	.054	0.03 (0.09)	.794
FO social awareness	0.32 (0.10)	.001	0.14 (0.10)	.139
FO self-efficacy	-0.05 (0.10)	.649	0.03 (0.09)	.754

Note. Each model refers to a separate regression analysis. All analyses include a covariate indicating whether trainees received an intervention from ATT. Analyses for the spring observations include the autumn ratings as a covariate, and were conducted on a smaller sample due to one participant withdrawing from the ATT programme (autumn: *N*=89, spring: *N*=88).

R^2 for H3B models of autumn ratings: FO utility =.15, accountability = .19, awareness = .24, self-efficacy = .15. R^2 for H3B models of spring ratings: FO utility =.33, accountability = .33, awareness = .35, self-efficacy = .33.

β (*SE*) indicates standardised coefficients and standard errors. FO = feedback orientation; GO = goal orientation.

Figure 4. H2B: Comparison of the standardised regression coefficients of the FO subscales on Autumn and Spring term classroom observation ratings



Note. β = standardised coefficient, error bars represent 95% CIs. FO = feedback orientation.

Exploratory analysis

First, we explored subsets of the classroom observation ratings from autumn 2023 and spring 2024 to see if FO and its subscales predicted individual domains of the rubric. As before, we fitted a graded response model (see Measures) to each subset of the data to create a single variable for each of four key domains from the autumn and spring terms: managing the climate for learning, planning and preparation, cognitive engagement, and assessing and responding. For each of these domains, we conducted a separate linear regression for the autumn and spring ratings. Regression models were identical to those conducted for H2, except that the dependent variable was one of the four key domains from the rubric listed above, and for the spring data we included the corresponding domain from the autumn ratings as a covariate. To account for multiple exploratory analyses, we used a Bonferroni-corrected alpha level of $p < .05/4$ analyses = .0125 for each dataset (autumn or spring). FO did not significantly predict any of the individual domains of classroom observations from the autumn or spring terms.

Second, we examined whether the feedback environment during trainees' school placements, and their feedback seeking behaviour, predicted their classroom observation ratings in the autumn term as well as improvements in ratings between then and spring. We again conducted the same linear regressions as in H2, with classroom observation ratings (as a single variable spanning all domains) from autumn or spring as the dependent variable. However, the independent variables were feedback environment and feedback seeking behaviour. We then conducted separate regression analyses with the supervisor and coworker subscales of feedback seeking behaviour and feedback environment, and the seven subscales of feedback environment relating to trainees' supervisors, as predictors. We used a Bonferroni-corrected alpha level of $p < .05/9$ analyses = .006 for each dependent variable. The full models and results are available at https://osf.io/ct73e/. Neither feedback seeking behaviour nor feedback environment predicted trainees' classroom observation ratings from the autumn term, nor did they predict improvements between the autumn and spring terms. This was also true when the scales were analysed separately in relation to trainees' supervisors and their coworkers, and when we examined the seven subscales of feedback environment relating to trainees' supervisors.

Hypothesis 3: Trainees' improvements in the simulator training exercise will predict their improvements in classroom observation ratings between autumn and spring terms

Confirmatory analysis

Improvement in participant skill and pupil time off task between the first and second simulator tasks of the simulator training exercise did not significantly predict classroom observation ratings from the autumn or spring terms (see Table 5).

Exploratory analysis

We also analysed whether teachers' performance in the simulator training exercise predicted an additional binary "classroom climate" measure. This measure specifically rated ATT trainees on their behaviour management and was measured halfway through the autumn term. It may therefore be more directly related (and closer in time) to the classroom simulator training exercise. Logistic regression indicated that neither improvements between the first and second simulator tasks in participant skill (OR = 0.88, Wald $\chi^2 = 0.24$, $p = .625$), nor in pupil time off task (OR = 0.76, Wald $\chi^2 = 0.94$, $p = .331$), predicted how well trainees managed their classroom climate.

Table 5. H3: Regression analyses of classroom observation ratings predicted by performance in the simulator training exercise

Simulator training exercise measure	Autumn classroom observation ratings			Spring classroom observation ratings		
	β (*SE*)	*p*	R^2	β (*SE*)	*p*	R^2
Participant skill	0.19 (0.12)	.114	.16	-0.11 (0.10)	.280	.35
Pupil time off task	0.22 (0.12)	.052	.077	-0.01 (0.11)	.926	.34

Note. All analyses include a covariate indicating whether trainees received an intervention from ATT. Analyses for the spring observations include the autumn ratings as a covariate, and were conducted on a smaller sample due to one participant withdrawing from the ATT programme (autumn: *N*=74, spring: *N*=73). β (*SE*) indicates standardised coefficients and standard errors.

Hypothesis 4: Trainees' feedback orientation will improve over their first term of training

Confirmatory analysis

We first analysed changes in FO using the standard summed variables. FO overall did not significantly improve over time although a trend was present (mean improvement = 1.18, *SD* = 6.95; see Table 6 and Figure 5). The only FO subdimension which significantly improved was FO self-efficacy (mean improvement = 1.09, *SD* = 2.82). As per the pre-registration, we also analysed the weighted variables. We found that FO did significantly improve over time (mean improvement = 1.38, *SD* = 7.05). As before, the subdimension of FO self-efficacy was the only subdimension which significantly improved (mean improvement = 1.05, *SD* = 2.67).

Sensitivity analysis

As the time between Survey 1 and 2 varied for individual participants, we analysed whether this impacted our results for H4. We conducted a regression analysis predicting the Survey 2 measure of FO using the Survey 1 measure, the time interval between the surveys and the interaction between both terms (for all FO measures using summed and weighted scores; see Table B1). No significant effect was observed for the time interval or interaction term (all *p*s > .05), indicating that the time interval between surveys did not impact how much participants improved in their feedback orientation.

Exploratory analysis

The literature demonstrates a reliable, positive relationship between FO and an individual's feedback environment, and with their feedback seeking behaviour (Katz et al., 2021). We investigated whether this was true of our trainees, who may encounter very different feedback environments in their school placements. We conducted regression analyses to explore whether the improvements we observed in FO (the overall measure and the self-efficacy subscale) could be predicted by trainees' feedback environment and feedback seeking behaviour, in relation to their supervisors and coworkers. We conducted separate regression analyses using the weighted version of FO and the self-efficacy subscale from Survey 2 as dependent variables, with weighted FO from Survey 1 as a covariate. We specifically analysed self-efficacy as this was the only FO subscale which showed a significant improvement between Survey 1 and 2. Feedback environment (and each of its seven subscales) and feedback seeking behaviour were the independent variables. We also analysed the feedback environment and feedback seeking variables in relation to participants' supervisors and coworkers separately. To account for multiple analyses, we used a Bonferroni-corrected significance level of $p < .05/9$ analyses = .006 for each dependent variable. Significant results (with standardised coefficients) are summarised here, but see Table B2 for full models and results.

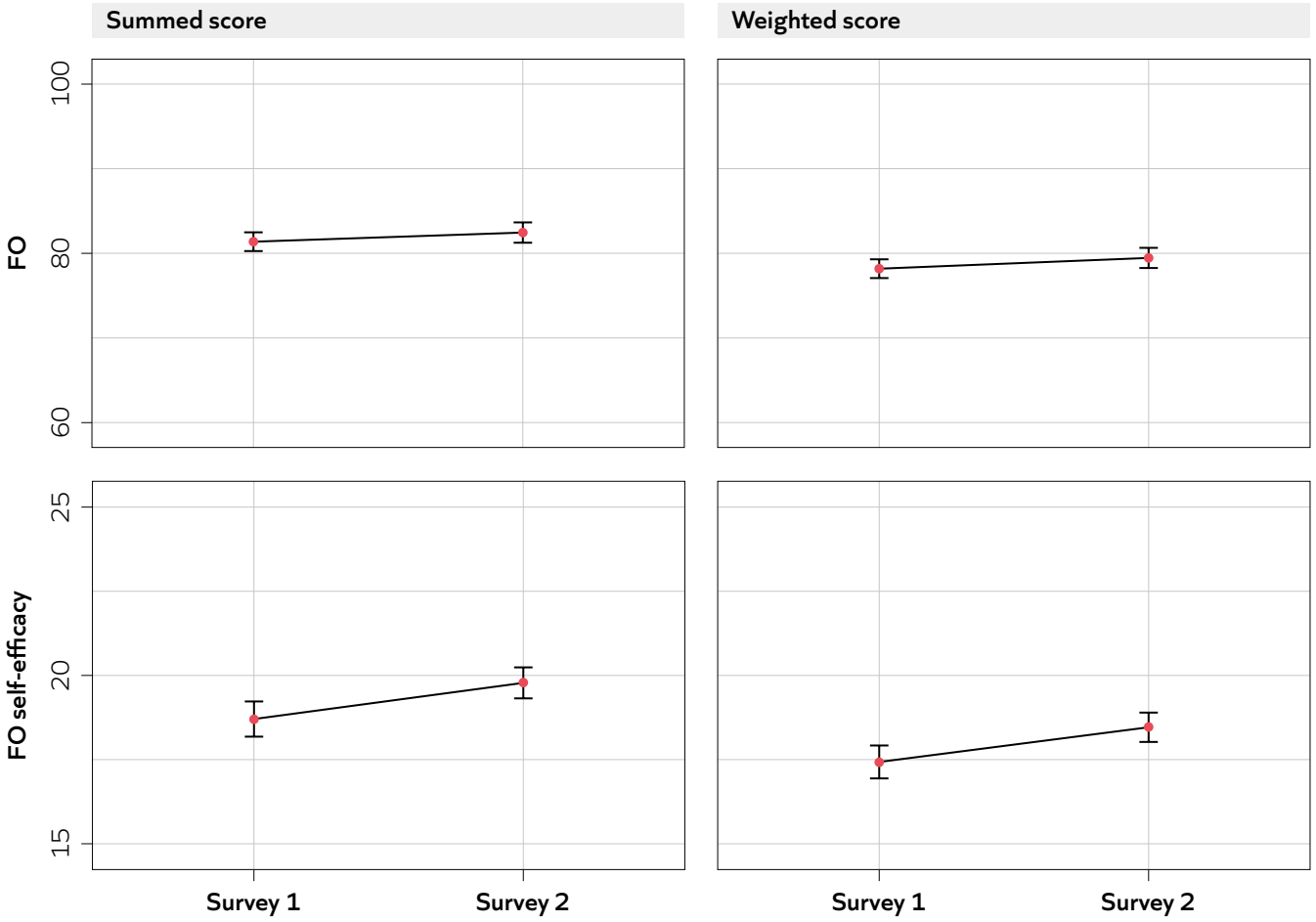
Feedback environment predicted trainees' improvement in FO ($\beta = .29$, $p < .001$), but feedback seeking behaviour did not ($\beta = .08$, $p = .255$). The feedback environment from trainees' supervisors, but not from their coworkers, predicted trainees' improvements in FO ($\beta = .24$, $p = .004$). Four subscales of supervisor feedback environment predicted FO overall: supervisors' credibility ($\beta = .27$, $p < .001$), the quality of their feedback ($\beta = .25$, $p < .001$), the extent to which they provided favourable feedback ($\beta = .33$, $p < .001$), and whether they promoted feedback seeking ($\beta = .21$, $p = .004$). Supervisor feedback also predicted the improvements in FO self-efficacy ($\beta = .28$, $p < .001$). Of the feedback environment subscales, supervisors' credibility ($\beta = .32$, $p < .001$), the quality of their feedback ($\beta = .28$, $p < .001$), the delivery of their feedback ($\beta = .27$, $p < .001$), the extent to which they provided favourable feedback ($\beta = .37$, $p < .001$), and whether they promoted feedback seeking ($\beta = .23$, $p = .001$) all predicted FO self-efficacy.

Table 6. H4: Paired *t*-tests comparing trainees' FO measured in Survey 1 and Survey 2

	FO summed			FO weighted		
	t (df)	***p***	**Cohen's *d***	***t (df)***	***p***	**Cohen's *d***
FO	1.97 (134)	.051	0.17	2.28 (134)	.024	0.20
FO utility	-0.42 (134)	.675	-0.04	-0.17 (134)	.865	-0.02
FO accountability	0.59 (134)	.553	0.05	0.64 (134)	.524	0.06
FO social awareness	0.27 (134)	.789	0.02	0.48 (134)	.635	0.04
FO self-efficacy	4.48 (134)	<.001	0.39	4.55 (134)	<.001	0.39

Note. FO = feedback orientation.

Figure 5. H4: Changes in summed and weighted FO and FO self-efficacy scores between the start (Survey 1) and end (Survey 2) of trainees' first term of training



Note. FO = feedback orientation. Error bars show 95% CIs. For changes in the other FO subscale scores, see Figure B2 in the Appendix.

Discussion

The present study investigated whether FO is relevant to novice teachers' training and development. We investigated whether it plays a role in how well novice teachers can implement feedback during a simulator training exercise, and whether it can explain why some teachers improve more than others over the course of their training. We also examined whether novice teachers become more responsive to feedback during the first term of their training – that is, whether their FO improves over a period when they likely receive a large amount of feedback from training providers and their school placements.

Did feedback orientation positively predict novice teachers' improvement in a classroom simulator training exercise? (Hypothesis 1)

Contrary to H1, FO did not predict novice teachers' improvement in behaviour management during the simulator training exercise. Neither the full FO scale, nor any of its subscales, consistently predicted teachers' improvement in the classroom simulator exercise. This is in contrast to the original theory of London and Smither (2002) and evidence from the literature, which has demonstrated that higher FO is consistently related to better job performance in a variety of sectors (Katz et al., 2021), and can predict university students' academic achievement (Abaci, 2014; Gjerde et al., 2022). Even though our participants were given explicit feedback about their behaviour management during the training, and a teacher educator modelled and explained strategies to help improve their skills, those with higher FO do not appear to have responded better to the coaching and feedback. We propose two possible reasons for this null finding. First, as originally theorised by London and Smither (2002), FO may be more relevant to how individuals improve over multiple cycles of feedback. In our study, teachers' improvement was measured following only one set of feedback in a single simulator training exercise that lasted 30 minutes. In comparison, Abaci (2014) and Gjerde (2022) observed a positive relationship between college students' FO and their academic performance in courses each lasting a full academic semester. FO may therefore be more relevant to learning over a longer period, with multiple cycles of feedback (London & Smither, 2002). Second, our sample of novice teachers were all relatively high in FO, and this lack of variance may explain the null finding here (this is discussed in more detail later). Nevertheless, we found that participants who were higher in FO were more satisfied with the feedback they received, which is in line with existing literature (Rasheed et al., 2015). This suggests that the former interpretation is more likely.

Did feedback orientation positively predict novice teachers' improvement in classroom observation ratings between autumn 2023 and spring 2024 terms? (Hypothesis 2)

Contrary to H2, the overall measure of FO did not predict improvements in classroom observation ratings between autumn and spring, and nor did any of the FO subscales. Also contrary to H2, the social awareness subscale predicted classroom observations but only for the autumn term – that is, it did not predict how much teachers improved. Further exploratory analyses did not reveal any potential relationships between FO and specific domains of teaching (for example, managing classroom climate, or planning and preparation). These results are again in contrast to studies from other professions, where FO has proved a reliable measure for predicting employees' job performance (Dahling et al., 2012; Elsner Twesme et al., 2021, 2021; Rasheed et al., 2015). This may be due to the complexity of teaching as a career, and the fact that feedback is one of many facets which influence teacher performance. Teachers may also be influenced by feedback from pupils in the form of summative assessment, while overarching directives from school or trust leaders may impact on the application of individual feedback, such as if a school implements a new policy to address challenging behaviour. In addition to these factors, our novice teachers will likely have received structured feedback and support from their coach or mentor to help them improve, for example in the form of interventions. For novice teachers at the start of their training, the level of feedback and support they receive may impact them more than their individual FO. However, these interpretations would need to be explored in future research.

A second explanation of our null finding may be the subjective nature of FO (and the scale we used to measure it), particularly that it could be subject to social desirability bias. All of our trainees tended to respond quite positively to the FOS, suggesting that it did not fully capture the variance of the underlying construct. Indeed, the average item rating in our sample was 4.23 (*SD* = 0.55), corresponding to the "mostly agree" category, while 20% participants had an average rating of 5 (that is, they responded "strongly agree" to most or all items). The lack of variance in our FO measure may therefore have contributed to the null findings reported here. Our item ratings for FO were indeed higher than those found in the original validation study for the FOS (*M* = 3.79; Linderbaum & Levy, 2010) and compared to a study of 69 public school teachers in the USA (*M* = 3.82; Hemler, 2020). A possible explanation of this difference is that the recruitment process of our teacher training providers might have favoured candidates high in FO. Alternatively, in a sample of novice teachers who are eager to make a good impression, or who are not yet aware of how they will respond to feedback in a teaching context, the FOS may not be a sensitive or objective enough measure to detect individual differences in FO. Future studies of individual differences in novice teachers may need to take into account the homogeneity and particular characteristics of this population.

Nevertheless, we unexpectedly found that novice teachers with greater social awareness of feedback had better classroom ratings during the autumn 2023 terms. The social awareness dimension of FO indicates an individual's sensitivity to how others view them (for example, Levy et al., 1995), and includes items related to making a good impression and an individual's awareness of what other people think of them. The finding may be explained by considering that novice teachers develop their teacher persona partly by observing those around them and thinking about how they are perceived. Developing a teacher persona or identity is an important part of the first few terms of teaching, both for career changing novice teachers and recent graduates (McLean, 1998). Novice teachers who are most aware of the way they are perceived are likely to want to respond to feedback to create positive perceptions of themselves. As this was an exploratory finding, it may warrant further investigation, particularly to help providers of initial teacher training support novice teachers to improve.

Did novice teachers' improvements in the simulator training exercise positively predict improvements in classroom observation ratings between autumn 2023 and spring 2024? (Hypothesis 3)

Contrary to H3, the extent to which teachers improved in the online simulator training did not predict improvements in their teaching skills between the autumn 2023 and spring 2024 terms. This indicates that a single training exercise dependent on feedback is not a good indicator of how much novice teachers are likely to improve during their training. It is worth noting here that none of the other psychological (conscientiousness, task mastery, goal orientation, working memory), behavioural (feedback seeking) or environmental (feedback environment) measures predicted improvements in novice teachers' classroom teaching. Hence, the reasons for which some teachers improve faster than others remain unclear, and are likely to involve factors not measured in this study.

Was feedback orientation measured at the end of novice teachers' first term higher than at the start of the academic year? (Hypothesis 4)

Theory has suggested that FO should improve following multiple cycles of feedback (Linderbaum & Levy, 2010; London & Smither, 2002), and in support of this we found that novice teachers' FO improved over their first term of training. However, this change was consistent only for the subdimension of feedback self-efficacy, which significantly improved in both the summed and weighted versions of participant scores. Feedback self-efficacy refers to how competent and confident individuals feel when processing and responding to feedback (Renn & Fedor, 2001), for example "I feel confident when responding to both positive and negative feedback". Improvements in this subdomain suggest that trainees felt more self-assured with regards to dealing with feedback. Interestingly, we found that this improvement was consistently predicted by the credibility and quality of the feedback that teachers received from their coach or mentor, as well as the extent to which they provided favourable (positive) feedback. Although this finding was exploratory, it is supported by existing literature which has shown a consistent link between feedback environment and feedback orientation (Katz et al., 2021). Our finding suggests that feedback from a coach or mentor may be important to build novice teachers' confidence and ability in responding to feedback. This may be an area that researchers or training providers wish to explore further, particularly if providing high quality feedback contributes to teachers' feedback self-efficacy.

Implications

The findings have several implications for practice and research. First, FO may not be a reliable measure for predicting improvement in teaching as it has been shown to be in other professions. This may reflect the complexity of teaching as a career, and the fact that feedback is only one of many facets which influence teacher performance. Particularly for novice teachers, other factors such as the support they receive from their schools and training providers may be more important than their FO. Our null results indicate a difference between the theory and its application to initial teacher training. Consequently, we would not recommend that training providers use FO as a tool to predict teachers' potential to improve or as a performance indicator. In general, when considering the use of any validated scale, teacher educators should consider whether there are studies which demonstrate the scale's effectiveness for novice teachers before using it.

Second, our findings show that FO can and does improve over time, similar to other personality traits (see also Jach et al., 2023). We found that teachers' self-efficacy regarding feedback improved over just one term of teaching. This suggests that coaching and feedback that teachers received during this time supported them to feel more confident and self-assured in dealing with feedback. While this did not influence their overall improvement in classroom teaching, it can still be seen as a positive outcome amid the challenges of recruitment and retention in teaching, especially as FO has been linked to teacher job satisfaction (Hemler, 2020).

Third, our exploratory findings point to the importance of an organisation's feedback environment. Teachers who experienced high quality and positive feedback from their coach or mentor had the largest increase in their self-efficacy regarding feedback. The quality and delivery of feedback from teachers' coaches or mentors may therefore be an area that training providers wish to focus on in the future. To reiterate, improving self-efficacy regarding feedback may not directly improve teaching, but it may have wider benefits, perhaps in terms of teachers' job satisfaction (Hemler, 2020).

Lastly, our results suggest several directions for future research. It is still unclear why some novice teachers improve faster than others. The null findings here suggest that the psychological and feedback-related constructs we tested are not potential candidates, at least for teachers in the initial stages of their training. Nevertheless, it remains important to understand who is likely to become a good teacher, and which skills or traits teacher educators should focus on to help trainees become better, faster. The present results indicate that trainees' immediate reactions to feedback (such as how satisfied they are with it), their social awareness, and the quality of feedback provided by their coaches and mentors may warrant further investigation. Researchers should continue to test alternative hypotheses to understand why some teachers improve more than others.

Limitations

Several limitations to our study should be noted. First, there were some limitations to the classroom observation data we analysed. The ATT rubric only uses three levels of ratings for trainees in their first year, and the aim is to raise all trainees' performance to the top level by the end of the year (indeed, a high proportion of participants training with ATT received interventions for this reason). The data therefore had limited variance, particularly for the spring. This was mitigated by using a graded response model; however, a limitation of this method is that a model is fitted to each individual dataset (that is, the model is guided by the data), meaning that items may have been weighted slightly differently for each time point. This limitation and the lack of variance may have contributed to the null findings reported here, particularly in relation to improvements in classroom teaching between the autumn and spring terms. Additionally, classroom observations in general, which are used to support mentors in making judgements for the ATT rubric, have limitations. By their nature, they can only capture a point-in-time measure of novice teacher performance in the classroom which may not represent the performance of the novice teacher in their unobserved lessons throughout the week (Ho & Kane, 2013). Additionally, judgements are subjectively made by the mentor with scaffolding provided by the rubric. This is mitigated by ATT staff providing quality assurance over this process, but cannot be mitigated entirely due to the nature of the measure. Future research examining improvements in novice teachers might consider measuring their performance in a variety of ways,

such as using multiple observers for observations, student achievement data (with appropriate controls) and student surveys (T. J. Kane et al., 2013).

Second, the amount of teaching experience participants gained before we conducted Survey 1 and the simulator training exercise varied. This was due to low recruitment numbers meaning we had to recruit participants from multiple training providers, over a longer period than intended. Unfortunately, due to participants misunderstanding the questions, the data we collected in Survey 2 was not reliable enough to conduct a sensitivity analysis to assess the impact of these differences. The differing amounts of experience may therefore partly explain some of the null results observed for the training exercise – that is, teaching experience may account for more variance than FO, but unfortunately we were unable to measure this accurately. Future studies of novice teachers should take into account the different amount of teaching they experience during their training.

Lastly, the improvements in FO that we observed were relatively small and showed substantial variance (see Figure B2). This may be due to the short time interval between data collection for Survey 1 and 2; we do not know if greater improvements would be visible over a longer time period, or if the change in FO would be stable if measured later in the year. Furthermore, the novice teachers in our sample already started with comparatively high average FO as reported in Survey 1, and as discussed earlier, this was higher than in previous studies that measured FO in teachers. This could have further limited the extent to which FO increased over time. The psychometric properties of the FOS are also important to consider. While estimates of internal consistencies passed the conventional cut-off, higher values of Cronbach's α are desirable when scores are to be assigned to an individual (Cronbach, 1951). As such, the FOS may not have the psychometric properties to reliably measure smaller-scale changes. To the best of our knowledge, no data is available yet on the standard error of measurement for the FOS. We were hence unable to determine whether the change observed in our sample was larger than what could be attributed to measurement errors. The test-retest reliability indicated only a moderate correlation between FO measured in Survey 1 and 2. It is unclear why the correlation coefficient is not higher, but this might have been caused by measurement errors affecting the scores at both surveys. Alternatively, FO may have indeed changed over time but the change itself is subject to large interindividual differences (see Figure B2). Although beyond the scope of this study, more research on FO is required to understand how the psychometric properties of the FOS may have affected our results. Future research should also confirm whether improvements in FO are long-lasting, and whether greater improvements can be observed over longer timeframes.

Conclusion

The present study is the first empirical investigation of FO in novice teachers. We found that FO has limited benefit as an indicator of teachers' improvement during their initial training, despite the large amount of feedback that novice teachers receive during this time. Our results provide the first evidence that FO can improve over a relatively short period of teacher training. Particularly, teachers' self-efficacy in responding to feedback seems to improve. Nevertheless, we observed much individual variation, and more research is needed to determine the role of feedback from coaches and mentors in developing positive responses to feedback in novice teachers. Overall, the question of why some teachers improve more than others remains largely unanswered by the present findings. Future research should investigate the extent to which novice teachers' social awareness impacts the quality of their classroom teaching, and how their feedback environment influences how effectively they respond to it.

References

Abaci, S. (2014). *Direct and indirect effects of feedback, feedback orientation, and goal orientations on students' academic performance in online learning* (3670284) [Doctoral dissertation, Indiana University]. ProQuest Dissertations and Theses. https://www.proquest.com/openview/9488350d56dfd9935a5ec9ea331aad01/1

Anseel, F., Beatty, A. S., Shen, W., Lievens, F., & Sackett, P. R. (2015). How Are We Doing After 30 Years? A Meta-Analytic Review of the Antecedents and Outcomes of Feedback-Seeking Behavior. *Journal of Management, 41*(1), 318–348.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press.

Banks, B., Sims, S., Curran, J., Meliss, S., Chowdhury, N., Altunbas, H. G., Alexandri, N., MacTavish, L., & Instone, I. (2024). *Decomposition and recomposition: Effects on novice teachers' enactment and transfer of behaviour management practices.* Ambition Institute. https://www.ambition.org.uk/research-decomposition-and-recomposition/

Bardelli, E., Ronfeldt, M., & Papay, J. P. (2023). Teacher Preparation Programs and Graduates' Growth in Instructional Effectiveness. *American Educational Research Journal, 60*(1), 183–216.

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher Preparation and Student Achievement. *Educational Evaluation and Policy Analysis, 31*(4), 416–440.

Brett, J. F., & Atwater, L. E. (2001). 360° feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology, 86*(5), 930–942.

Cavanaugh, C. (2016). *Beyond Cheerleaders and Checklists: The effects of the feedback environment on employee self-development* (akron1468076228) [Doctoral dissertation, University of Akron]. OhioLINK Electronic Theses and Dissertations Center. http://rave.ohiolink.edu/etdc/view?acc_num=akron1468076228

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software, 48*, 1–29.

Cherasaro, T. L., Brodersen, R. M., Reale, M. L., & Yanoski, D. C. (2016). *Teachers' Responses to Feedback from Evaluators: What Feedback Characteristics Matter? REL 2017-190.* Regional Educational Laboratory Central. https://eric.ed.gov/?id=ED570162

Cohen, J., Wong, V., Krishnamachari, A., & Berlin, R. (2020). Teacher Coaching in a Simulated Environment. *Educational Evaluation and Policy Analysis, 42*(2), 208–231.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Dahling, J. J., Chau, S. L., & O'Malley, A. (2012). Correlates and Consequences of Feedback Orientation in Organizations. *Journal of Management, 38*(2), 531–546.

Doherty-Bigara, J., & Gilmore, L. (2015). Development of the Dimensions of Adult Mastery Motivation Questionnaire. *The Australian Educational and Developmental Psychologist, 32*(2), 142–157.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41,* 1040–1048.

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*(3), 613–628.

Elsner Twesme, E., Werner, J. M., & Simha, A. (2021). Can We Talk? Improving Employee Contextual Performance Using Motivating Language and Feedback Orientation. *International Journal of Business Communication,* 23294884211038745.

Giersch, J., & Dong, C. (2018). Principals' Preferences When Hiring Teachers: A Conjoint Experiment. *Journal of Educational Administration, 56*(4), 429–444.

Gjerde, K. P., Skinner, D., & Padgett, M. (2022). Importance of Goal and Feedback Orientation in Determining Feedback Effectiveness. *Journal of the Scholarship of Teaching and Learning, 22*(3), 55–75.

Goldhaber, D. (2019). Evidence-Based Teacher Preparation: Policy Context and What We Know. *Journal of Teacher Education, 70*(2), 90–101.

Gregory, J. B., & Levy, P. E. (2012). Employee feedback orientation: Implications for effective coaching relationships. *Coaching: An International Journal of Theory, Research and Practice, 5*(2), 86–99.

Guskey, T. R., & Link, L. J. (2022). What Teachers Really Want When It Comes to Feedback. *Educational Leadership, 79*(7), 42–48.

Harris, D. J., Vine, S. J., Wilson, M. R., McGrath, J. S., LeBel, M.-E., & Buckingham, G. (2018). Action observation for sensorimotor learning in surgery. *British Journal of Surgery, 105*(13), 1713–1720.

Hemler, S. M. (2020). *Feedback Receptivity, Teacher Job Satisfaction, and Organizational Commitment* (27962909) [Doctoral dissertation, University of Charleston - Beckley]. ProQuest Dissertations and Theses. https://www.proquest.com/docview/2406662225/abstract/21854896F40B484CPQ/1

Herold, D. M., Parsons, C. K., & Rensvold, R. B. (1996). Individual Differences in the Generation and Processing of Performance Feedback. *Educational and Psychological Measurement, 56*(1), 5–25.

Ho, A. D., & Kane, T. J. (2013). *The Reliability of Classroom Observations by School Personnel.* Bill & Melinda Gates Foundation. https://usprogram.gatesfoundation.org/-/media/dataimport/resources/pdf/2016/12/met-reliability-of-classroom-observations-research-paper.pdf

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of Individual Feedback on Behavior in Organizations. *Journal of Applied Psychology, 64*(4), 349–371.

Jach, H. K., Bardach, L., & Murayama, K. (2023). How Personality Matters for Education Research. *Educational Psychology Review, 35*(3), 94.

JASP Team. (2023). *JASP* (Version 0.18.1) [Computer software].

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working Memory, Attention Control, and the N-back Task: A Question of Construct Validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 615–622.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment.* Bill & Melinda Gates Foundation. https://usprogram.gatesfoundation.org/-/media/dataimport/resources/pdf/2016/12/met-validating-using-random-assignment-research-paper.pdf

Katz, I. M., Moughan, C. M., & Rudolph, C. W. (2023). Feedback orientation: A meta-analysis. *Human Resource Management Review, 33*(4), 100986.

Katz, I. M., Rauvola, R. S., & Rudolph, C. W. (2021). Feedback environment: A meta-analysis. International *Journal of Selection and Assessment, 29*(3–4), 305–325.

Kraft, M. A., Papay, J. P., & Chi, O. L. (2020). Teacher Skill Development: Evidence from Performance Ratings by Principals. *Journal of Policy Analysis and Management, 39*(2), 315–347.

Kraft, & Papay, J. P. (2014). Can Professional Environments in Schools Promote Teacher Development? *Explaining Heterogeneity in Returns to Teaching Experience. Educational Effectiveness and Policy Analysis, 36*(4), 476–500.

Lechermeier, J., & Fassnacht, M. (2018). How do performance feedback characteristics influence recipients' reactions? A state-of-the-art review on feedback source, timing, and valence effects. *Management Review Quarterly, 68*(2), 145–193.

Levy, P. E., Albright, M. D., Cawley, B. D., & Williams, J. R. (1995). Situational and Individual Determinants of Feedback Seeking: A Closer Look at the Process. *Organizational Behavior and Human Decision Processes, 62*(1), 23–37.

Linderbaum, B. A., & Levy, P. E. (2010). The Development and Validation of the Feedback Orientation Scale (FOS). *Journal of Management, 36*(6), 1372–1405.

London, M., & Smither, J. W. (2002). Feedback orientation, feedback culture, and the longitudinal performance management process. *Human Resource Management Review, 12*(1), 81–100.

London, M., Smither, J. W., & Adsit, D. J. (1997). Accountability: The Achilles' Heel of Multisource Feedback. *Group & Organization Management, 22*(2), 162–184.

McLean, S. V. (1998). Becoming a Teacher: The Person in the Process. In R. P. Lipka & T. M. Brinthaupt (Eds.), *The Role of Self in Teacher Development*. State University of New York Press.

Mills, L. M., O'Sullivan, P. S., ten Cate, O., & Boscardin, C. (2023). Investigating feedback orientation in medical learners. *Medical Teacher, 45*(5), 492–498.

Quintelier, A., De Maeyer, S., & Vanhoof, J. (2020). The role of feedback acceptance and gaining awareness on teachers' willingness to use inspection feedback. *Educational Assessment, Evaluation and Accountability, 32*(3), 311–333.

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.2) [Computer software].

Rasheed, A., Khan, S.-U.-R., Rasheed, M. F., & Munir, Y. (2015). The Impact of Feedback Orientation and the Effect of Satisfaction With Feedback on In-Role Job Performance. *Human Resource Development Quarterly, 26*(1), 31–51.

Renn, R. W., & Fedor, D. B. (2001). Development and field test of a feedback seeking, self-efficacy, and goal setting model of work performance. *Journal of Management, 27*(5), 563–583.

Roberts, B. W., Lejuez, C., Krueger, R. F., Richards, J. M., & Hill, P. L. (2014). What is conscientiousness and how can it be assessed? *Developmental Psychology, 50*(5), 1315–1330.

Ronfeldt, M., & Campbell, S. L. (2016). Evaluating Teacher Preparation Using Graduates' Observational Ratings. *Educational Evaluation and Policy Analysis, 38*(4), 603–625.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*, 1–36.

Sims, S., Fletcher-Wood, H., Godfrey-Fausset, T., & Meliss, S. (2023). *Modelling evidence-based practice in initial teacher training: Causal effects on teachers' skills, knowledge and self-efficacy.* Ambition Institute. https://www.ambition.org.uk/modelling-evidence-based-practice-in-initial-teacher-training/

Smither, J. W., London, M., & Reilly, R. R. (2005). Does Performance Improve Following Multisource Feedback? A Theoretical Model, Meta-Analysis, and Review of Empirical Findings. *Personnel Psychology, 58*(1), 33–66.

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113,* 117–143.

Steelman, L. A., Levy, P. E., & Snell, A. F. (2004). The Feedback Environment Scale: Construct Definition, Measurement, and Validation. *Educational and Psychological Measurement, 64*(1), 165–184.

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096–1104.

Stoet, G. (2017). PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology, 44*(1), 24–31.

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education, 48*(6), 1273–1296.

Tonidandel, S., Quiñones, M. A., & Adams, A. A. (2002). Computer-adaptive testing: The impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology, 87,* 320–332.

Tuytens, M., & Devos, G. (2014). How to activate teachers through teacher evaluation? *School Effectiveness and School Improvement, 25*(4), 509–530.

Vandewalle, D. (1997). Development and Validation of a Work Domain Goal Orientation Instrument. *Educational and Psychological Measurement, 57*(6), 995–1015.

VandeWalle, D., Ganesan, S., Challagalla, G. N., & Brown, S. P. (2000). An integrated model of feedback-seeking behavior: Disposition, context, and cognition. *Journal of Applied Psychology, 85,* 996–1003.

Weiss, J. A., Outland, N., Plummer, G., Zervos, L., Carmichael-Tanaka, N., & Kang, B. (2023). The stable individual differences driving employee coachability behaviours. *International Journal of Evidence Based Coaching and Mentoring, 21*(2), 102–117.

Appendix

Appendix A: Survey questionnaires

Feedback Orientation Scale (FOS; Linderbaum & Levy, 2010)

Administer these items using a 5-point scale from strongly disagree to strongly agree.

Utility

1. Feedback contributes to my success at work.
2. To develop my skills at work, I rely on feedback.
3. Feedback is critical for improving performance.
4. Feedback from supervisors can help me advance in a company.
5. I find that feedback is critical for reaching my goals.

Accountability

1. It is my responsibility to apply feedback to improve my performance.
2. I hold myself accountable to respond to feedback appropriately.
3. I don't feel a sense of closure until I respond to feedback.
4. If my supervisor gives me feedback, it is my responsibility to respond to it.
5. I feel obligated to make changes based on feedback.

Social Awareness

1. I try to be aware of what other people think of me.
2. Using feedback, I am more aware of what people think of me.
3. Feedback helps me manage the impression I make on others.
4. Feedback lets me know how I am perceived by others.
5. I rely on feedback to help me make a good impression.

Feedback Self-Efficacy

1. I feel self-assured when dealing with feedback.
2. Compared to others, I am more competent at handling feedback.
3. I believe that I have the ability to deal with feedback effectively.
4. I feel confident when responding to both positive and negative feedback.
5. I know that I can handle the feedback that I receive.

Achievement Goal Questionnaire – Revised (AGQ-R; Elliot & Murayama, 2008)

Administer these items using a 5-point scale from strongly disagree to strongly agree.

Mastery approach: items 1, 3, 7
Mastery avoidance: items 5, 9, 11
Performance approach: items 2, 4, 8
Performance avoidance: 6, 10, 12

1. My aim is to completely master the material presented in this class.
2. I am striving to do well compared to other students.
3. My goal is to learn as much as possible.
4. My aim is to perform well relative to other students.
5. My aim is to avoid learning less than I possibly could.
6. My goal is to avoid performing poorly compared to others.
7. I am striving to understand the content of this course as thoroughly as possible.
8. My goal is to perform better than the other students.
9. My goal is to avoid learning less than it is possible to learn.
10. I am striving to avoid performing worse than others.
11. I am striving to avoid an incomplete understanding of the course material.
12. My aim is to avoid doing worse than other students.

Feedback seeking (adapted from VandeWalle et al., 2000)

Please complete this section thinking about your current school placement. The term **supervisor** here refers to your in-school mentor, tutor or classroom teacher. If one person regularly monitors or observes you, consider this person your supervisor. If this is more than one person, please consider them both.

7-point scales anchored by 1 (never) and 7 (daily).
1 = never, 2 = less than once a month, 3 = monthly, 4 = 2/3 times a month, 5 = weekly, 6 = more than once a week, 7 = daily

How frequently have you sought feedback from your supervisor regarding:

a) overall job performance
b) technical aspects of the job (subject knowledge, pedagogies, practices)
c) values and attitudes of the school (including policies)
d) role expectations
e) social behaviours (including professional relationships)

Feedback Environment Scale (Steelman et al., 2004)

Please complete this section thinking about your current school placement. The term **supervisor** here refers to your in-school mentor, tutor or classroom teacher. If one person regularly monitors or observes you, consider this person your supervisor. If this is more than one person, please consider them both.

Co-workers refers to other colleagues you work closely with in school, such as your department or phase.
Performance review could refer to a weekly review of progress, a progress meeting or similar.
Job performance can refer to anything related to your teaching role.
Please consider **deadlines** in terms of school matters, not university deadlines.

7-point Likert-type scale ranging from 1 = strongly disagree to 7 = strongly agree.

Source credibility:

Supervisor source

1. My supervisor is generally familiar with my performance on the job.
2. In general, I respect my supervisor's opinions about my job performance.
3. With respect to job performance feedback, I usually do not trust my supervisor.
4. My supervisor is fair when evaluating my job performance.
5. I have confidence in the feedback my supervisor gives me.

Co-worker source

1. My co-workers are generally familiar with my performance on the job.
2. In general, I respect my co-workers' opinions about my job performance.
3. With respect to job performance feedback, I usually do not trust my co-workers.
4. My co-workers are fair when evaluating my job performance.
5. I have confidence in the feedback my co-workers give me.

Feedback quality:

Supervisor source

1. My supervisor gives me useful feedback about my job performance.
2. The performance feedback I receive from my supervisor is helpful.
3. I value the feedback I receive from my supervisor.
4. The feedback I receive from my supervisor helps me do my job.
5. The performance information I receive from my supervisor is generally not very meaningful.

Co-worker source

1. My co-workers give me useful feedback about my job performance.
2. The performance feedback I receive from my co-workers is helpful.
3. I value the feedback I receive from my co-workers.
4. The feedback I receive from my co-workers helps me do my job.
5. The performance information I receive from my co-workers is generally not very meaningful.

Feedback delivery:

Supervisor source

1. My supervisor is supportive when giving me feedback about my job performance.
2. When my supervisor gives me performance feedback, he or she is considerate of my feelings.
3. My supervisor generally provides feedback in a thoughtless manner.
4. My supervisor does not treat people very well when providing performance feedback.
5. My supervisor is tactful when giving me performance feedback.

Co-worker source

1. My co-workers are supportive when giving me feedback about my job performance.
2. When my co-workers give me performance feedback, they are usually considerate of my feelings.
3. My co-workers generally provide feedback in a thoughtless manner.
4. In general, my co-workers do not treat people very well when providing performance feedback.
5. In general, my co-workers are tactful when giving me performance feedback.

Favourable feedback:

Supervisor source

1. When I do a good job at work, my supervisor praises my performance.
2. I seldom receive praise from my supervisor.
3. My supervisor generally lets me know when I do a good job at work.
4. I frequently receive positive feedback from my supervisor.

Co-worker source

1. When I do a good job at work, my co-workers praise my performance.
2. I seldom receive praise from my co-workers.
3. My co-workers generally let me know when I do a good job at work.
4. I frequently receive positive feedback from my co-workers.

Unfavourable feedback:

Supervisor source

1. When I don't meet deadlines, my supervisor lets me know.
2. My supervisor tells me when my work performance does not meet organizational standards.
3. On those occasions when my job performance falls below what is expected, my supervisor lets me know.
4. On those occasions when I make a mistake at work, my supervisor tells me.

Co-worker source

1. When I don't meet deadlines, my co-workers let me know.
2. My co-workers tell me when my work performance does not meet organizational standards.
3. On those occasions when my job performance falls below what is expected, my co-workers let me know.
4. On those occasions when I make a mistake at work, my co-workers tell me.

Source availability:

Supervisor source

1. My supervisor is usually available when I want performance information.
2. My supervisor is too busy to give me feedback.
3. I have little contact with my supervisor.
4. I interact with my supervisor on a daily basis.
5. The only time I receive performance feedback from my supervisor is during my performance review.

Co-worker source

1. My co-workers are usually available when I want performance information.
2. My co-workers are too busy to give me feedback.
3. I have little contact with my co-workers.
4. I interact with my co-workers on a daily basis.

Promotes feedback seeking:

Supervisor source

1. My supervisor is often annoyed when I directly ask for performance feedback.
2. When I ask for performance feedback, my supervisor generally does not give me the information right away.
3. I feel comfortable asking my supervisor for feedback about my work performance.
4. My supervisor encourages me to ask for feedback whenever I am uncertain about my job performance.

Co-worker source

1. My co-workers are often annoyed when I directly ask them for performance feedback.
2. When I ask for performance feedback, my co-workers generally do not give me the information right away.
3. I feel comfortable asking my co-workers for feedback about my work performance.
4. My co-workers encourage me to ask for feedback whenever I am uncertain about my job performance.

Big Five Inventory-2 (BFI-2) Conscientiousness (Soto & John, 2017)

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

1 = disagree strongly; 2= disagree a little; 3 = neutral, no opinion; 4 = agree a little; 5 = agree strongly

I am someone who…

1. tends to be disorganised
2. tends to be lazy
3. is dependable, steady
4. is systematic, likes to keep things in order
5. has difficulty getting started on tasks
6. can be somewhat careless
7. keeps things neat and tidy
8. is efficient, gets things done
9. is reliable, can always be counted on
10. leaves a mess, doesn't clean up
11. is persistent, works until the task is finished
12. sometimes behaves irresponsibly

Dimensions of Adult Mastery Motivation Questionnaire (DAMMQ; Doherty-Bigara & Gilmore, 2015)

7-point Likert scale ranging from 1 (not at all typical) to 5 (very typical)

1. I persist with a task even if I feel it is difficult.
2. If I am unsuccessful with a difficult task I know that I can gain the skills needed to try it again.
3. I practise new skills over and over until I am satisfied.
4. I work at a new challenge until I feel I can do it well.
5. I attempt difficult tasks even if I have some uncertainty about whether I will be able to complete them.
6. I like to build on my existing skills even if it will be difficult for me.
7. I prefer to challenge myself with difficult tasks even if I am unsure that I will be able to complete them.
8. I explore all ways to solve a problem.
9. I like a challenge when learning new skills.
10. I find challenging tasks to be more interesting than easy ones.
11. I enjoy being challenged by difficult tasks.
12. I choose to do tasks that I think will be challenging for me.
13. I feel proud of myself when I am successful.
14. I feel excited when I realise that I am making progress with a difficult task.
15. I feel a sense of achievement when I complete a difficult task.
16. I begin to enjoy difficult tasks as I begin to develop new skills.
17. I am good at the things I do.
18. I solve problems quickly.
19. I believe I have the skills and abilities to complete challenges that I am faced with.
20. I usually am able to complete tasks even if they are difficult.
21. I can become completely absorbed in a challenging task.
22. I often lose track of time when I am working on a challenging task.
23. I generally persist with a difficult task for a long time.
24. I can become completely immersed in tasks that are developing my skills.

Feedback satisfaction

Adapted from (Rasheed et al., 2015; Tonidandel et al., 2002)

Please rate your level of agreement with each of the following statements on a scale from 1 ("strongly disagree") to 5 ("strongly agree")

1. I feel that I am valued by this feedback.
2. I think my feedback exactly represents my output.
3. I have learnt a lot to improve my performance through my feedback.
4. I believe that the feedback process is fair and justified.

Working memory

We measured working memory using the n-back task – a widely used test in psychology (see for example M. J. Kane et al., 2007). We adapted a computerised version freely available online at PsyToolkit (Stoet, 2010; 2017). Participants viewed a sequence of letters presented one-by-one on screen (each for max 2000ms; 15 different letters used) and were asked to decide whether they saw the same letter two trials ago (that is, $n = 2$ back). If yes, they indicated this by pressing the letter 'm' on their keyboard. The task started with a practice block during which feedback was provided (green bars appear above and below the letter for correct responses, and red bars for incorrect). This was followed by two task blocks without feedback. Each block consisted of 30 trials. We calculated individual working memory as accuracy rate (hit rate – false alarm rate) according to standard signal detection theory.

Appendix B: Additional results

Table B1. H4 Sensitivity analysis: Regression to predict FO at Survey 2 using Survey 1, the time interval between surveys and the interaction between both

	Summed scores			Weighted scores		
	β (*SE*)	*t*	*p*	β (*SE*)	*t*	*p*
FO model	R^2 **= .32**			R^2 **= .30**		
(Intercept)	0 (0.07)	0.01	.991	0 (0.07)	0	.997
Time interval	0.04 (0.07)	0.54	.593	0.05 (0.07)	0.75	.456
FO – Survey 1	0.56 (0.07)	7.77	<.001	0.54 (0.07)	7.39	<.001
Time interval x FO	0.03 (0.07)	0.41	.681	0.01 (0.07)	0.09	.930
FO utility model	R^2 **= .18**			R^2 **= .18**		
(Intercept)	0 (0.08)	0	.997	0 (0.08)	-0.02	.983
Time interval	0.09 (0.08)	1.17	.244	0.1 (0.08)	1.24	.216
FO utility – Survey 1	0.42 (0.08)	5.26	<.001	0.42 (0.08)	5.24	<.001
Time interval x FO utility	-0.03 (0.08)	-0.32	.747	-0.04 (0.08)	-0.54	.59
FO accountability model	R^2 **= .29**			R^2 **= .28**		
(Intercept)	0 (0.07)	0.01	.994	0 (0.07)	0.04	.972
Time interval	-0.04 (0.07)	-0.53	.598	-0.03 (0.07)	-0.44	.663
FO accountability – Survey 1	0.55 (0.07)	7.38	<.001	0.53 (0.07)	7.16	<.001
Time interval x FO accountability	0 (0.08)	-0.05	.958	-0.02 (0.08)	-0.28	.783
FO social awareness model	R^2 **= .35**			R^2 **= .41**		
(Intercept)	0.01 (0.07)	0.12	.905	0.01 (0.07)	0.19	.853
Time interval	-0.04 (0.07)	-0.52	.601	-0.03 (0.07)	-0.5	.621
FO social awareness – Survey 1	0.57 (0.07)	8.00	<.001	0.61 (0.07)	8.88	<.001
Time interval x FO social awareness	0.06 (0.07)	0.87	.387	0.08 (0.07)	1.15	.253
FO self-efficacy model	R^2 **= .30**			R^2 **= .30**		
(Intercept)	0 (0.07)	-0.02	.987	0 (0.07)	-0.03	.976
Time interval	0.1 (0.07)	1.32	.188	0.11 (0.07)	1.45	.149
FO self-efficacy – Survey 1	0.55 (0.07)	7.48	<.001	0.54 (0.07)	7.41	<.001
Time interval x FO self-efficacy	-0.04 (0.07)	-0.49	.628	-0.05 (0.07)	-0.74	.458

Note. The dependent variable here is FO measured at Survey 2. β (*SE*) = standardised regression coefficient (standard error); FO = feedback orientation.

Table B2. H4 Exploratory regression analyses: changes in FO and FO self-efficacy (between Survey 1 and 2) predicted by feedback seeking behaviour and feedback environment

	FO (Survey 2)			FO self-efficacy (Survey 2)		
	β (*SE*)	*t*	*p*	β (*SE*)	*t*	*p*
Model 1	R^2 = .38			R^2 = .39		
Survey 1 FO baseline measure	.47 (0.07)	6.70	< .001	.52 (0.07)	7.40	<.001
Feedback seeking	.08 (0.07)	1.14	.285	.03 (0.07)	0.45	.653
Feedback environment	.28 (0.07)	3.99	<.001	.31 (0.07)	4.57	<.001
Model 2	R^2 = .38			R^2 = .40		
Survey 1 FO baseline measure	.49 (0.07)	6.91	<.001	.53 (0.07)	7.75	<.001
Feedback environment – supervisor	.24 (0.08)	3.09	.002	.29 (0.08)	3.93	<.001
Feedback environment – coworkers	0.10 (0.08)	1.30	.196	.08 (0.08)	1.05	.295
Model 3	R^2 = .27			R^2 = .39		
Survey 1 FO baseline measure	.51 (0.07)	7.25	<.001	.52 (0.07)	7.62	<.001
Supervisor credibility	.27 (0.07)	3.92	<.001	.32 (0.07)	4.69	<.001
Model 4	R^2 = .35			R^2 = .36		
Survey 1 FO baseline measure	.54 (0.07)	7.67	<.001	.54 (0.07)	7.83	<.001
Supervisor feedback quality	.25 (0.07)	3.53	<.001	.28 (0.07)	4.01	<.001
Model 5	R^2 = .33			R^2 = .36		
Survey 1 FO baseline measure	.54 (0.07)	7.57	<.001	.56 (0.07)	8.04	<.001
Supervisor feedback delivery	.19 (0.07)	2.66	.009	.27 (0.07)	3.82	<.001
Model 6	R^2 = .40			R^2 = .43		
Survey 1 FO baseline measure	.51 (0.07)	7.56	<.001	.53 (0.07)	8.10	<.001
Supervisor favourable feedback	.33 (0.07)	4.82	<.001	.37 (0.07)	5.66	<.001
Model 7	R^2 = .30			R^2 = .30		
Survey 1 FO baseline measure	.52 (0.07)	7.06	<.001	.52 (0.07)	6.98	<.001
Supervisor unfavourable feedback	.11 (0.07)	1.44	.152	.12 (0.07)	1.58	.116
Model 8	R^2 = .32			R^2 = .33		
Survey 1 FO baseline measure	.52 (0.07)	7.25	<.001	.54 (0.07)	7.56	<.001
Supervisor availability	.16 (0.07)	2.18	.031	.20 (0.07)	2.77	.007
Model 9	R^2 = .34			R^2 = .34		
Survey 1 FO baseline measure	.53 (0.07)	7.43	<.001	.53 (0.07)	7.52	<.001
Supervisor promotes feedback seeking	.21 (0.07)	2.89	.004	.23 (0.07)	3.32	.001

Note. All measures of FO and FO self-efficacy are weighted based on factor analysis (see Methods). FO = feedback orientation.

Figure B1. Binscatter plots showing the relationship between FO and improvements in participant skill and pupil time off task in the simulator training exercise

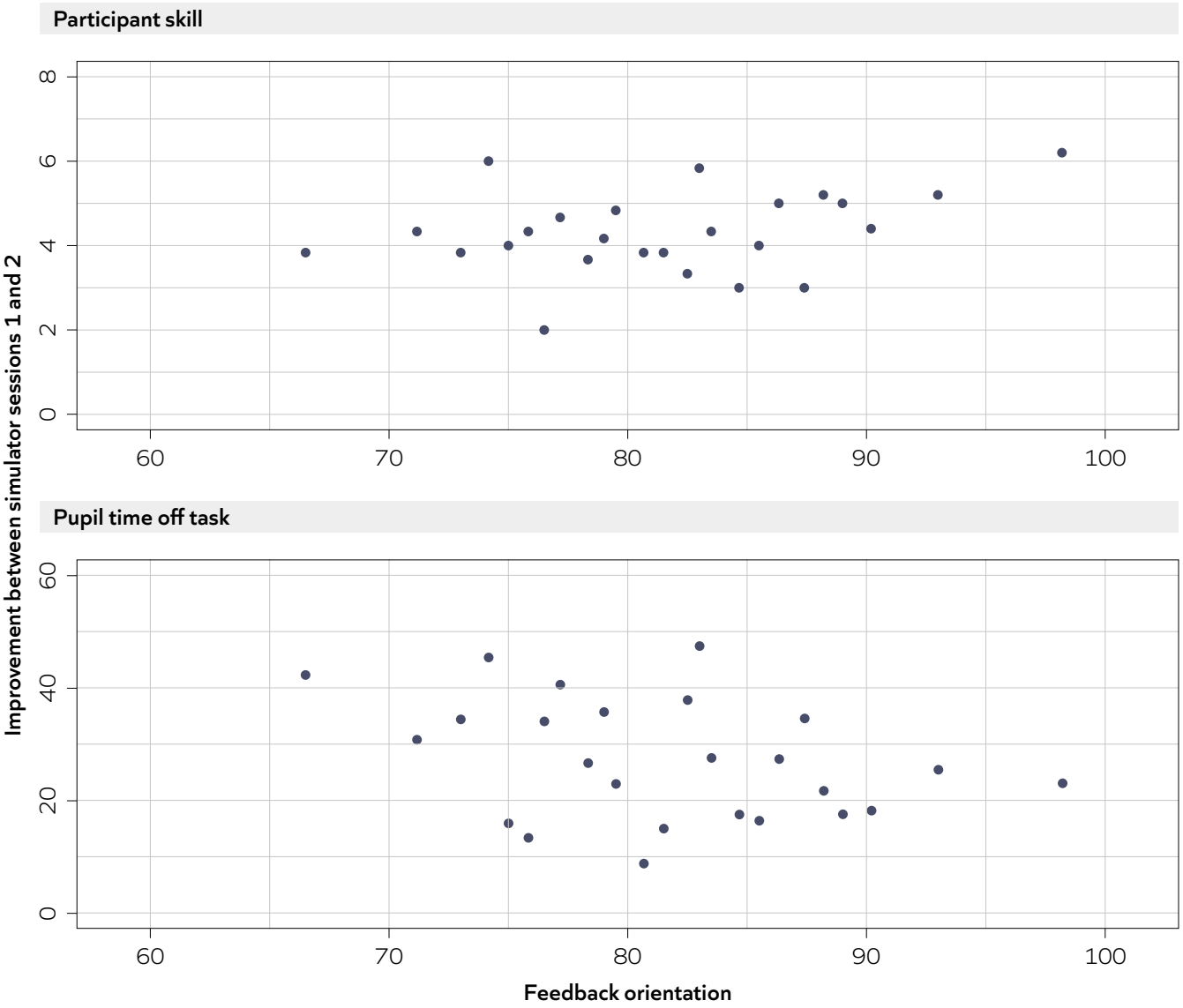
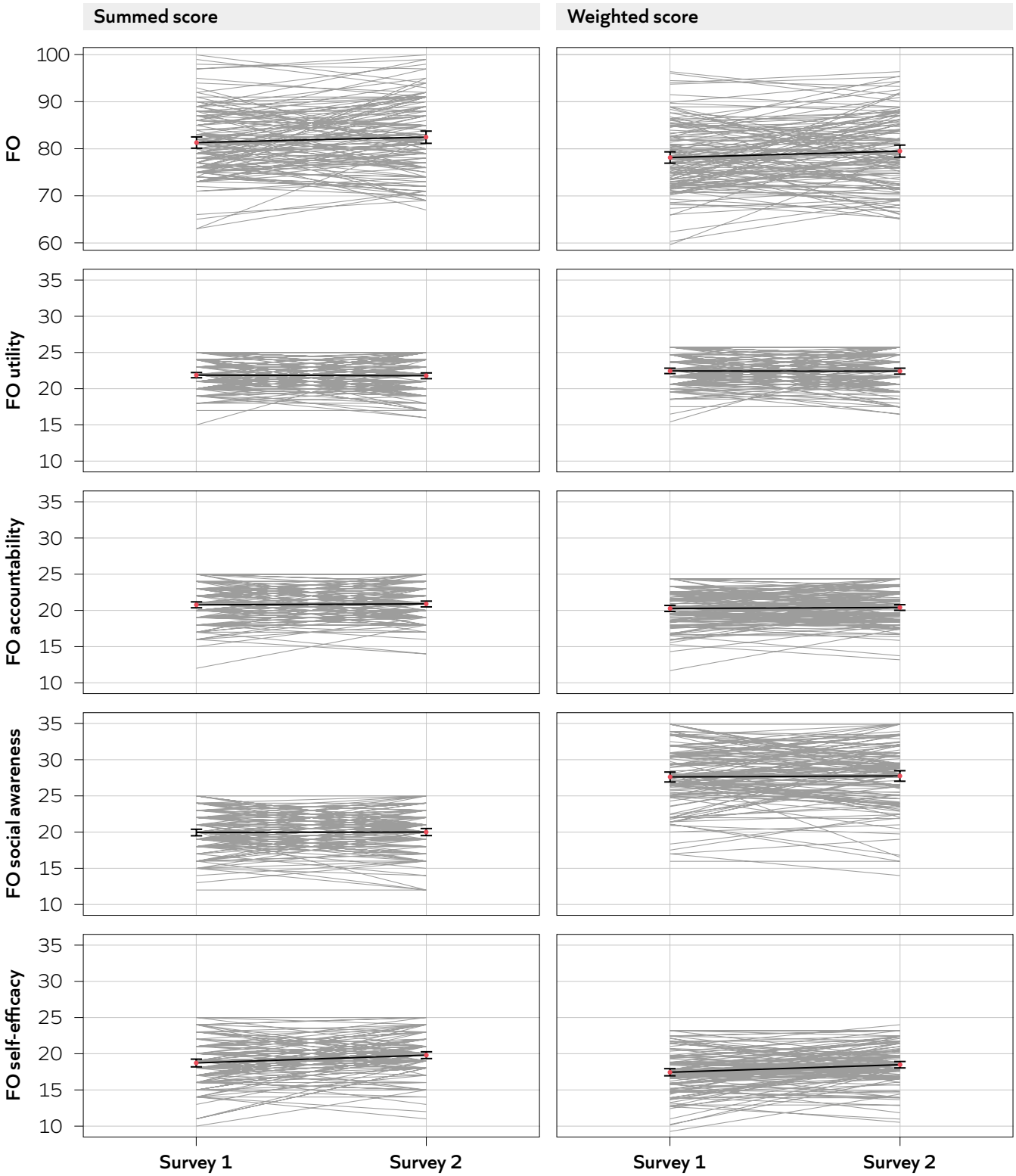


Figure B2. Changes in summed and weighted FO and its subscales between the start (Survey 1) and end (Survey 2) of trainees' first term of training



Note. Mean values are plotted in coral. Grey lines indicate changes for individual participants. Error bars show 95% CIs. FO = feedback orientation.

A great teacher changes the future every day. They can be the critical factor in a child's success, especially for those who have had a tough start in life. At Ambition Institute, we support teachers and school leaders at every stage of their careers, helping them to keep getting better. We are a charity providing training and professional development based on the most rigorous research and evidence about what really works. Together, we're shaping the future of education to give every child the best start in life.

Learn more

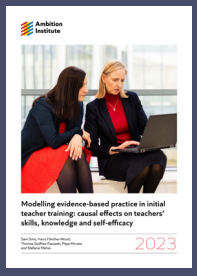
Get in touch today to find out how you can nominate your staff, or yourself, for one of our programmes.

ambition.org.uk
info@ambition.org.uk
020 3668 6865
@Ambition_Inst
@Ambition.Inst

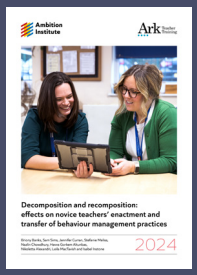
Ambition Institute
2 Bridge Wharf,
156 Caledonian Road,
London, N1 9UU

Discover more

Evidence is at the heart of what we do at Ambition Institute. Through our research we want to push forward our collective knowledge of effective professional development and share what works, to turn evidence into action. We aim to answer important questions faced by the sector, help teachers and leaders to address disadvantage, and show the impact of evidence-informed teaching.



Does adding modelling to professional development help to bridge the theory-practice gap? This research looked at whether video models helped initial teacher trainees to develop their retrieval practice skills.



This research explored whether breaking down (decomposing) and then recombining teaching practices into a different sequence of teaching (recomposing) helped initial teacher trainees apply what they learnt to new contexts.

Ambition Institute

© Ambition Institute 2024. Ambition Institute is a registered charity, number 1146924.