Teaching and Teacher Education 129 (2023) 104136

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate

Research paper

The effectiveness of the Conversation Analytic Role-Play Method (CARM) on interactional awareness: A feasibility randomized controlled trial with student teachers

Rein Ove Sikveland [a, *], Thomas Moser [b], Marit Skarbø Solem [c], Karianne Skovholt [c]

[a] Department of Language and Literature, Norwegian University of Science and Technology (NTNU), Dragvoll Campus, Edvard Bulls veg 1, 7491, Trondheim, Norway
[b] University of Stavanger (UiS), Norwegian Centre for Learning Environment and Behavioural Research in Education, Department Porsgrunn, Kjølnes Ring 30, 3918, Porsgrunn, Norway
[c] Department of Languages and Literature Studies, University of South-Eastern Norway (USN), Vestfold Campus, Raveien 215, 3184, Borre, Norway

ARTICLE INFO

Article history:
Received 18 July 2022
Received in revised form
18 February 2023
Accepted 31 March 2023
Available online 28 April 2023

Keywords:
Interactional awareness
Pedagogical competence
Conversation analysis
Randomized controlled trial
Communication skills training
Video analysis

ABSTRACT

This study tested the effectiveness of communication skills training intervention CARM (Conversation Analytic Role-play Method) on student teachers at a university in Norway. A feasibility randomized controlled trial (n = 41) found significant positive effects (d = 1.09) of the intervention on interactional awareness. Qualitative and quantitative results from an acceptability and satisfaction questionnaire revealed that the CARM training proves to be a highly appropriate training method in teacher education. This study shows how communication skills training may be implemented and evaluated using controlled experimental designs, and highlights some of the challenges and limitations of RCTs in educational research.

1. Introduction

Teacher education programmes globally have faced increasing requirements to provide instruction methods that bridge teacher training practices on campus with real-world teaching practices teachers will experience in schools. Scholars agree there is an urgent need for more innovative instruction methods that involve evidence-based, case-oriented, explorative, and student-active teaching methods (e.g., Darling-Hammond, 2006; Jahreie & Ottesen, 2010; Ministry of Education and Research, 2008–2009, 2012; Gore et al., 2017), and teacher educators have further specified the need of *communication skills training* (CST) into teacher education programs (Gisewhite, Jeanfreau & Holden, 2021; Hovdenak & Wiese, 2017; Hunt, Simonds & Cooper, 2002; Ortega & Fuentes, 2015; Simonds et al., 2008).

Despite the great emphasis on CST for teachers, few attempts have been made globally to implement and evaluate innovative new techniques into teacher training programmes (Ortega & Fuentes, 2015). Communication skills training is almost non-existent in both international and Norwegian teacher education (Nesje, 2020), despite such methods having been shown to motivate and increase students' self-efficacy before their first in-service period, prepare their practical competence, and bridge the gap between theory and practice (Kaufman & Ireland, 2016; Ministry of Education and Research, 2016–2017; Wiesbeck et al., 2017; Nesje, 2020).

A challenge for teacher education is to offer training activities that make teaching methods and experiences relevant and transferable to the activities and situations student teachers encounter in actual school contexts (Hovdenak & Wiese, 2017). Part of the challenge is to build a knowledge base which pinpoints what the relevant transferable skills are, i.e., the key features that constitute teachers' tacit expertise in their everyday practice. While there is a growing body of research describing teachers' tacit expertise (see

* Corresponding author. Department of Language and Literature, Norwegian University of Science and Technology (NTNU), Dragvoll campus, Edvard Bulls veg 1, 7491, Trondheim, Norway.
E-mail address: rein.o.sikveland@ntnu.no (R.O. Sikveland).

https://doi.org/10.1016/j.tate.2023.104136

e.g., Sert, 2013; Skovholt et al., 2021; Waring, 2007; Waring & Creider, 2021), there is need for more research; in particular, we need more empirical evidence on the implementation of CSTs ('interventions') and what the outcomes are for the student teachers who have undergone the training (i.e., the 'effectiveness' of such interventions).

1.1. The implementation and effectiveness of CST interventions in teacher training programmes

Implementation of CST in teacher education to date comprise experimental and intervention studies that investigate effects of instruction methods such as role-play, simulations, computer-based simulations, and video-feedback, designed to support student teachers in their daily practice and overall professional development (Ferry et al., 2004; Dalgarno et al., 2016; Fischetti et al., 2022; Theelen, van den Beemt, & den Brok, 2019; Simonds et al., 2008; Hoogendijk et al., 2018). In general, these studies demonstrate positive effects of the CST on practitioners' communication skills. They argue the CST provides a valuable opportunity for student teachers to practice and enhance their practical experience, to build self-confidence and prepare novices for their first "real" teaching practicum, and also functions as an assessment tool to offer feedback and reflection based upon repeated practice (Kaufman & Ireland, 2016; Fischetti et al., 2022).

Despite the general positive outcome of CST interventions, the studies reveal three overall concerns regarding CST intervention design and evaluation. First, some studies are based on more or less "inauthentic" scenarios where student teachers simulate, or role play, *imagined* situations, often based on a predefined scenario (Simonds et al., 2008; Lindset & Aune, 2020; Lindstøl, 2016; Fischetti et al., 2021). The advantage of simulation-based interventions is that they provide a scenario for practicing in an environment where the stakes for success/failure are reduced. The risk, however, is that the simulation is not based on actual practices that take places in real life settings (Stokoe, 2014). In general, though findings from some of the previous studies on simulations in classroom revealed that the students' experiences were useful to prepare for their future practice in real classrooms (e.g. Fukkink, Trienekens & Kramer, 2011; Gartmeier et al., 2015), some also reported that the activities differed from the reality of the classroom (e.g. Lindset & Aune, 2020; Lindstøl, 2016) and that future research should address "actual classroom behaviour" (Simonds et al., 2008, p. 63). Therefore, some teacher educators address this concern about reduced authenticity using video-recordings of actual real-life encounters as a basis for the training, whereby teachers or student teachers reflect on their own and others' authentic classroom communication (Carpenter, 2021; Waring & Creider, 2021). This method is referred to as "micro-teaching", or "video-feedback", a method that "allows course participants to look at themselves 'from a distance' and with space for reflection, thereby giving them a realistic picture of their own skills, or self-image (…). Through repeated playing of the videotape, this method also allows a detailed analysis of a person's behaviour" (Fukkink et al., 2011, p. 46).

A key question is whether CST as video-feedback and other simulation methods have an impact on professionals' practice. In the context of surgeon's ward rounds in hospital, White, Ward and Hibberd (2021) video-recorded one surgeon's ward rounds and provided a CA-informed CST for one-to-one training. In this training, specific target skills were emphasized and the analysis of post-training recordings showed that the practitioner modified the verbal behaviour after training. In another study, Carpenter (2021) implemented CA-based interactional evidence with a group of student teachers, using repeated cycles of reflective analysis. The intervention focused on turn-by-turn observations of authentic elicitations in the classroom, related to securing attention, dealing with unresponsiveness and with student contributions. The author recorded student teachers before and after the intervention to discuss strategies for improvement, finding that student teachers improved their elicitation methods following the intervention.

The second concern is whether the CSTs were based on *predefined targeted ("best practice") communication skills*. Such targeted communication skills may range from micro behaviours such as the use of head nodding, hand gestures, silence (Lindset & Aune, 2020) to broader or more holistic skills related to behavioural dimensions like "immediacy" (Simonds et al., 2008), "establishing a supportive interpersonal relationship", "advancing a joint problem solution" and "structuring the conversation in a pro-active and transparent manner" (Gartmeier et al., 2015, p. 445), or involve a more general communication model that "comprises different stages in the interactional encounter, such as initiating the session, gathering information and closing the session" (Fukkink et al., 2011, p. 46). Some studies did not use any systematic approach or targeting communication skills in the CST experiment (Simonds et al., 2008), which makes it hard to assess the outcome of the CST. According to Fukkink et al. (2011), the outcomes of video-feedback are considerably greater if participants use a standard evaluation form that gives an overview of the desired target behaviour which forms part of the training program, because such a form "structures the observation, thereby focusing the participants' attention on the aspects of their own behaviour that are central to the program" (Fukkink et al., 2011, p. 56). According to their meta-analysis, the video feedback method has a statistically significant effect on the interaction skills of professionals (Fukkink et al., 2011). However, these effects were stronger for positive outcomes, i.e., biased towards more desired behaviours as opposed to less desired ones.

The third concern is how *effectiveness is evidenced*, and what can be concluded based on the range of different study designs and qualitative and quantitative outcome assessments. Some studies use a debriefing talk with the students as outcome assessment (Lindset & Aune, 2020), others conduct pre- and post-intervention surveys to check students' attitude towards the training (Dalgarno et al., 2016; Gartmeier et al., 2015; Simonds et al., 2008). For example, Dalgarno et al. (2016) conducted a post-intervention questionnaire where student teachers responded to questions asking them to rate the degree to which they found the activity "confusing, difficult, irrelevant, boring, interesting, easy to use, useful and enjoyable" (p. 137). Overall, there is considerable heterogeneity in both study design and outcome assessment that calls for systematic review. An overall concern with previous CST studies in teacher education have been a lack of control groups, and not many have used experimental designs. One exception is a randomized controlled trial (RCT) ($n = 168$) by Gartmeier et al. (2015) assessing communication competence following three training conditions. Their study featured (i) contrastive video cases as e-learning, (ii) role-play with a trained actor including video-feedback and (iii) their combination, and found a strong overall effect, and that e-learning alone proved more effective than role-play alone.

Given the extant discussion of the literature and the issues it raises, our study focuses on one specific type of CST intervention: the Conversation Analytic Role-play Method (CARM; described in detail in the Methods section). CARM is based on conversation analytic evidence, which has a long track-record for using authentic video/audio in real time to present the audience (learners) with key interactional practice within a particular professional setting (see e.g., Stokoe, 2014; Stokoe & Sikveland, 2017). Our implementation and evaluation of CARM as a CST in student teacher training addresses the three concerns raised above. First, CARM builds

communication training materials based upon *authentic* interaction. Second, *the targeted communication skills are not pre-defined* but based on evidence-based conversation analytic research using audio- and video materials as a basis for the training. Third, the effectiveness of CARM is, for the first time, evaluated using a feasibility randomised controlled trial design (i.e., an experimental paradigm) implemented with student teachers as participant group.

As the problems with previous CST studies in teacher education have been a lack of experimental designs, the current study used an experimental design to increase control and reduce acquiesce bias, e.g., reducing the bias that participants respond based on their perceptions of how they think the researcher wants them to respond. We conceived CARM as a 'complex intervention' because of the high number of interlocking parts needed to ensure the success of the delivery and uptake of this intervention. The UK Medical Research Council (MRC) has been at the forefront in developing guidelines for the developing and evaluation of complex intervention (Skivington et al., 2021). And keeping with the MRC's guidelines, we decided that we needed to conduct a feasibility RCT, before we launched a full-scale RCT, to determine the feasibility and acceptability of the intervention and to identify any signals of efficacy of the intervention. Our aims for this feasibility trial were to.

- Explore the practicalities, acceptability, and satisfaction of delivering the intervention
- Determine whether a CARM intervention improves interactional awareness amongst student teachers compared to current training

2. Methodology

2.1. Design

This study is a feasibility randomized controlled trial (RCT) designed (i) to assess the feasibility and acceptability and (ii) to evaluate the effectiveness of the CARM intervention to improve student teachers' pedagogical competence based on a scale of 'interactional awareness' as outcome measure. Thus, this study aims to explore whether there is realism in evaluating a CARM intervention in teacher education using a RCT design. In case of positive effects and despite a small sample size, applying a rigorous experimental can lay an appropriate foundation for the implementation of a further large-scale study that allows for causal statements on a stronger statistical basis. As this is a feasibility trial we report on the practicalities, acceptability, and satisfaction of delivering the intervention in the results section.

The trial was embedded within the teaching curriculum for students in one class as part of an "Educational Science" module. The CARM training was a newly developed *additional* part of the curriculum, which was *voluntary*, i.e., the student teachers (henceforth; STs) were not obliged to take part and they were not going to be assessed on the content of the CARM training. The training could only be delivered in small groups, which therefore enabled an *opportunistic crossover design*, with the experimental group receiving the CARM intervention first (week 2) and after the baseline assessment (week 1), and the control group receiving the intervention later (week 4) and after both groups had completed the outcome assessment (week 3; see Fig. 1).

2.2. Participants

All STs were students studying for a Master's degree in Pedagogics as part of a teacher training course at a university in Norway. Therefore, all STs in one class were eligible and participated in the trial. There were no exclusion criteria. All STs received a participant information sheet about the study in advance. STs were recruited and informed about the study by the module convenor, and two reminder emails were sent to all about completing the study questionnaires (outlined below) before a deadline. STs were also informed that participating in the CARM training was optional, and therefore if they attended the training, this was considered as implied consent. The participants were informed that they could withdraw from the training at any point without having to give a reason and were assured that there would be no negative repercussions for not participating. The trainer (i.e., the person who delivered the intervention) was not involved in recruiting participants.

All STs had comparable levels of educational achievement, as they were all undertaking the second year of the same Master's programme. All participants had either four or five years of teacher education prior to this second year of the Master's. The participants had either opted to take Mathematics or Norwegian as their core subject for the Master's programme. Besides the different core subjects the participants undertook the same modules in the programme. Both trial groups had STs of different genders and a combination of those with Mathematics and Norwegian as additional/core subjects (see Section 3.1.1).

2.3. Intervention

Like video-feedback, CARM uses real-life, authentic recordings as basis for learning and reflection (Stokoe, 2014). The CARM intervention presents video excerpts and associated transcripts to the audience. These excerpts are selected based on key interactional challenges and opportunities identified through the conversation analytic research. The interactional challenges and opportunities are presented as 'projects' on the 'interactional racetrack' (Stokoe, 2014). The content of the intervention was

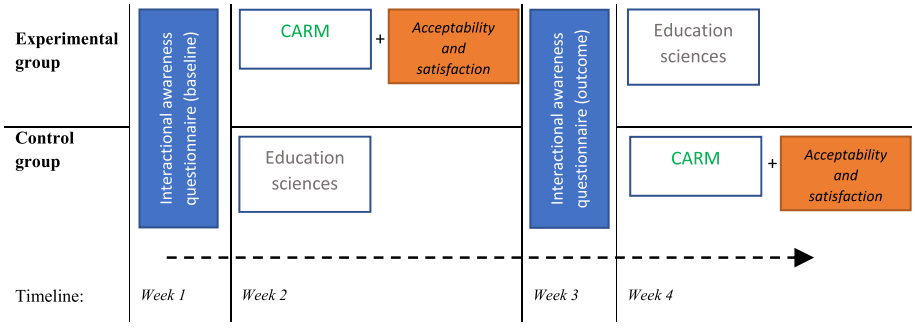

Fig. 1. Intervention components and timeline.

organized in four main sections, to capture the stages on the interactional racetrack of asking questions and providing feedback in the classroom and during oral exams. These four sections were (i) the initiation of an encounter, (ii) formulating a problem or a question, (iii) pursuing the answer of the problem or question, and (iv) close the encounter or sequence. Within each section there were a set of *trainables* (i.e., targeted communication skills). The trainables were developed in an iterative process, uncovering member practices involved in teachers eliciting and evaluating student answers (see Robinson & Heritage, 2014 on the 'pre-intervention' phase of interventionist CA), and were the basis for developing the interactional awareness scale and are summarized in the Appendix (see also Section 2.5).

In the CARM intervention, each trainable is presented using audio-visual transcripts of contrasting examples that are stopped at particular moments to target the audience's reflections on what is going on and what they expect the teacher (or students) to do next. Overall, CARM as communication training is designed to facilitate reflection and awareness on how language and interactional conduct affects the ensuing interaction.

The CARM intervention was delivered in one session, lasting approximately 3 h, with two 15 min breaks. The intervention was delivered using Microsoft PowerPoint presentations in front of the group of STs. The intervention was delivered by one trainer (RO) who is an expert in conversation analysis and who co-developed the training. He has 8 years of experience in delivering CARM training to multiple audiences. To ensure the fidelity of the delivery of the intervention, two other experts in conversation analysis and CARM training (KS & MS) observed the delivery of the intervention *in situ*, taking free notes. The notes were used to ensure that the trainer delivered the intervention as planned, but also to record any issues that arose during the delivery of the intervention to inform the refinement of the CARM training to be used in future interventions and trials.

The CARM intervention delivered in this trial and the outcome assessments were pilot-tested on 10 non-trial student teachers from the same university, and the feedback received was used to further refine the intervention and the choice and development of the outcome measure.

2.4. Control condition

Participants in the control condition (i.e., those not initially receiving the CARM training) received lessons in Educational science, according to the lesson plan. Students in the experimental group received the equivalent lesson in Education science when the control group took part in the CARM training (see Fig. 1).

2.5. Outcome assessment

As shown in Fig. 1, all STs completed the outcome assessment online, at baseline (week 1) one week *before* the experimental group received the CARM intervention (week 2) while the control group did the control condition (Educational science), and again one week *after* the intervention (week 3). All STs were sent a link by email to complete the questionnaire within a week, and everyone received a reminder to do so. On the day of the CARM intervention, both groups also completed an acceptability and satisfaction questionnaire, and a demographics questionnaire (see Section 3.1.1).

We developed and piloted a bespoke 9-item questionnaire (in Norwegian; see English translation in appendix) assessing participants' awareness of aspects to do with their pedagogical competence through interacting with students, in terms of how their practices affect trajectories of actions. For this study we defined and measured such competence as 'interactional awareness'. Interactional awareness has previously been assessed in the classroom context, regarding teachers' sensitivity to their role in particular stages of a lesson (Walsh, 2011). In our study, we applied a similar definition of interactional awareness, but targeting specific interactional environments in place of 'stages of a lesson'. Interactional awareness was thus implemented as an outcome assessment – a questionnaire - developed by the research team based on the key interactional practices presented as trainables (see Section 2.3).

The questionnaire was validated in a multi-stage process. First the trainables were developed and accepted by the research group, considering whether and how the questions corresponded to each trainable and whether they were conceived as answerable to the general student teacher. The CARM workshop and interactional awareness questionnaire was then piloted for an audience of student teachers. To further validate the questions the students provided feedback using 'talk aloud' methods, through which the pilot participants reasoned about their understanding of each question and how they would go about answering it. Through this exercise it would become apparent whether and how the participant understood the question and/or found it relevant to the workshop they had just observed. As the research group initially had considered self-efficacy as a construct of interest for outcome assessment, we included a modified version of the Norwegian Teacher Self-Efficacy Scale (Skaalvik & Skaalvik, 2007) in the pilot workshop. However, in the 'talk-aloud' sessions STs displayed considerable confusion about items related to self-efficacy and the research team felt that its inclusion might adversely affect completion rates of the questionnaire. Therefore, we opted to only assess aspects related to interactional awareness for evaluating the intervention. We regarded interactional awareness as more proximal to the communication skills training at hand, and self-efficacy a too distal construct for the current purposes.

In the questionnaire, STs were asked to indicate the extent to which they agreed with items on a Likert scale ranging from 1 ("to a very low extent") to 7 ("to a very high extent"). See Appendix for an overview of the questionnaire content and trainables.

Acceptability of and satisfaction with the CARM training were implemented to inform future work. Acceptability and satisfaction were assessed using a bespoke questionnaire conducted at the end of the training program for both groups. Questions addressed statements regarding (i) "the usefulness of the workshop for student teachers", (ii) "the organisation of the workshop", (iii) "the content of the workshop", and (iv) "the level of difficulty of the workshop", using a Likert scale ranging from 1 ("very much disagree") to 7 ("very much agree"). Participants completed this questionnaire anonymously and dropped them in a box for the assessors to collect later. We describe the feedback scores from the acceptability questionnaire using descriptive statistics. Open-ended (free-text) questions were included to invite more nuanced feedback. These questions addressed (v) "what worked well in the workshop and which I would recommend the trainers to keep doing", (vi) "what worked less well in the workshop and which I would recommend the trainers to change or stop doing", (vii) "what I thought was missing from the workshop and which I would recommend including in future training". We identified common themes using thematic analysis, which means that after organizing the answers into overarching themes, we singled out sub-themes that captured each overarching theme (see Clarke, Braun & Hayfield, 2015).

2.6. Data analysis

This study assesses the effectiveness of the CARM intervention on student teachers' interactional awareness. For the purpose of

evaluating the intervention, we compared the results from the interactional awareness questionnaire, comparing the scores of the nine question items at baseline (i.e., before either group received the CARM intervention) with those at outcome (i.e., following the experimental group - and not the control group - had received the CARM intervention; see Fig. 1 for an overview of the intervention timeline). We ran a One-Sample-Kolmogorov-Smirnov Test and inspected Q-Q plots to determine whether the data were normally distributed. Given that this is an ordinal-level measure and a multidimensional scale, assessing different aspects of interactional awareness, we decided not to group the data (e.g., through factor analysis), but to report the scores for each item separately. Statistical tests revealed it was not appropriate to calculate one total score for the questionnaire. The data were not normal distributed. Effect sizes (eta squared, η2) (see Thompson, 2006, pp. 317–319) were calculated by the open accessible tool provided by Lenhard and Lenhard (2016). The effect sizes were interpreted as small ≥0.01; medium >0.06; large ≥0.16; in accordance with rules of thumb.

We performed a statistical power analysis to provide information about the actual statistical power for this study (based on 0.8 power, meaning 80% certainty that the observed effects really exist). Average effect size based on the nine questions (see Table 2) was calculated and transformed to Cohen's d (Cohen, 1988), which we used in the statistical power analysis, applying G*Power version 3 (Faul et al., 2007).

Ethical approval was obtained from 02.01.2019 to 31.12.2028 (ref no. 448873).

3. Results

Our aim of the study was two-fold, (i) to explore the practicalities, acceptability, and satisfaction of delivering the intervention, and (ii) determine the efficacy of the CARM intervention to improve interactional awareness amongst student teachers compared to current training. We have followed the CONSORT guidelines for reporting feasibility and pilot trials (ref: https://www.equator-network.org/reporting-guidelines/consort-2010-statement-extension-to-randomised-pilot-and-feasibility-trials/), and according to these guidelines we present the findings on each of these components in turn, starting with the feasibility (Section 3.1), with reference to the CONSORT chart showing the flow of the participants through the trial (Fig. 2).

3.1. Feasibility

The evaluation of practicalities, acceptability, and satisfaction of delivering the intervention will be discussed through three sections *Recruitment and sampling (3.1.1)*, *Data collection and completion (3.1.2)* and *Acceptability of intervention (3.1.3).*

3.1.1. Recruitment and sampling

The recruitment process was successful: all participants received the relevant information (in class and by email) and the process resulted in almost all students engaging with the study. The sample comprised 41 student teachers (21 in the experimental group and 20 in the control group). Everyone in the class was allocated to either to the experimental or the control group. Both

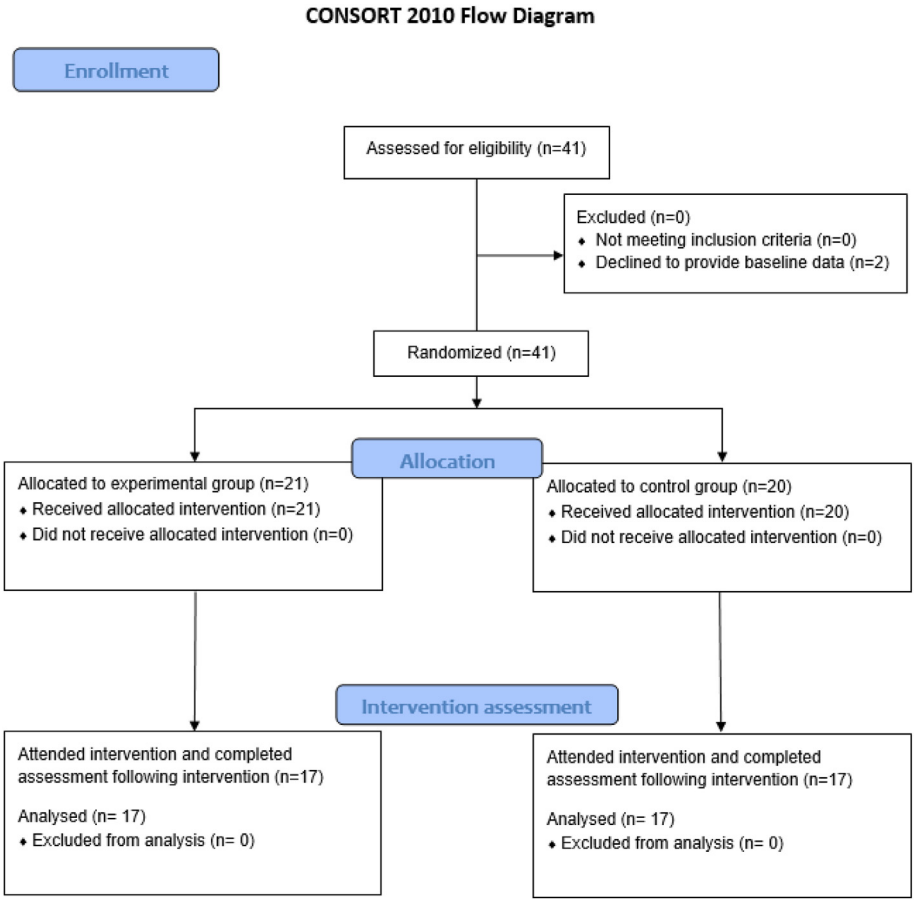

Fig. 2. CONSORT diagram showing the flow of the participants though the trial.

Table 1
Participant characteristics in the two trial groups.

	Experimental group	Control group
Total (n)	21	20
Women (n)	15 (71%)	17 (85%)
Men (n)	6 (29%)	3 (15%)
Mean age (SD)	30.6 (7.0)	25.8 (3.2)
Core subject mathematics (n)	13 (9 women, 4 men)	15 (12 women, 3 men)
Core subject Norwegian (n)	8 (6 women, 2 men)	5 (5 women, 0 men)

groups were relatively well balanced in terms of core subject composition (i.e., Mathematics or Norwegian). See the flowchart for the recruited participants summarized in Fig. 2.

Table 1 gives an overview of participant characteristics. Gender of participants was relatively balanced between experimental and control groups. However, the experimental group had twice as many men as the control group (6 vs. 3). This is perhaps a chance occurrence given the small sample size, which is likely to be rectified in a larger trial. The age distribution was between 20 and 44. All age groups were represented in both experimental and control groups, with somewhat higher average age and standard deviation in the experimental group.

3.1.2. Data collection and completion

36/41 (88%) of participants allocated to the intervention attended the session(s). All participants in the experimental group and all but two in the control group completed the baseline questionnaire. 17 participants (81%) in the experimental group and 17 participants (85%) in the control group completed the outcome measure. Participants were not required to give a reason for not responding.

3.1.3. Acceptability of intervention

The student teachers reported high levels of acceptability on a 7-point scale in relation to (i) "the usefulness of the workshop for student teachers" (M = 5.8, SD = 0.9), (ii) "the organisation of the workshop" (M = 6.4, SD = 0.9), (iii) "the content of the workshop" (M = 6.4, SD = 0.8), and, with some more variability in responses to (iv) "the level of difficulty of the workshop" (M = 5.4, SD = 1.7) (see Fig. 3). In addition, participants were invited to respond to three open-ended (free-text) questions: (v) "what worked well in the workshop and which I would recommend the trainers to keep doing", (vi) "what worked less well in the workshop and which I would recommend the trainers to change or stop doing", (vii) "what I thought was missing from the workshop and which I would recommend including in future training". All 36 participants who attended the intervention responded to the evaluation questionnaire. All participants responded to the four questions to be answered on a Likert scale. Most of the participants responded to the free-text questions as well: in the experimental group 2, 16/19 participants responded to all questions (including the free-text ones), and in the control group, 12/17 participants responded to all questions (including the free-text ones).

Based on the results of the survey, the participants understood the training content and found the training suitable and well-designed/organised. Responses in the free text option of the survey revealed that most of the participants found the training well-structured ("Nice and well-structured presentation. I liked the 'roadmap' metaphor") and easy to follow (e.g., "the trainer was pleasant to listen to, suitable tempo, and responsive"). One central sub-theme regarding the overarching theme of usefulness of the workshop, was the use of authentic data. The participants highlighted the use of 'real' examples and providing time for reflection and discussion as particularly useful. Participants commented there were "clear questions to guide group discussions", and described the use of authentic examples as important to "make the content close to actual practice". One participant wrote: "It is extremely valuable to see and hear authentic situations: it turns [the training] into extended practice". Some were more specific in expressing increased interactional awareness resulting from the training: "The training makes teachers more aware about how one gets the most out of encounters with students".

Another emerging theme was how the training gave a change in understanding: one participant wrote that although they had some previous knowledge on the concepts presented, such as the differentiation between 'closed-ended' and 'open-ended' questions, their understanding was refined and changed following the CARM training: "I now look differently upon the function closed and open questions", and "the course makes us more aware about how to get most out of interactions with students". These examples provide further evidence on the change in awareness following the training.

Participants wrote that the amount of instruction was appropriate, with suitable examples, which opened up opportunities for reflection and sharing ideas and strategies. Four participants wished for more time for reflection and discussion (e.g., "the course could have benefited from more time to discuss the examples provided in

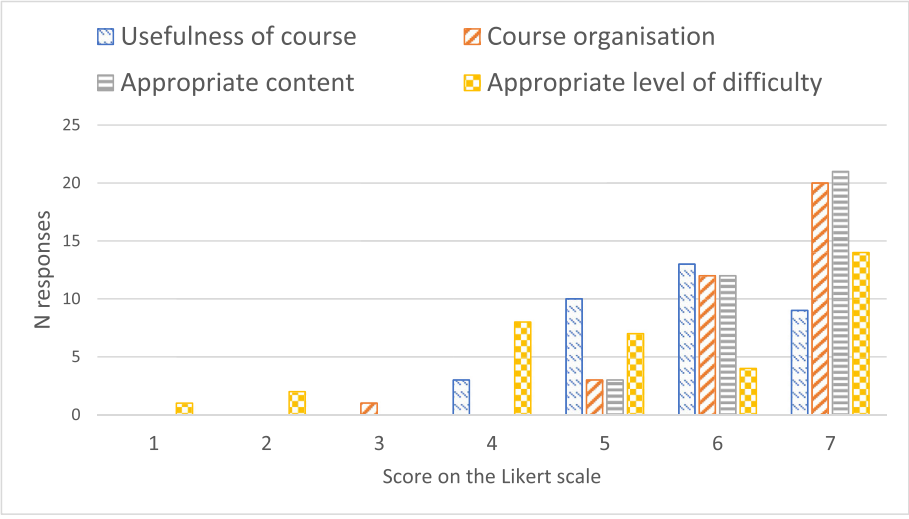

Fig. 3. Acceptability of the intervention.

groups"), and four sought more 'real' examples to discuss ways of engaging students in the next turn ("I wanted to see more examples of different ways of asking questions", "the groups could have analysed a conversation on their own to activate learning"). The opinions highlighting the need for more discussion and examples indicate that the reflective exercises provided were well received as part of the training, but that perhaps more time was needed to get familiarized with the materials, and to 'activate' learning as pointed out by one participant. In contrast, one participant sought more input and conclusions from the trainers, on what was best practice ("it was good to reflect on the examples, but I missed some sort of confirmation/guidance from the trainers on what was right"). This latter opinion (though only given by one participant) suggests there may be differing expectations amongst participants as to whether to understand the trainables as items of dos and don'ts, or as items of pedagogical practice for the participants to reflect upon.

Amongst further critical opinions, five participants highlighted the need for more references to relevant theory or literature, and four participants highlighted that the difficulty level was rather low, with comments that the content could have been more advanced, with "pedagogically more demanding situations". These comments perhaps reflects the more variable score on question (iv) in the questionnaire (appropriate level of difficulty). Further positive feedback within the theme 'further development' was received, including suggestions the training "could take place earlier in the teacher training programme", and that this form of instruction "ought to inform teacher education more broadly, not just as a free-standing course".

3.2. Efficacy

Table 2 presents the mean ranks for the experimental and control group and provides information about differences between the groups at baseline and intervention respectively.

At the baseline assessment, a Mann-Whitney *U* Test revealed no significant differences between the experimental and the control group on the nine items covering the concept of attentional awareness (see Table 2). Thus, we can assume that the two groups of student teachers did not differ systematically in relation to their attentional awareness before the intervention; thus, possible group-differences post-intervention may be attributed to the intervention.

After the intervention, as hypothesized, the experimental and control groups differed significantly (Mann-Whitney *U* Test, $p < 0.05$) on eight out of nine items. The item that did not show differences between two groups was item 7, "… how follow-up questions (e.g., "How is it?") can be used to support the student in deepening the answer?". The differences between the groups in the remaining eight items revealed relatively large effects (eta squared, $\eta 2 = 0.14$ - 0.37).

Furthermore, we explored the changes from baseline to outcome assessment for the experimental and control groups separately, to assess whether changes over time are associated with group affiliation. As Table 3 indicates, the experimental group had a

Table 2
Mean ranks for experimental and control group and tests of differences between the two groups at baseline (pre-intervention) and outcome (post-intervention) respectively (Mann-Whitney *U* Test).

		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Baseline (pre-intervention)										
Mean Rank	Experimental group	19.76	19.24	19.88	21.5	20.76	20.24	18.50	22.52	18.93
	Control group	20.28	20.89	20.14	18.78	19.11	19.72	21.75	17.06	21.25
Mann-Whitney U		184.00	173.00	186.50	167.00	173.00	184.00	157.50	136.00	166.50
Z		−.145	−.461	−.073	−.632	−.461	−.144	−.946	−1.515	−.650
p (2-tailed)		.885	.645	.942	.527	.645	.886	.344	.130	.516
Effect size (η^2)[a]		.054	.028	.062	.018	.028	.054	.006	.003	.017
Outcome (post-intervention)										
Mean Rank	Experimental group	23.56	21.18	22.38	22.56	21.32	21.85	19.88	23.50	23.00
	Control group	11.44	13.82	12.62	12.44	13.68	13.15	15.12	11.50	12.00
Mann-Whitney U		41.50	82.00	61.50	58.50	79.50	70.50	104.00	42.50	51.00
Z		−2.373	−1.724	−.929	−1.311	−.262	−.812	−1.026	−1.930	−1.069
p (2-tailed)		.000	.023	.002	.003	.023	.008	.141	.000	.001
Effect size (η^2)[b]		.370	.136	.240	.258	.147	.191	.057	.363	.305

Q1 to Q9 represents the 9 individual test items.
[a] Eta squared, between groups at baseline (pre-intervention);
[b] Eta squared, between groups at outcome (post-intervention);

Table 3
Examination of changes from baseline (pre-intervention) to outcome (post-intervention) (Wilcoxon-Matched Pairs Signed Rank Test) for experimental and control groups. Descriptive values (mean ranks) are presented in Table 2.

	Q1 T2-T1	Q2 T2-T1	Q3 T2-T1	Q4 T2-T1	Q5 T2-T1	Q6 T2-T1	Q7 T2-T1	Q8 T2-T1	Q9 T2-T1
Experimental group (n = 21/17)									
Z	−2.850	−1.926	−2.317	−3.139	−1.456	−2.285	−1.413	−3.222	−3.310
p (2-tailed)	.004	.054	.021	.002	.145	.022	.158	.001	.001
Effect size (η^2) [a]	.478	.218	.316	.580	.125	.307	.117	.611	.644
Control group (n = 20/17)									
Z	−2.373	−1.724	−.929	−1.311	−.262	−.812	−1.026	−1.930	−1.069
p (2-tailed)	.018	.085	.353	.190	.794	.417	.305	.054	.285
Effect size (η^2) [b]	.331	.175	.051	.101	.004	.039	.062	.219	.067

Q1 to Q9 represent the 9 individual test items.
[a] Eta squared, within experimental group from baseline to outcome;
[b] Eta squared, within control group from baseline to outcome;

statistically significant increase in six out of nine items, while the control group only revealed one significant change, a deterioration from baseline to intervention in the case of item 1 "… how the teacher interacts with the students to offer assistance with group work?".

Overall, the intervention shows a statistically significant effect on the outcome measure interactional awareness.

3.3. Power analysis

The average effect size of the nine question was 0.229 (eta squared), which is equivalent to a Cohen's d of 1.09 (Cohen, 1988). Using this value for Cohen's d in the statistical power analysis (Faul et al., 2007), we found an actual power of 0.807 for this study. Thus, despite the small sample size, the differences between groups at outcome (post-intervention) seem to be of sufficient credibility, indicating that the effect of the intervention exists.

4. Discussion

This article explored the practicalities of delivering and evaluating the efficacy of the communication training intervention CARM. The study was designed as a feasibility randomized controlled study, to (i) explore the practicalities, acceptability, and satisfaction of delivering the intervention, and (ii) determine whether a CARM intervention improves interactional awareness amongst student teachers compared to current training. This was the first time a CARM intervention has been evaluated in a randomized controlled trial.

The sample was representative of student teacher populations in Norway. Though the numbers are too small for statistical comparison, the overall gender balance was reasonably representative of the gender balance nationally for student teachers in Norway. Overall, the ratio is 78% women and 22% men, compared to 80.7% women and 19.3% men in the national statistic for 2021 (teacher education year 1–7 – Norwegian Directorate for Higher Education and Skills, 2021). The age distribution was also comparatively similar to the national statistic with approximately 70% of the students in their 20s, but with more students in their late 20s than in their early 20s compared to the national statistic (general Master's level - Norwegian Directorate for Higher Education and Skills, 2021).

All but two participants (who were allocated to the control group) completed the baseline assessment, 36/41 allocated participants attended the intervention, and 34 participants completed the outcome assessment. Because the data were anonymous and because participants did not need to explain why they did not wish to complete the assessment, we were unable to identify these participants and their reason for not completing. 88% of participants received the intervention and control conditions as originally planned. However, in our analyses, we have only used the data from those who completed both baseline and outcome assessments (83% of allocated participants). It appears that the assessment was acceptable to the participants. Future studies, however, may wish to consider making completion of baseline and outcome measures a prerequisite to participation to avoid having missing data.

The CARM intervention was implemented as intended, with the independent fidelity observers noting that all elements of the training were delivered. The fidelity observers also noted that the final part of the intervention, which covered the closing stage of the 'interactional racetrack' (see Appendix), was not given enough time, compared to the earlier stages of the racetrack. Therefore, in future CARM delivery, this needs to be attended to. CARM instructors could be provided with a clearer time schedule. For our study, we only had one trainer who delivered the CARM training. In a larger study or to deliver the intervention at scale, more trainers would need to be trained in delivering CARM. This, however, may create variability in the delivery of the intervention, and therefore, fidelity assessments would need to be implemented to ensure consistency of training across groups.

The student teachers demonstrated high levels of acceptability and satisfaction with the CARM training. Satisfaction scores combined with free-text feedback showed that student teachers were satisfied with the content and their learning from the training. They highlighted the relevance of using examples representing actual teacher practice, the importance of clear tasks for reflection and discussion, and they felt that the course was well presented and organized, which made it easy to follow. Overall comments were encouraging to continue to develop and use this form of communication training. However, participants also pointed out that there was not sufficient time to consider examples and phenomena in group discussion and reflection exercises. Based on this feedback, and a further way to evaluate fidelity of the CARM intervention, future CARM interventions could implement checklists for, e.g., how many examples participants should discuss, and for how long and/or how many interactional phenomena they should be able to identify and consider per trainable. The student feedback also suggested they had different expectations regarding the framework for the intervention and the extent to which this ought to be theoretically based. So far it has been implicit in CARM training that participants - through engaging with real life examples and relating them to their own (theoretical) knowledge and experience - may be able to consider the value of this knowledge and experience and perhaps view them in a new light. In future implementations of CARM interventions one may work on making this point even clearer, to support the participants to identify the very practices that research has shown to affect trajectories of action. Doing so would perhaps ensure the authenticity of the training materials are closely tied to reflective practices developed (see e.g., Waring & Creider, 2021).

In terms of the efficacy of the CARM intervention, we found a significant increase in the student teachers' interactional self-awareness on areas of conduct addressed in the training. The increase that was observed in the experimental group was not observed in the control group, therefore, we are relatively confident that this difference was caused by the intervention. The power analysis revealed, despite the small group size, that the trustworthiness of the findings is sufficient. Thus, an upscaling of the sample size is not required to claim with reasonable certainty that the results of this small-scale study are robust. Nevertheless, it may be useful to roll out the intervention on a larger scale and include various higher education programs in teacher education. This is to evaluate whether the CARM intervention is sustainable in real life, under varying conditions, and with a less focus on the feasibility of the implementation.

In this study, our measure of 'interactional awareness' address participants' awareness of their pedagogical competence through interacting with students, in terms of how their pedagogical practices affect trajectories of actions. Our results suggest that, though student teachers report some level of interactional awareness on how language and conduct influences teaching and assessment prior to intervention, this awareness significantly increased after the intervention. This increase in interactional awareness was also evidenced in the qualitative part of the evaluation (see Section 3.1.3), in which the participants reported how their conceptions relevant to pedagogical practice were better defined following the

CARM intervention, including the variability and use of different question formats. This is in line with previous approaches to evaluating CSTs in teacher training, showing the value of directing participants' attention to key skills in their own behaviour (e.g., Fukkink et al., 2011). Based on the overall results, we argue that the CARM training proves to be a highly appropriate and useful training method in teacher education.

Interactional awareness relevant to pedagogical practice also proves to be a useful construct to address and pinpoint interactional decisions and what their consequences are in educational contexts (see e.g., Walsh, 2011). The evidence base for what constitute interaction skills and awareness relevant to pedagogical practice was conversation analysis (CA), a method which has proved to be a rigorous and sustainable method for pinning down the building blocks of interaction, and promises to be key to future development and evaluation of interventions such as CST (see Robinson & Heritage, 2014, on 'interventionist CA'). Though well suited to unpack teachers' interaction practice *in situ*, few studies use CA in CST in teacher education so far. While interactional awareness is a relevant construct, and self-reported awareness is a valid and feasible measure for CSTs, the ultimate litmus test of the effectiveness of a CARM (or other CST) intervention is change in actual interactional practice, and outcomes. Future studies may indeed focus on changes in actual encounters, comparing interactional behaviours before and after the training, or in RCTs.

Interactional practices can be measured, however, reducing practice to 'effective' and 'ineffective' may deflect from developing core reflective practice, i.e., developing awareness of the variability of linguistic, interactional and pedagogical practices and how they may have different consequences to what happens next in the interaction. For example, rather than instructing teachers to "avoid closed questions", which is stereotypical rather than evidence-based, our training material demonstrated what constitutes different question types and their consequences in different activities and sequential environments. The student teachers developed their awareness and understanding of how their pedagogical competence builds on the flexible use of different interactional practices, rather than being advised to avoiding some (for example closed questions), and using other (for example open-ended questions). That there had been a shift in awareness on what constitutes pedagogical practice skills *in situ* was evident in the free-text responses from the students (see Section 3.1.3).

4.1. Limitations of study

For the purposes of evaluation, we compared the self-reported data from student teachers in the experimental and control groups at baseline (pre-intervention) and outcome (post-intervention) and the changes in their scores for each of the nine items respectively. According to our assumption, all items represent a common concept of interactional awareness, and, by that, are not independent. Thus, we have been cautious when interpreting the findings and do not consider the findings based on the statistical analyses to be conclusive. However, we believe that they provide a trustworthy indication of the possible effects of the intervention.

RCTs are considered the 'gold standard' for determining the effectiveness of interventions in healthcare settings (Bothwell et al., 2016), and this research design has found purchase in other fields also, including education (e.g., Gore et al., 2017; Hoogendijk et al., 2018), policy research (e.g., Shemilt et al., 2004) and has gained prominence in economics following the award of the 2019 Nobel prize to three 'randomistas' (Callaway, 2019). But while the RCT design has been adapted for non-medical fields, they are not without limitations, for example in education research, where control is harder to achieve than in medical research (see Sullivan, 2011). The delivery of a CST intervention may be sensitive to contextual factors not controlled for, and it is difficult to blind participants to their assigned group, even with crossover designs. Notwithstanding these limitations, RCTs are still a useful method to evaluate the effectiveness of CST interventions (see e.g. Gore et al., 2017). Future CARM training and evaluation of the CARM intervention may allow for more time for reflection and discussion on each trainable, and including a range of quantitative and qualitative outcome measures (self-reported and objectively measured behaviours), and with a larger cluster of participants.

The fidelity of the delivery of the intervention was assessed by two other experts in conversation analysis and CARM training *in situ*, when the intervention was being delivered. While this process revealed that there were no fidelity issues, it was time-consuming to evaluate and assess the delivery of the intervention. Therefore, when the study is expanded to a full trial with increased number of clusters, we may need to consider a more systematic way of documenting intervention fidelity to make such fidelity assessments quicker, and whether fidelity assessments can be completed using samples of video recordings of the delivery of the intervention. We may also need to develop other fidelity checklists as part of future trials, including what exercises the participants should do, and what they should achieve within each trainable. Such checklists would also be of use to other researchers in the field of interventionist CA (see e.g. Robinson & Heritage, 2014). A future definitive trial could use a step-wedge design (Hemming et al., 2015) to again ensure that all participants get the intervention – but within a random and sequential order, with clusters moving from control to experimental groups until all clusters receive the intervention. This will also enable a systematic implementation of the CARM intervention.

Overall, this study shows that the implementation of CARM as a training intervention is highly suitable for the training of student teachers, based on empirical findings from conversation analytic research on interactions *in situ*. Future teacher education may benefit from involving CARM as one way of practicing communication skills training side by side with other training methods, such as simulations and other ways of learning through practice – to ensure plurality of teaching method and innovation.

Funding

This work was supported by The Research Council of Norway (NFR) under grant 273417.

Data availability

Data will be made available on request.

Appendix

Bespoke 9-item questionnaire addressing interactional awareness

On a scale from 1 to 7, where 1 represents "to a very small degree", and 7 represents "to a very high degree", participants were asked to judge their self-awareness based on the 9 items listed in the left column of the table below. The middle column describes the analytic basis for each trainable as summarized in the CARM intervention. The right column shows where we put each trainable on the 'interactional racetrack' which structured the content of the training into four stages.

Items on the 9-item questionnaire addressing interactional awareness. *(To what extent have you reflected on …)*	Description of each trainable, as basis for interactional awareness questionnaire	Stage on 'interactional racetrack'
1. … *how a teacher interacts with students to offer assistance in group work?*	Awareness of different ways to be accessible for student assistance in the classroom: • A teacher may be accessible by standing by the student desk, or by standing/walking across the classroom.	Initiate
2. … *how reprimanding students affects the proceeding conversation?*	Awareness of how opening questions affect the ensuing classroom encounter: • Opening questions (e.g., "how is it going?") that opens for a problem formulation from the student without making them accountable for doing so. • Using process-oriented questions (e.g., "which topic have you chosen to work on?") compared to sanctioning student behaviour (e.g., "haven't you got started yet").	Initiate
3. … *how teachers use questions to influence the opportunities for students to answer?*	Awareness of how different questions give opportunities for different types of answers: • A question/problem can set expectations for a simple or more elaborate answer, depending on how it is formulated. • Some questions (e.g., "have you found [the relevant online resource]?") indicate what should be done next (e.g., "then you can start writing down some of the main points").	Formulate
4. … *how 'closed' question formulations (e.g., yes/no questions) can be an effective method for getting the conversation started?*	• Awareness of how question formats making relevant simple answers (e.g., "who wrote [a specified novel]?") constrains the type of answer but also gets students going with something they know, as a starting point for building a more elaborate answer.	Formulate
5. … *advantages and disadvantages with 'open' question formulations (e.g., "can you tell us about X …") in oral exams?*	• Awareness of how question formats making relevant a more elaborate answer (e.g., "tell me about X …") may put students in charge of defining an answer in their own terms. However, putting the students in charge may be a challenge for students who lack the competence, especially during oral exams.	Formulate
6. … *how some question formulations may highlight the student's lack of knowledge?*	• Awareness of how some question formats (e.g. "what do you know?", "what have you read about?") may expose the student's lack of relevant knowledge.	Formulate
7. … *how follow-up questions (e.g., "how?") can be used to support the student in deepening the answer?*	Awareness of the difference between claiming and demonstrating understanding: • Asking "why?" as a follow-up question gives the student the opportunity to elaborate, without challenging the student, especially when the student already has shown they know something on the subject. • Making students accountable for not (yet) having provided an account (e.g., "you need to explain why") are found in classrooms but not during oral exams.	Pursue
8. … *how breaking down questions into smaller pieces can help candidates who are struggling to answer questions during the oral exam?*	Awareness of how separating questions with longer answers into smaller parts can help students, especially during oral exams: • To avoid students ending up with short, one-word answers. • Productive follow-up questions build on what students have already shown they know. • Multi-unit questions may provide useful hints to students, but may also constrain the student's opportunity to expand upon an overall answer as they typically address the final question.	Pursue
9. … *how the conclusion of group discussions set the framework for further group work?*	Awareness of how the closing of a group talk conversation in the classroom may be *instructive* and *encouraging* for further work, but to different extent: • When an instruction for further work is less explicit, it may be harder to monitor progress, but it may also hand over more initiative to the students.	Close

References

Bothwell, L. E., Greene, J. A., Podolsky, S. H., & Jones, D. S. (2016). Assessing the gold standard—Lessons from the history of RCTs. *New England Journal of Medicine, 374*(22), 2175–2181. https://doi.org/10.1056/NEJMms1604593

Callaway, E. (2019). Randomistas' who used controlled trials to fight poverty win economics Nobel. *Nature*. https://doi.org/10.1038/d41586-019-03125-y

Carpenter, L. B. (2021). Supporting student–teacher development of elicitations over time: A conversation analytic intervention. *Classroom Discourse*, 1–19. https://doi.org/10.1080/19463014.2021.1946112

Clarke, V., Braun, V., & Hayfield, N. (2015). Thematic analysis. In J. A. Smith (Ed.), *Qualitative psychology: A practical guide to research methods* (pp. 222–248). Sage Publications Ltd.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.

Dalgarno, B., Gregory, S., Knox, V., & Reiners, T. (2016). Practising teaching using virtual classroom role plays. *Australian Journal of Teacher Education (Online), 41*(1), 126–154. https://doi.org/10.3316/informit.184882045640447

Darling-Hammond, L. (2006). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education, 57*, 120–138. https://doi.org/10.1177/0022487105283796

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146

Ferry, B., Kervin, L., Cambourne, B., Turbill, J., Puglisi, S., Jonassen, D., et al. (2004). Online classroom simulation: The 'next wave' for preserviceteacher education? In R. Atkinson, C. McBeath, D. Jonas-Dwyer, & R. Phillips (Eds.), *Beyond the comfort zone: Proceedings of the 21st ASCILITE conference* (pp. 294–302). Perth, Western Australia: Australian Society for Computers in Learning in Tertiary Education.

Fischetti, J., Ledger, S., Lynch, D., & Donnelly, D. (2022). Practice before practicum: Simulation in initial teacher education. *The Teacher Educator, 57*(2), 155–174. https://doi.org/10.1080/08878730.2021.1973167

Fukkink, R. G., Trienekens, N., & Kramer, L. J. (2011). Video feedback in education and training: Putting learning in the picture. *Educational Psychology Review, 23*(1), 45–63. https://doi.org/10.1007/s10648-010-9144-5

Gartmeier, M., Bauer, J., Fischer, M. R., Hoppe-Seyler, T., Karsten, G., Kiessling, C., … Prenzel, M. (2015). Fostering professional communication skills of future physicians and teachers: Effects of e-learning with video cases and role-play. *Instructional Science, 43*(4), 443–462. https://doi.org/10.1007/s11251-014-9341-6

Gisewhite, R. A., Jeanfreau, M. M., & Holden, C. L. (2021). A call for ecologically-based teacher-parent communication skills training in pre-service teacher education programmes. *Educational Review, 73*(5), 597–616. https://doi.org/10.1080/00131911.2019.1666794

Gore, J., Lloyd, A., Smith, M., Bowe, J., Ellis, H., & Lubans, D. (2017). Effects of professional development on the quality of teaching: Results from a randomised controlled trial of Quality Teaching Rounds. *Teaching and Teacher Education, 68*, 99–113. https://doi.org/10.1016/j.tate.2017.08.007

Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J., & Lilford, R. J. (2015). The stepped wedge cluster randomised trial: Rationale, design, analysis, and reporting. *BMJ: British Medical Association* (Vol. 350,, h391. https://doi.org/10.1136/bmj.h391

Hoogendijk, C., Tick, N. T., Hofman, W. H. A., Holland, J. G., Severiens, S. E., Vuijk, P., &

van Veen, A. F. D. (2018). Direct and indirect effects of Key2Teach on teachers' sense of self-efficacy and emotional exhaustion, a randomized controlled trial. *Teaching and Teacher Education, 76*, 1–13. https://doi.org/10.1016/j.tate.2018.07.014

Hovdenak, S. S., & Wiese, E. (2017). Fronesis: Veien til profesjonell lærerutdanning? *Uniped, 40*(2), 170–184. https://doi.org/10.18261/issn.1893-8981-2017-02-06

Hunt, S., Simonds, C., & Cooper, P. (2002). Communication and teacher education: Exploring a communication course for all teachers. *Communication Education, 51*(1), 81–94. https://doi.org/10.1080/03634520216497

Jahreie, C. F., & Ottesen, E. (2010). Learning to become a teacher. In V. Ellis, A. Edwards, & P. Smagorinsky (Eds.), *m cultural-historical Perspectives on teacher Education and development* (pp. 131–145). London: Routledge.

Kaufman, D., & Ireland, A. (2016). Enhancing teacher education with simulations. *TechTrends, 60*(3), 260–267. https://doi.org/10.1007/s11528-016-0049-0

Lenhard, W., & Lenhard, A. (2016). Computation of effect sizes. *Psychometrica*. Retrieved from https://www.psychometrica.de/effect_size.html.

Lindset, M., & Aune, I. (2020). Simulering som pedagogisk metode i lærerutdanning. *Skandinavisk tidsskrift for yrker og profesjoner i utvikling, 5*(1), 46–70. https://doi.org/10.7577/sjvd.3452

Lindstøl, F. (2016). Tenk hvis: om fiksjonserfaringer som utgangspunkt for refleksjon i lærerutdanningen. *Acta didactica Norge, 10*(3), 1–21. https://doi.org/10.5617/adno.2964

Ministry of Education and Research. (2008–2009). *Meld.St. 11. Læreren rollen og utdanningen*. Retrieved from: https://www.regjeringen.no/contentassets/dce0159e067d445aacc82c55e364ce83/no/pdfs/stm200820090011000dddpdfs.pdf.

Ministry of Education and Research. (2016–2017). *Meld. St.16. Kultur for kvalitet i høyere utdanning.* Retrieved from: https://www.regjeringen.no/no/dokumenter/meld.-st.-16-20162017/id2536007/.

Nesje, K. (2020). Virtuelle klasseromssimuleringer i lærerutdanningen. In E. Elstad (Ed.), *Lærerutdanning i nordiske land* (pp. 243–255). Universitetsforlaget.

Norwegian directorate for higher education and skills.(2021). Retrieved from: https://dbh.hkdir.no/tall-og-statistikk/statistikk-meny/indeks.

Ortega, G. J. L., & Fuentes, R. A. (2015). Communication skills training in trainee primary school teachers in Spain. *Journal of Teacher Education for Sustainability, 17*(1), 85–97. http://hdl.handle.net/10481/39418.

Robinson, J. D., & Heritage, J. (2014). Intervening with conversation analysis: The case of medicine. *Research on Language and Social Interaction, 47*(3), 201–218. https://doi.org/10.1080/08351813.2014.925658

Sert, O. (2013). 'Epistemic status check' as an interactional phenomenon in instructed learning settings. *Journal of Pragmatics, 45*(1), 13–28. https://doi.org/10.1016/j.pragma.2012.10.005

Shemilt, I., Harvey, I., Shepstone, L., Swift, L., Reading, R., Mugford, M., … Robinson, J. (2004). A national evaluation of school breakfast clubs: Evidence from a cluster randomized controlled trial and an observational analysis. *Child: Care, Health and Development, 30*(5), 413–427. https://doi.org/10.1111/j.1365-2214.2004.00453.x

Simonds, B. K., Lippert, L. R., Hunt, S. K., Angell, M. E., & Moore, M. K. (2008). Communication and diversity: Innovations in teacher education. *Communication Teacher, 22*(2), 56–65. https://doi.org/10.1080/17404620802040536

Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology, 99*(3), 611–625. https://doi.org/10.1037/0022-0663.99.3.611

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., … Moore, L. (2021). A new framework for developing and evaluating complex interventions: Update of medical research Council guidance. *BMJ: British Medical Association, 374*, n2061. https://doi.org/10.1136/bmj.n2061

Skovholt, K., Solem, M. S., Vonen, M. N., Sikveland, R. O., & Stokoe, E. (2021). Asking more than one question in one turn in oral examinations and its impact on examination quality. *Journal of Pragmatics, 181*, 100–119. https://doi.org/10.1016/j.pragma.2021.05.020

Stokoe, E. (2014). The conversation analytic role-play method (CARM): A method for training communication skills as an alternative to simulated role-play. *Research on Language and Social Interaction, 47*(3), 255–265. https://doi.org/10.1080/08351813.2014.925663

Stokoe, E., & Sikveland, R. O. (2017). The conversation analytic role-play method. In S. Pink, V. Fors, & T. O'Dell (Eds.), *Theoretical scholarship and applied practice* (pp. 73–96). Berghahn Books.

Sullivan, G. M. (2011). Getting off the "gold standard": Randomized controlled trials and education research. *Journal of Graduate Medical Education, 3*(3), 285–289. https://doi.org/10.4300/JGME-D-11-00147.1

Theelen, H., Van den Beemt, A., & den Brok, P. (2019). Using 360-degree videos in teacher education to improve preservice teachers' professional interpersonal vision. *Journal of Computer Assisted Learning, 35*(5), 582–594. https://doi.org/10.1111/jcal.12361

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York: Guilford.

Walsh, S. (2011). *Exploring classroom discourse; language in action.* London: Routledge.

Waring, H. Z. (2007). The multi-functionality of accounts in advice giving. *Journal of SocioLinguistics, 11*(3), 367–391. https://doi.org/10.1111/j.1467-9841.2007.00328.x

Waring, H. Z., & Creider, S. C. (2021). *Micro-reflection on classroom communication: A FAB framework*. Sheffield, UK: Equinox Publishing Limited.

White, S. J., Ward, K., & Hibberd, E. (2021). A pilot of modified Conversation Analytic Role-play Method for one-to-one clinical communication training. *Patient Education and Counseling, 104*(11), 2748–2755. https://doi.org/10.1016/j.pec.2021.03.023

Wiesbeck, A. B., Bauer, J., Gartmeier, M., Kiessling, C., Möller, G. E., Karsten, G., Ficher, M. R., & Prenzel, M. (2017). Simulated conversations for assessing professional conversation competence in teacher parent and physician-patient conversations. *Journal for Educational Research Online, 9*(3), 82–101. https://doi.org/10.25656/01:15302