

This article was downloaded by: [University Library Utrecht]

On: 14 January 2015, At: 04:43

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Scandinavian Journal of Educational Research

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/csje20>

Teachers' Questions and Responses during Teacher-Student Feedback Dialogues

Agaath Dekker-Groen^a, Marieke Van der Schaaf^a & Karel Stokking^a

^a Department of Educational Sciences, Utrecht University

Published online: 06 Aug 2014.



CrossMark

[Click for updates](#)

To cite this article: Agaath Dekker-Groen, Marieke Van der Schaaf & Karel Stokking (2014): Teachers' Questions and Responses during Teacher-Student Feedback Dialogues, Scandinavian Journal of Educational Research, DOI: [10.1080/00313831.2014.937359](https://doi.org/10.1080/00313831.2014.937359)

To link to this article: <http://dx.doi.org/10.1080/00313831.2014.937359>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Teachers' Questions and Responses during Teacher-Student Feedback Dialogues

Agaath Dekker-Groen, Marieke Van der Schaaf, and Karel Stokking
Utrecht University

In vocational education, students have to develop competences for reflection to self-regulate their development during their career. Students' reflection can be supported by teachers interacting with students and giving them prompts. In this study, 46 videotaped feedback dialogues of 23 teachers and their individual students were analysed. A total of 16 teachers participated in a training programme and 7 teachers formed a control group. The study aimed to investigate teachers' questions and responses and students' reactions, and the effects of the training programme. Analysis showed that, in the post-test, the trained group used a broader repertoire than the control group. The trained group, on average, used more hybrid prompts, deep reasoning questions, and prompt responses, which potentially stimulate students to elaborate and reflect.

Keywords: feedback dialogues, interaction, prompt questions and responses, reflection

In vocational education, students have to develop competences for continuous learning and development during their career. A competence is a capacity to enact specific combinations of knowledge, skills, and attitudes needed to adequately perform tasks in jobs (Lizzio & Wilson, 2004). Competences for continuous learning and development demand students to use self-regulation processes. These self-regulation processes consist of three cyclical phases: forethought, performance, and reflection (Zimmerman, 2002). Reflection is not only seen as a phase during self-regulation, but also as an essential skill for students to improve their learning. This article focuses on how to support reflection skills as part of nursing students' self-regulation processes.

The development of students' reflection skills can be stimulated by means of feedback dialogues (cf. Carless, 2006; Van der Schaaf, Baartman, Prins, Oosterbaan, & Schaap, 2011). We view feedback as information about how successful students' functioning is and how to bridge the gap between their actual performance and certain learning goals aimed at improving their competences (Sadler, 2010). During a feedback dialogue, both the teacher and the students have an active role. Students bring in arguments about how and why they acted in a certain way and the teacher poses questions and gives responses.

Agaath Dekker-Groen, Department of Educational Sciences, Utrecht University; Marieke Van der Schaaf, Department of Educational Sciences, Utrecht University; Karel Stokking, Department of Educational Sciences, Utrecht University.

Correspondence concerning this article should be addressed to Agaath Dekker-Groen, Utrecht University, Educational Sciences, Heidelberglaan 1, Utrecht, 3584 CS, Netherlands. E-mail: a.m.dekker@uu.nl

Through the exchange of information and discussion between the teacher and the students in the dialogue, students receive feedback that is meant to support their reflection (Lee, 2005). Feedback dialogues can be part of a classroom conversation or a one-to-one situation comparable with a tutor session.

The effectiveness of feedback dialogues depends on whether students use the received information to become more competent. The prerequisite is that students understand the feedback and interpret it accurately (Van der Schaaf et al., 2011). Prompting can facilitate this understanding (Davis, 2003; Ifenthaler, 2012; Thillmann, Küsting, Wirth, & Leutner, 2009). In feedback dialogues, prompts are basically teachers' questions and responses to students. Prompts can induce learning strategies for self-regulation, including reflection skills (Gé & Land, 2004; Nückels, Hübner, & Renkl, 2009).

Little is known about how teachers pose questions and give responses in feedback dialogues to support a student's reflections. Getting more insight into the use of categories and sequences of questions and responses in feedback dialogues and the effects of this is relevant because these dialogues are an important pedagogical tool in educational settings. Moreover, results about effective teacher behaviour can be used to train teachers in posing questions to students in feedback dialogues (Branch & Paranjape, 2002).

Teachers' Prompts and Students' Self-Regulation

In feedback dialogues that aim to support the development of students' reflection skills, teachers and students exchange information and teachers support students when necessary. Such dialogues are prone to the so-called "assistance dilemma" in teacher-student interaction, pertaining to the degree of external support in relation to the student's degree of self-regulation in learning how to succeed (Ifenthaler, 2012; Koedinger & Aleven, 2007). If students experience difficulties in their learning, the teacher's additional information and support can be helpful but should be in balance with the students' need to figure out insights and solutions themselves (Shuell, 1996; Thompson, 2009). Self-regulation skills can help students to figure out things themselves. According to Zimmerman (2002), self-regulation concerns setting goals, using learning strategies to achieve these goals, monitoring one's performance for progress, restructuring the context and managing one's time, attributing causation to results, and adapting future strategies. In his model of self-regulation, Zimmerman distinguishes three cyclical phases: forethought, performance, and reflection. In each phase of the model, teachers can support students by giving prompts in the form of questions and responses, which can be rather generic or specific (Davis, 2003; Ifenthaler, 2012). The content of prompts should be related to the processes, which should be activated. Therefore, Thillmann et al. (2009) distinguish between generic and specific prompts. Generic prompts are effective in focusing students' attention to content and strategies they already possess but do not adequately use (production deficiency), for example, "Which step can help to solve this problem?" Specific prompts contain contextualised directed information, for example, "You may use the protocol" given in case of a lack in knowledge or skills for self-regulation (availability deficiency). Prompts can be very effective by focusing students on certain cognitive processes and evoking specific reflective activities (Berthold, Nückles, & Renkl, 2007; Butler, 1998; Van den Boom, Paas, & Van Merriënboer, 2007). In feedback dialogues, prompts are often given to direct students' reflection. Examples of such prompts are: "How does this affect you in your functioning?", "Why is this so important for you?", "What else can you do in this situation?"

Questions and Responses in Feedback Dialogues

In feedback dialogues, teachers' questions and responses may constitute support for students' reflection. Students are challenged the most by teacher-led conversations but not by teacher-dominated ones (Chin, 2006). To enhance teacher-student interaction, teachers should give students enough room to talk (Blatt, Confessore, Kallenberg, & Greenberg, 2008). They also should pose questions. The function of a question is to elicit a verbal response from the person to whom the question is addressed (Kearsley, 1976). The student determines how a question is conceived and which response will be given. As a consequence, there is not a one-to-one relation between a question and response. The response depends on the social context and the foregoing questions and responses in the dialogue (Thomas Farrar, 1986). To promote students' reflection, teachers can pose questions as follow up on clues they notice in students' thoughts or feelings (Branch & Paranjape, 2002). Therefore, we assume that in feedback dialogues to support nursing students' reflections, the teachers' questions on students' responses may engage students and elicit self-feedback that can support their reflection skills. Question and response categories that are often used in dialogue studies in medical and other educational contexts might be useful as categories in this study.

In a feedback dialogue, a response can be an answer to a question, a statement, or a comment (Chin, 2006). Each of these three response categories can be positive, neutral, negative, or corrective (Blatt et al., 2008). In dialogues, both the teacher and the student can react with an acceptance (positive), a denial (negative), or a neutral response. The fourth option, a corrective response, pertains to the content of feedback and corresponds with corrective feedback (e.g., Chi, 1996). In addition, teachers can give prompt responses, for example, a hint to activate students (e.g., Davis, 2003).

Questions in feedback dialogues can be explicitly stated as a question, for example: "Did you consider that?" or as a statement that is posed as if it was a question, for example: "Then you did not think about it?" Chin (2006) used a categorisation of open and closed questions, based on the function of the questions. According to Chin, questions are closed if they focus at information gathering, while open questions are meant to stimulate a dialogue or a discussion or to get the receiver's view. Chin's categorisation links two categories that are often used in learning situations. One category concerns questions aimed at the recall of information (lower-order questions) and the other concerns questions that focus on the development of a critical thinking process (higher-order questions). More differentiated than Chin's dichotomy of "lower-higher", are the categories in Bloom, Engelhart, Furst, Hill, and Krathwohl (1956) taxonomy, elaborated by Krathwohl (2002): remember, understand, apply, analyse, evaluate, and create. Teachers' questions may support students in cognitive, affective, and regulative processes.

Graesser and Person (1994) used the Graesser-Person-Huber question scheme, in which the names of question categories refer to the kind of responses a teacher expects. Besides requests and hybrid questions, they distinguished short- and long-answer questions. An example of a request is: "Please tell me about that". A hybrid question combines two or more question categories. For example the question: "Was it difficult for you?" is a closed question, which can be answered with "Yes" or "No" (short answer), but can also be intended to stimulate reasoning as a result of which a student may be encouraged to give a longer answer. Questions that are useful to get a short answer (so called "short-answer questions") are, for example, verification questions, such as: "Did you do the assignment?" and questions about quantitative data, such as: "How many days did you miss of your internship?" It is likely

that students rely on their memory (e.g., recall of information) to answer such short-answer questions (Chin, 2006). As this need for recall clarifies the function of the category, instead of the name “short-answer question”, from now on we will use “recall question”.

Graesser and Person (1994) spoke of “long-answer questions” when an answer was expected to be longer than one word or one sentence, for instance questions that focus on how something happened or what made that the student performed well. Within long-answer questions, they distinguished two categories. The first category concerned questions posed to get a definition, an example, a comparison, an interpretation, or a judgment. The function of this question category seems to be diagnosing students’ knowledge and getting their opinions in cases where little reasoning is necessary, for example, “How are you going?” Therefore, we will use “surface reasoning” for the first long-answer question category. The second category concerned questions posed to get information about cause and effect, purposefulness, or expectations, and instrumental or procedural questions. These questions intend to let students elucidate patterns of deep reasoning. From now on we will use “deep reasoning” for the second long-answer question category. We suppose that deep reasoning questions, for example, “How do you plan to handle this next time?”, aim to evoke reflection because they intend to stimulate thinking about experiences and processes in order to learn from them.

Considering the potential benefit of prompting for self-regulated learning in interactive learning environments (Ifenthaler, 2012; Thillmann et al., 2009), a combination of prompt responses and questions and certain of Graesser and Pearson’s (1994) question categories might be fruitful to induce students to use learning strategies that they do not spontaneously use (Berthold et al., 2007; Gé & Land, 2004).

Sequences of Teacher and Student Questions and Responses

In the literature about dialogues, different types of sequences are distinguished concerning the function that questions and responses can fulfil. An example is the sequence: teacher Initiation, student Response, teacher Evaluation (the IRE sequence) (Van Zee & Minstrell, 1997). In this sequence, teachers’ information-seeking questions are often meant to understand students (Flammer, 1981). Teachers can also start a dialogue to give feedback. In that case, the sequence is: teacher Initiation, student Response, teacher Feedback/Follow-up (the IRF sequence) (Chin, 2006; Mercer, 2004). When this sequence starts with a question by the teacher, followed by an answer of a student, the sequence is: teacher Question, student Answer, teacher Feedback (the QAF pattern).

In the studies referred to above, the focus was on the teachers’ or tutors’ behaviour in a lesson situation and on questions related to the content of a course, and not to the students’ reflection. In our study, the focus is on teachers’ behaviour in feedback dialogues with individual students. Both the teacher and the student may pose questions and give responses. The dialogue is a process in which teacher utterances and student utterances follow each other continuously. According to Flammer (1981), a teacher’s utterances can be seen as interventions because the teacher uses the student’s information to pose questions, for example: “Which alternative solutions might be successful in such circumstances?” Moreover, teachers give responses to support students’ reflection.

To get an insight into the role of teachers in feedback dialogues, we will analyse two types of sequences: two-sequence (Teacher utterance-Student utterance, to be abbreviated as T-S) and three-sequence (Teacher utterance-Student utterance-Teacher utterance, T-S-T). For

both, the teacher's and the student's utterances question categories and response categories will be distinguished.

Context, Aim, and Research Questions

The study was carried out in the context of senior-secondary vocational education for nursing in the Netherlands. The goal of Dutch senior secondary vocational education is to deliver students to the labour market as well as to prepare and stimulate them to continue their educational career in higher professional education. This study aims to investigate: (1) teachers' and students' questions and responses within feedback dialogues intended to stimulate students' self-regulation and competence development, and (2) the effects of a training programme on teachers' repertoire of questions and responses.

Feedback dialogues between a teacher and a student are used to study the interaction. In this one-to-one situation, other students cannot interfere on the dialogue. The research questions are: Which questions and responses are given in feedback dialogues between nursing teachers and individual students? What are the effects of the training programme on the questions and responses of teachers and individual students in feedback dialogues?

Methods

Design

In the study, a quasi experimental pre-post test design was used, including an experimental group and a control group to explore teachers' questions and responses during feedback dialogues and to examine the effects of a teacher training programme in using prompts, posing deep reasoning questions, and giving feedback (Dekker-Groen, Van der Schaaf, & Stokking, 2013). Both the pre-test and the post-test consisted of the registration of a feedback dialogue of a teacher with an individual student. The pre-test was prior to the execution of a teacher training programme and the post-test after completion of a programme. The time lag between the pre-test and post-test was between 12 and 15 months. The experimental group participated in a training programme. From now on we will call this group the 'training group'.

The intervention is a teacher training programme in which teachers learn to pose students questions and give prompts and feedback to stimulate students' reflection skills. The teacher training programme for supporting nursing students' reflection skills development consisted of group meetings, periods of independent practice, and individual meetings.

In three group meetings of 90 minutes each, information and instruction were given about the function of questioning, general and specific questions, question categories, and their relation with feedback and reflection. Moreover, video fragments of representative feedback dialogues between teachers and students were used as a means for the teachers to discuss with colleagues and to exchange experiences (Van Es & Sherin, 2008). Each group meeting was followed by a period of independent practice of new behaviour in interaction with students. In the three individual meetings of 45 minutes each, video interaction analysis was used to give teachers insight into their actions (Roter et al., 2004). The researcher, also a video interaction coach, discussed with each teacher video fragments of the teacher's feedback dialogue with a student, and guided the teacher to reflect on experiences and to formulate new goals. After each individual meeting, there was another period of independent practice.

Participants

Six schools of senior secondary nursing education participated in the study. These schools were spread over the country. A purposeful sample of teachers for the training group was composed based on the selection criteria of being interested in developing their own competences and having enough time to participate in the training. For both the training group and the control group, further selection criteria were: being responsible for teaching and guiding students for a minimum number of 20 hours per week, and not being close to retirement. The training group consisted of 12 female and 4 male teachers, on average 51 years old, and with 13 years of teaching experience. The seven teachers in the control group were all female, on average 47 years old, with 16 years of teaching experience. The mean differences in age and experience between the two groups were not significant. The teachers of the control group worked on other locations of the same schools as the teachers of the training group. The control group did not receive any information about the research questions nor about the training programme and did not have contact with the teachers from the training group. Students who participated in the study were in their second year of nursing education, mostly female (79% or more per group), and, on average, 20 years old. Students in the pre-test did not significantly differ in age, gender, and type of education compared to the students in the post-test.

Instrumentation

Coding schemes.

Based on the literature described above, a coding scheme was developed for coding teachers' and students' questions and responses in the feedback dialogues. Also, a coding map was developed in which we included a description and an example of each code and the coding rules. For example, the context of questions and responses had to be taken in account. Due to this coding map, it was clear what was understood by a code and when a code had to be used (Chi, 1997).

Question categories in both the teacher and the student coding scheme were based on the Graesser-Person-Huber scheme described above (Graesser & Person, 1994). In the teacher scheme, these question categories were combined with prompt questions (Berthold et al., 2007; Davis, 2003; Gé & Land, 2004; Ifenthaler, 2012). Seven categories were distinguished: (1) Recall, (2) Hybrid, (3) Hybrid prompt, (4) Request, (5) Surface reasoning, (6) Deep reasoning, and (7) Reasoning prompt. The decision of coding quite a number of different teacher question categories is in line with Thomas Farrar (1986), who states the need to match the complexity of questions as well as their meanings in the social context and their relations with the kind of dialogue.

As response categories, the following three coding categories were used for both teachers' and students' utterances: Accept (positive), Deny (negative), and Neutral response (Blatt et al., 2008). The Neutral response was also used when an affirmation or a critical remark was made less strong (e.g., "I agree but ..."). In the teacher coding scheme, the category Prompt was added (see, e.g., Berthold et al., 2007; Davis, 2003; Gé & Land, 2004; Nückels et al., 2009). An example of a Prompt response is: "So you will have the propensity to take responsibilities but people around you also will." See Appendix 1 for the coding scheme.

Segmentation and inter-rater agreement.

All videotaped feedback dialogues were transcribed verbatim (Mercer, 2004). Next, the transcripts were imported in the program Multi Episode Protocol Analysis (MEPA) (Erkens, 2005). Speech turn-taking was used as the segmentation criterion, because this fits both the natural course of the conversation and the research questions (Chi, 1997). Moreover, turn-taking is a unit of segmentation that can be objectively applied. All words spoken in one turn will be called an utterance, here to be called a segment. Each segment of a teacher and of a student was coded with a question category or a response category.

Two research assistants independently coded 100 and 102 segments randomly chosen from three conversations (in total 302 allotted codes, 150 to teachers' segments and 152 to students' segments), each of a different feedback dialogue between a teacher and a student who did not participate in the study but were from the same schools as in the sample for the study.

Inter-rater agreement and percentage agreement were computed. Cohen's Kappa for the teachers' segments was 'fair' (.56) and for the students' segments 'good' (.74) (Banerjee, Capozzoli, McSweeney, & Sinha, 1999; Strijbos & Stahl, 2007). Percentage agreement of the allotted codes for the teachers was 63%, and for the students 85%.

Procedure and measures.

Two feedback dialogues per teacher were videotaped, one before and one after the intervention. In the training group, data gathering took place in the summer of 2009 and in the summer and autumn of 2010. In the control group, for each teacher, the length of the period between the two dialogues was similar. All teachers received the same written request to videotape a feedback dialogue with a student who participated in their mentor group. The teachers asked the students' consent about the use of the videotape for research purposes.

In both teacher groups (training and control), only one feedback dialogue in the pre-test concerned feedback on how successful a practical task in the curriculum was, all the other dialogues were about the student's progress in the study and during the internship. All teachers used the student's portfolio or a written reflective report as input for the feedback dialogue. In the post-test the feedback dialogues of all teachers were about the progress in the study and the internship. The videotaped feedback dialogues in the training group lasted, on average, 18 minutes in the pre-test and 24 minutes in the post-test. For the control group, this was 23 minutes and 27 minutes, respectively.

Means and standard deviations of the number and duration of utterances in the videotaped feedback dialogues were computed. Differences in number and duration of utterances and in words between the training group and the control group were computed and tested for statistical significance using two-tailed independent *t*-tests, and within each condition between pre-test and post-test by using two-tailed paired *t*-tests. In addition, for the significant results, the effect sizes were computed.

The research assistants independently coded in MEPA all 46 feedback dialogues of 16 teachers in the training group and 7 teachers in the control group. This concerned a total of 10,185 utterances. Utterances could be questions or responses. Next, the frequencies of the teacher utterances and the student utterances were computed separately for the training group and the control group. Percentages were also calculated to control for differences in

number of utterances in the dialogues between the groups and within both groups between pre-test and post-test.

Sequences of two utterances were analysed to determine how often a certain teacher utterance was followed by a certain student utterance (the T-S sequence). In MEPA, four different scores of these T-S sequences were computed: Fr-score (succession frequency), E-score (expected succession frequency), Z-score, and K-score (Kappa score). The Z-score was computed to determine whether there was a statistically significant difference between the frequency and the expected value of a certain student utterance that followed on a certain teacher utterance (Jeong, 2005), and K-scores are corrected Z-scores. In the computation of the K-score, a correction is made for differences in the number of segments within and between groups, resulting in scores between -1 and 1 (Wampold & Margolin, 1982). Negative K-scores indicate that the sequence occurs less often than expected, positive scores indicate that the sequence occurs more often than expected.

Additionally, analyses of sequences of three utterances were computed: Teacher utterance-Student utterance-Teacher utterance (the T-S-T sequence) (Chin, 2006; Mercer, 2004). See [Appendix 2](#) for an example of how 17 successive utterances were divided into sequences and coded.

Some utterances and many sequences occurred infrequently. Therefore, we decided to use as threshold for further analysing and reporting results a minimum total frequency of 16 in the training group and of 7 in the control group. So we ensured that in both groups, on average, the sequence occurred at least once per dialogue. This threshold should be reached on at least one of the two moments of measurement (pre-test and post-test).

To analyse the effects of the training, Ancova was used to control for differences in pre-test percentages of teacher utterances between the training group and the control group (Huck & McLean, 1975). Ancova was also used for analysing differences in percentages of T-S sequences and T-S-T sequences. These analyses on the differences between the two groups were preceded by a bootstrap procedure because the sample was small and the data were not normally distributed (Field, 2009). Bootstrapping was carried out by creating 2000 bootstrap samples with replacement. Besides the F value of the Ancova, both the p value and partial η^2 of the bootstrap were computed.

Results

Time Frame, Turns, and Words in the Dialogues

The time frame of the videotaped feedback dialogues in the training group in the pre-test was 6–31 minutes and in the post-test 13–38 minutes. For the control group, this was 8–36 minutes in the pre-test and 15–43 minutes in the post-test. Two-tailed paired t -tests only revealed a significant difference between pre-test and the post-test of the training group ($t [15] = -2.87, p = .012, \eta^2 = .36$). Due to the choice for turn-taking as the segmentation criterion, the numbers of teacher and student turns are (almost) the same. The teachers always initiated the dialogues. The number of teacher turns in the dialogues in the training group in the pre-test was 38–300 and in the post-test 75–517. Two-tailed paired t -tests on the number of teacher turns only revealed a significant difference between the pre-test and the post-test of the training group ($t [15] = -2.77, p = .014, \eta^2 = .34$) (see [Table 1](#)).

Two-tailed paired t -tests on the average number of words the teachers in the training group used per turn between pre-test and post-test revealed a significant difference ($t [15] = 2.13,$

Table 1

Mean and Standard Deviation of Time Frame, Number of Turns, and Number of Words in the Dialogues

| | Pre-test | | | | Post-test | | | |
|----------------------------------|----------------|-----------|---------------|-----------|----------------|-----------|---------------|-----------|
| | Training group | | Control group | | Training group | | Control group | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Time frame in minutes | 17.4 | 7.5 | 22.6 | 10.6 | 24.1 | 8.5 | 26.6 | 8.9 |
| Number of turns | 155.4 | 72.2 | 254.6 | 171.9 | 249.6 | 131.1 | 274.7 | 120.3 |
| Number of teacher words per turn | 19.2 | 22.1 | 15.3 | 19.5 | 14.4 | 16.2 | 17.0 | 22.2 |
| Number of student words per turn | 17.1 | 22.8 | 14.2 | 21.5 | 18.2 | 26.2 | 15.1 | 23.2 |

$p = .04$, $\eta^2 = .23$). The number of words per minute used by teachers of the training group in the post-test (not shown in Table 1) was 9 fewer than in the pre-test (on average 85 words in the pre-test and 76 in the post-test), for their students in the post-test it was 21 words more than in the pre-test (on average 75 words in the pre-test and 96 words in the post-test). The teachers and students of the control group used about the same number of words in the pre-test and in the post-test, the teachers on average 86 words in the pre-test and 88 words in the post-test, and their students on average 80 words in the pre-test and 78 words in the post-test.

Teacher and Student Utterances during Feedback Dialogues

For the coding of the teacher utterances, in total 11 categories were distinguished (see Appendix 1). See Table 2 for the frequencies and percentages of these teacher categories. In the pre-test, the number of questions the teachers of the training group used was about the same as the number of responses (utterances other than questions), of which they used Recall questions and Accept responses the most. In the post-test, 48% of their utterances were questions and 52% were responses. A comparative analysis of the percentages of pre-test and post-test showed that, in the post-test, teachers of the training group used especially more Prompt and Deep reasoning questions and Prompt and Neutral responses, and fewer Recall, Surface reasoning questions, and Accept responses.

In the control group, 25% of teachers' utterances in the pre-test were questions, mostly Recall questions, and 75% were responses, of which the most were Neutral. In the post-test, 31% of their utterances were questions and 69% were responses. A comparison of the percentages on teachers' utterances of pre-test and post-test in the control group showed that the differences in question categories were very small. This did not count for Recall questions (pre-test: 14.7%, post-test 21.7%).

For the coding of the students' utterances, in total seven categories were distinguished (see Appendix 1). In both the pre-test and the post-test, the students of both groups in the experiment hardly posed questions. Only 2.2–5.3% of students' utterances were questions; three to seven questions per dialogue. These questions were Recall questions (82% or more) and Surface reasoning questions. Students mostly used Neutral and Accept responses and hardly any Deny responses.

Table 2

Frequencies^a, Percentages, and Ancovas with Bootstrap of Teachers' Utterances, Post-test Mean% and SD Corrected for Pre-test Differences

| Teacher utterance | Pre-test | | | | Post-test | | | | | | | | |
|-------------------------------|-----------|-------|-----------|-------|-----------|-------|-----------|-------|-------------------------------------|------------------------------------|-----------|----------|----------------------|
| | Training | | Control | | Training | | Control | | Training <i>M%</i> (<i>SD</i>) | Control <i>M%</i> (<i>SD</i>) | Bootstrap | | |
| | Frequency | % | Frequency | % | Frequency | % | Frequency | % | | | <i>F</i> | <i>P</i> | <i>Part</i> η^2 |
| Teacher question | | | | | | | | | | | | | |
| Recall | 230 | 18.4 | 131 | 14.7 | 269 | 13.4 | 209 | 21.7 | 13.38 (5.69) | 23.16 (4.15) | 15.61 | .00* | .44 |
| Hybrid | 124 | 9.9 | 37 | 4.2 | 191 | 9.6 | 33 | 3.4 | 9.62 (5.17) | 3.50 (2.92) | 5.02 | .03** | .20 |
| Hybrid prompt | 63 | 5.0 | 11 | 1.2 | 114 | 5.7 | 8 | 0.8 | 6.44 (3.57) | .94 (0.61) | 13.06 | .00* | .41 |
| Request | (7) | (0.6) | (1) | (0.1) | (13) | (0.7) | (0) | (0.0) | | | | | |
| Surface reasoning | 122 | 9.8 | 32 | 3.6 | 135 | 6.8 | 48 | 5.0 | 7.96 (8.42) | 6.07 (5.52) | 0.46 | .49 | .02 |
| Deep reasoning | 71 | 5.7 | 8 | 0.9 | 173 | 8.7 | (4) | (0.4) | 9.43 (4.42) | .88 (1.76) | 9.12 | .00* | .35 |
| Reasoning prompt | (11) | (0.9) | (5) | (0.6) | 69 | 3.5 | (0) | (0.0) | | | | | |
| Teacher response ^b | | | | | | | | | | | | | |
| Neutral | 176 | 14.1 | 371 | 41.6 | 463 | 23.1 | 389 | 40.4 | 22.0 (8.17) | 38.0 (6.24) | 18.66 | .00* | .48 |
| Deny | (2) | (0.2) | 9 | 1.1 | (2) | (0.1) | 10 | 1.0 | | | | | |
| Accept | 356 | 28.5 | 261 | 29.3 | 412 | 20.6 | 251 | 26.0 | 18.22 (9.58) | 24.89 (7.49) | 0.62 | .40 | .03 |
| Prompt | 87 | 7.0 | 25 | 2.8 | 160 | 8.0 | 12 | 1.2 | 7.86 (5.30) | 1.70 (2.18) | 3.96 | .01** | .18 |
| Total ^c | 1229 | 98.4 | 885 | 99.4 | 1986 | 99.4 | 960 | 99.5 | | | | | |

Notes: ^a Frequencies below the threshold (training group ≥ 16 , control group ≥ 7) and their percentages in parentheses; ^b teacher utterance other than a question; ^c total of questions and responses without numbers between parentheses. * $p < .01$, ** $p < .05$.

T-S Sequences.

Each T-S sequence consists of a pair of two utterances (questions or responses), starting with a teacher utterance followed by a student utterance. Eleven different teacher categories and seven different student categories allowed for 77 different sequences. Of these 77 sequences, 16 occurred on average at least one time per dialogue in the training group (the threshold), of which 15 sequences in both the pre-test and the post-test and one sequence only in the post-test. For the control group, this concerned 11 of these 15 sequences in both the pre-test and the post-test. The 16 sequences of the training group and the 11 sequences of the control group that reached the threshold consisted of combinations of nine respectively six teacher categories with mainly Student Accept and Neutral responses, and one Student Deny response. See Table 3 for the 15 T-S sequences that included a Student Accept or a Student Neutral response. Besides frequencies (Fr) per sequence per group (training and control) per measurement moment (pre-test and post-test), the expected values (row total of nine rows x column total of two columns/overall total) (E), Z-scores (Z), and their adjusted scores (K: Kappa values) are also given.

In both the pre-test and post-test of both groups, the sequence Teacher Recall-Student Accept appeared more than expected and the sequence Teacher Recall-Student Neutral was found less than expected. In the training group in the pre-test, both Teacher Surface and Deep reasoning questions, and still more in the post-test, these two reasoning questions and, moreover, Teacher Reasoning prompts occurred more than expected, in combination with Student Neutral response. For the control group, this was only the case for Surface reasoning.

In both the pre-test and post-test, the training group showed the Teacher Neutral-Student Accept sequence the most, and more than expected. In the post-test, the training group showed two other sequences much more than expected, namely Teacher Accept-Student Neutral and Teacher Prompt-Student Accept, of which the first two sequences were shown about three times as often as the third. In both the pre-test and the post-test, the control group showed the Teacher Neutral-Student Accept sequence far the most frequent, followed by Teacher Accept-Student Neutral. These teachers used Prompt responses less than expected in the post-test.

T-S-T Sequences.

Each T-S-T sequence consists of three utterances (questions or responses), starting with a teacher utterance followed by a student utterance, and ending with a teacher utterance. In the pre-test of the training group, 20 different T-S-T sequences, together 514 sequences, occurred on average at least once per dialogue (the threshold). In the pre-test of the control group, 25 different T-S-T sequences, together 672 sequences, occurred on average at least one time per dialogue (the threshold). In the post-test of the training group, 36 different T-S-T sequences, together 1075 sequences, occurred on average at least one time per dialogue (the threshold). In the post-test of the control group, 29 different T-S-T sequences, together 750 sequences, occurred on average at least one time per dialogue (the threshold). See Table 4 for the frequencies of combinations of teacher questions and teacher responses in the 46 different T-S-T sequences (the S consisting of a Student Accept or Neutral response) that reached the threshold.

In the post-test compared to the pre-test in the training group and the control group, respectively, 20 and 6 different new T-S-T sequences reached the threshold. In the training group, 18 of these 20 new sequences contained at least a Deep reasoning question, a Reasoning prompt, or a Prompt response from teachers. In the control group, the sequences

Table 3
Frequencies, Expected Frequencies, Kappa- and Z- Scores of T-S Sequences^a

| | Student response | | | | | | | |
|-------------------|--|---|---|--|--|---|--|---|
| | Accept | | | | Neutral | | | |
| | Pre-test | | Post-test | | Pre-test | | Post-test | |
| | Training | Control | Training | Control | Training | Control | Training | Control |
| Teacher question | | | | | | | | |
| Recall | Fr = 117 E = 77.57 Z = 6.40** K = .36 | Fr = 75 E = 57.85 Z = 3.62** K = .37 | Fr = 174 E = 108.41 Z = 8.99** K = .47 | Fr = 110 E = 87.86 Z = 3.81** K = .27 | Fr = 71 E = 110.42 Z = -6.40** K = -.51 | Fr = 29 E = 46.15 Z = -3.62** K = -.30 | Fr = 75 E = 140.59 Z = -8.99** K = -.60 | Fr = 59 E = 81.14 Z = -3.81** K = -.25 |
| Hybrid | Fr = 53 E = 44.57 Z = 1.73* K = .13 | Fr = 17 E = 19.47 Z = -0.86 K = -.16 | Fr = 70 E = 76.19 Z = -0.99 K = -.06 | Fr = 18 E = 15.08 Z = 1.11 K = .21 | Fr = 55 E = 63.43 Z = -1.73* K = -.19 | Fr = 18 E = 15.53 Z = 0.86 K = .13 | Fr = 105 E = 98.81 Z = 0.99 K = .16 | Fr = 11 E = 13.92 Z = -1.11 K = -.19 |
| Hybrid prompt | Fr = 18 E = 22.28 Z = -1.21 K = -.14 | (Fr = 5) | Fr = 38 E = 44.85 Z = -1.40 K = -.12 | (Fr = 5) | Fr = 36 E = 31.72 Z = 1.21 K = .19 | (Fr = 5) | Fr = 65 E = 58.15 Z = 1.40 K = .15 | (Fr = 3) |
| Surface reasoning | (Fr = 10) | (Fr = 2) | (Fr = 4) | (Fr = 5) | Fr = 101 E = 65.20 Z = 7.27** K = .78 | Fr = 28 E = 13.31 Z = 5.50** K = .88 | Fr = 123 E = 71.70 Z = 9.50** K = .93 | Fr = 42 E = 22.57 Z = 5.84** K = .80 |
| Deep reasoning | (Fr = 5) | (Fr = 2) | (Fr = 9) | (Fr = 0) | Fr = 63 E = 39.94 Z = 5.86** K = .82 | (Fr = 6) | Fr = 156 E = 93.16 Z = 10.32** K = .87 | (Fr = 4) |
| Reasoning prompt | (Fr = 2) | (Fr = 1) | (Fr = 7) | (Fr = 0) | (Fr = 8) | (Fr = 3) | Fr = 58 E = 36.70 Z = 5.42** K = .75 | (Fr = 0) |

(Continued)

Table 3
Continued

| | Student response | | | | | | | |
|-------------------------------|---|--|---|--|---|---|--|---|
| | Accept | | | | Neutral | | | |
| | Pre-test | | Post-test | | Pre-test | | Post-test | |
| | Training | Control | Training | Control | Training | Control | Training | Control |
| Teacher response ^b | | | | | | | | |
| Accept | Fr = 58 E = 68.09 Z = -1.73* K = -.10 | Fr = 80 E = 134.61 Z = -8.42** K = -.51 | Fr = 112 E = 174.16 Z = -7.05** K = -.28 | Fr = 75 E = 110.05 Z = -6.81** K = -.40 | Fr = 107 E = 96.91 Z = 1.73* K = .15 | Fr = 162 E = 107.39 Z = 8.42** K = .41 | Fr = 288 E = 225.84 Z = 7.05** K = .36 | Fr = 154 E = 109.95 Z = 6.81** K = .37 |
| Neutral | Fr = 168 E = 138.65 Z = 3.89** K = .15 | Fr = 250 E = 195.79 Z = 7.70** K = .35 | Fr = 302 E = 196.37 Z = 11.48** K = .41 | Fr = 224 E = 186.11 Z = 5.26** K = .22 | Fr = 168 E = 197.35 Z = -3.89** K = -.21 | Fr = 102 E = 156.21 Z = -7.70** K = -.28 | Fr = 149 E = 254.63 Z = -11.84** K = -.54 | Fr = 134 E = 171.89 Z = -5.26** K = -.20 |
| Prompt | Fr = 32 E = 33.84 Z = -0.43 K = -.04 | Fr = 19 E = 13.35 Z = 2.36* K = .53 | Fr = 111 E = 67.49 Z = 7.35** K = .50 | Fr = 8 E = 6.24 Z = 1.03 K = .31 | Fr = 50 E = 48.16 Z = 0.43 K = .05 | Fr = (5) E = 87.51 Z = -7.35** K = -.64 | Fr = 44 E = 87.51 Z = -7.35** K = -.64 | (Fr = 4) |
| Total ^c | Fr = 463 | Fr = 455 | Fr = 829 | Fr = 445 | Fr = 659 | Fr = 363 | Fr = 1075 | Fr = 411 |

Notes: ^a Frequencies below the threshold (training group ≥ 16 , control group ≥ 7) in parentheses; ^b teacher utterance other than a question; ^c these totals are without the numbers given above between parentheses.

* $p < .05$, ** $p < .01$.

Fr = frequency; E = expected frequency; Z = Z-score; K = kappa-score.

Table 4
Frequencies of Teacher Questions and Responses in T-S-T Sequences

| | | | Student response | | | | | | | |
|-------------------|----------------|-------------------------------|------------------|---------|-----------|---------|----------|---------|-----------|---------|
| | | | Accept | | | | Neutral | | | |
| | | | Pre-test | | Post-test | | Pre-test | | Post-test | |
| | | | Training | Control | Training | Control | Training | Control | Training | Control |
| Teacher question | S ^b | Teacher question | | | | | | | | |
| Recall | S ^b | Recall | 23 | 22 | 30 | 27 | | | | 17 |
| Recall | S ^b | Hybrid | | | 16 | | | | | |
| Recall | S ^b | Surface reasoning | 17 | | 18 | 8 | | | | |
| Surface reasoning | S ^b | Recall | | | | | 22 | | 18 | 11 |
| Surface reasoning | S ^b | Hybrid | | | | | 18 | | 19 | |
| Surface reasoning | S ^b | Surface reasoning | | | | | | | 17 | 8 |
| Deep reasoning | S ^b | Recall | | | | | | | 22 | |
| Deep reasoning | S ^b | Deep reasoning | | | | | | | 20 | |
| Teacher question | S ^b | Teacher response ^a | | | | | | | | |
| Recall | S ^b | Accept | | 17 | 27 | 26 | | 11 | 17 | 18 |
| Recall | S ^b | Neutral | 35 | 24 | 41 | 40 | 23 | 8 | | 15 |
| Hybrid | S ^b | Accept | | | | | | | 26 | |
| Hybrid | S ^b | Neutral | | 7 | | | 19 | | 20 | |
| Hybrid prompt | S ^b | Accept | | | | | | | 17 | |
| Surface reasoning | S ^b | Accept | | | | | 16 | 11 | 24 | 11 |
| Surface reasoning | S ^b | Neutral | | | | | | 7 | | 9 |
| Deep reasoning | S ^b | Accept | | | | | | | 35 | |
| Deep reasoning | S ^b | Neutral | | | | | | | 29 | |

(Continued)

Table 4
Continued

| | | | Student response | | | | | | | |
|-------------------------------|----------------|-------------------------------|------------------|---------|-----------|---------|----------|---------|-----------|---------|
| | | | Accept | | | | Neutral | | | |
| | | | Pre-test | | Post-test | | Pre-test | | Post-test | |
| | | | Training | Control | Training | Control | Training | Control | Training | Control |
| Teacher response ^a | S ^b | Teacher question | | | | | | | | |
| Accept | S ^b | Recall | | | 15 | | 22 | 33 | 24 | 26 |
| Accept | S ^b | Hybrid | | | | | | 7 | 25 | |
| Accept | S ^b | Hybrid prompt | | | | | | | 19 | |
| Accept | S ^b | Surface reasoning | | | | | | 8 | | |
| Accept | S ^b | Deep reasoning | | | | | | | 25 | |
| Neutral | S ^b | Recall | 29 | 29 | 29 | 37 | 26 | 9 | | 29 |
| Neutral | S ^b | Hybrid | 18 | 7 | 28 | 10 | | | 16 | |
| Neutral | S ^b | Hybrid prompt | | | 23 | | | | | |
| Neutral | S ^b | Surface reasoning | | | | 10 | | | | |
| Neutral | S ^b | Deep reasoning | | | 20 | | | | | |
| Teacher response ^a | S ^b | Teacher response ^a | | | | | | | | |
| Accept | S ^b | Accept | | 20 | | 11 | 23 | 76 | 109 | 68 |
| Accept | S ^b | Neutral | 21 | 48 | 34 | 42 | 24 | 29 | 51 | 49 |
| Accept | S ^b | Prompt | | | | | | | 19 | |
| Neutral | S ^b | Accept | 16 | 41 | 38 | 32 | 28 | 28 | 33 | 41 |
| Neutral | S ^b | Neutral | 60 | 159 | 109 | 128 | 58 | 46 | 41 | 54 |
| Neutral | S ^b | Prompt | | | 28 | | | | | |
| Prompt | S ^b | Neutral | | | 28 | | | | | |
| Prompt | S ^b | Prompt | | | 21 | | | | | |
| Total | | | 219 | 374 | 490 | 386 | 279 | 273 | 626 | 356 |

Notes: ^a Teacher utterance other than a question; S^b student utterance is an Accept or Neutral response, see column headings.

contained at least a Recall question, a Surface reasoning question, an Accept response, or Neutral response from teachers. In the post-test compared to the pre-test in the control group, three sequences did not reach the threshold anymore, two with a Hybrid question and one with a Surface reasoning question.

Effects of the Training on the Occurrence of Utterances and Sequences

Ancova analyses were conducted on the data of question and response categories percentages of single teacher utterances and T-S and T-S-T sequences, which reached the threshold of occurrence on average at least once per dialogue in both groups at least on one of the two moments one of the two moments of measurement (pre-test and post-test). See Table 2 for the outcomes of 8 of 12 Ancovas concerning percentages of single teacher utterances that reached the threshold. Six of these Ancovas revealed a significant difference between the two groups on the post-test, controlled for their pre-test frequencies: (1) teachers in the training group used more Hybrid, Hybrid prompt, Deep reasoning questions, and Prompt responses than teachers in the control group, (2) teachers in the training group used fewer Recall questions and Neutral Responses than the control group.

Concerning the T-S sequences, the threshold was not reached by both groups for Surface reasoning-Student Accept. Further, the threshold was not reached by the control group for three sequences starting with a question (Hybrid, Deep reasoning, and Reasoning prompt) all in combination with Accept or Neutral response, and for one sequence of two response categories (Prompt-Neutral). See Table 5 for the outcomes of the 11 Ancovas concerning percentages of the T-S sequences, which reached the threshold in both groups.

Table 5
Ancovas with Bootstrap of T-S Sequences in the Dialogues, Post-test Mean % and SD, Corrected for Pre-test Differences

| Teacher-Student sequence | | Training M% (SD) | Control M% (SD) | F | Bootstrap | |
|-------------------------------|------------------|---------------------|--------------------|-------|-----------|---------------------|
| | | | | | p | Part η ² |
| Teacher question | Student response | | | | | |
| Recall | Accept | 8.99 (1.19) | 12.48 (1.48) | 2.50 | .10 | .11 |
| Recall | Deny | 0.91 (.28) | 2.72 (.27) | 13.14 | .01** | .40 |
| Recall | Neutral | 3.51 (.56) | 6.90 (1.31) | 8.49 | .02** | .30 |
| Hybrid | Accept | 3.64 (.63) | 1.88 (.58) | 1.62 | .17 | .08 |
| Hybrid | Neutral | 5.22 (1.00) | 1.19 (.71) | 6.23 | .02** | .24 |
| Surface reasoning | Neutral | 7.45 (2.09) | 5.30 (2.06) | 0.33 | .55 | .02 |
| Teacher response ^a | Student response | | | | | |
| Accept | Accept | 5.40 (.84) | 7.47 (1.00) | 0.54 | .54 | .03 |
| Accept | Neutral | 12.45 (1.80) | 15.74 (2.20) | 1.21 | .20 | .06 |
| Neutral | Accept | 14.49 (1.84) | 22.36 (1.44) | 3.21 | .07 | .14 |
| Neutral | Neutral | 6.92 (.93) | 13.60 (2.27) | 9.83 | .02** | .33 |
| Prompt | Accept | 5.48 (.89) | .76 (.58) | 10.35 | .01* | .34 |

Notes: * $p < .01$, ** $p < .05$.
^a Teacher utterance other than a question.

Five Ancovas revealed a significant difference in percentages of T-S sequences between the two groups on the post-test, controlled for the pre-test percentages: (1) The training group used more Hybrid question-Student Neutral response sequences and Prompt response-Student Accept response sequences than the control group, (2) The training group used fewer Teacher Recall questions followed by, respectively, Student Deny and Neutral responses and Teacher Neutral response-Student Neutral response sequences than the control group.

With regard to the T-S-T sequences, the threshold was not reached by both groups for sequences containing a Reasoning prompt question and by the control group for sequences containing a Hybrid prompt a Deep reasoning question or a Prompt response. See Table 6

Table 6

Ancovas with Bootstrap of T-S-T Sequences in the Dialogues, Post-test Mean % and SD, Corrected for Pre-test Differences

| Teacher -Student-Teacher sequence | | | Training M% (SD) | Control M% (SD) | F | Bootstrap | |
|-----------------------------------|----------------------|-------------------------------|---------------------|--------------------|-------|-----------|---------------|
| | | | | | | p | Part η^2 |
| Teacher question | Student ^b | Teacher question | | | | | |
| Recall | Accept | Recall | 1.40 (.42) | 3.19 (.94) | 3.99 | .12 | .17 |
| Recall | Deny | Recall | .24 (.11) | 1.05 (.14) | 15.87 | .01* | .44 |
| Recall | Accept | Surface reasoning | 1.18 (.33) | 1.04 (.47) | 0.57 | .46 | .03 |
| Surface reasoning | Neutral | Recall | 1.14 (.29) | 1.38 (.68) | 0.10 | .79 | .01 |
| Surface reasoning | Neutral | Surface reasoning | .96 (.74) | 1.03 (.68) | 0.34 | .44 | .02 |
| Teacher question | Student ^b | Teacher response ^a | | | | | |
| Recall | Accept | Accept | 1.48 (.32) | 2.76 (.49) | 4.07 | .08 | .17 |
| Recall | Neutral | Accept | .81 (.23) | 2.05 (.44) | 7.66 | .05** | .28 |
| Recall | Accept | Neutral | 2.02 (.45) | 4.46 (.50) | 7.10 | .01** | .26 |
| Recall | Neutral | Neutral | .51 (.20) | 1.79 (.60) | 7.24 | .01** | .27 |
| Surface reasoning | Neutral | Accept | 1.35 (.33) | 1.45 (.70) | 0.01 | .90 | .00 |
| Teacher response ^a | Student ^b | Teacher question | | | | | |
| Accept | Neutral | Recall | .86 (.28) | 3.00 (.65) | 11.81 | .02** | .37 |
| Accept | Neutral | Hybrid | 1.13 (.27) | .46 (.17) | 2.10 | .09 | .10 |
| Neutral | Accept | Recall | 1.48 (.33) | 4.20 (.63) | 17.08 | .00* | .46 |
| Neutral | Neutral | Recall | .70 (.18) | 3.29 (.81) | 20.12 | .01* | .50 |
| Neutral | Accept | Hybrid | 1.50 (.41) | 1.11 (.55) | 0.24 | .61 | .01 |
| Teacher response ^a | Student ^b | Teacher response ^a | | | | | |
| Accept | Neutral | Accept | 4.61 (1.18) | 6.48 (1.88) | 0.16 | .69 | .01 |
| Accept | Accept | Neutral | 1.82 (.53) | 4.24 (.57) | 2.34 | .25 | .11 |
| Accept | Neutral | Neutral | 2.28 (.56) | 5.18 (.90) | 3.81 | .10 | .16 |
| Neutral | Accept | Accept | 1.59 (.54) | 3.43 (.40) | 0.13 | .75 | .01 |
| Neutral | Neutral | Accept | 1.25 (.35) | 4.00 (.62) | 15.97 | .00* | .44 |
| Neutral | Accept | Neutral | 4.86 (1.53) | 11.62 (2.04) | 1.77 | .15 | .08 |
| Neutral | Neutral | Neutral | 1.95 (.39) | 5.21 (1.47) | 8.22 | .01** | .29 |

Notes: * $p < .01$, ** $p < .05$.

^a Teacher utterance other than a question; ^b student utterance.

for the outcomes of the 22 Ancovas concerning percentages of T-S-T sequences, which reached the threshold in both groups.

Nine of these Ancovas revealed a significant difference between the two groups on the post-test, controlled for the pre-test T-S-T sequences, which were all T-S-T sequences less used by the training group than the control group: (1) Seven Ancovas concerned T-S-T sequences with at least a Teacher Recall question, (2) Two Ancovas concerned T-S-T sequences in which teachers used, first, a Neutral response and, second, an Accept response, or used two Neutral responses and students used Neutral responses.

Discussion and Conclusion

Feedback dialogues in vocational education were explored and the effects of a training programme were determined. It was assumed that teachers' prompts and reasoning questions could be very effective to evoke students' reflection to self-regulate their learning and development.

Exploration of Dialogues and Determination of Training Effects

In the control group, the average duration of the feedback dialogues and the average number of turns remained about the same, while both teachers and students used longer elaborations in the post-test than in the pre-test. In the training group, the average duration of the feedback dialogues and the average number of turns both increased, while the average number of words teachers used per turn decreased. The additional time for interaction and the decrease of the average number of words by the teachers were two reasons why students in the training group were given more time to reflect. Therefore, we conclude that verbal dominance in the training group shifted somewhat from teacher to student (Blatt et al., 2008; Chin, 2006). These results might be due to the training programme, in which teachers were stimulated to pose open questions in feedback dialogues and to listen more and talk less (Roter et al., 2004). They also learned that their own long elaborations are not useful for supporting students' reflections but, instead, active listening and caring for wait time are important (Good & Brophy, 2000). Moreover, in the post-test, the teachers of the training group used in sequences more question categories and prompts that are assumed to stimulate learning strategies for self-regulation, including reflection (e.g., Davis, 2003; Ifenthaler, 2012).

Utterances in Dialogues

Per question and response category, a frequency of occurrence of, on average, at least once per dialogue was used as a threshold for including a category in further analysis. Based on this threshold, teachers were shown to use six of seven question categories, while students used only Recall questions. Recall questions and Accept and Neutral responses were used far the most by the teachers. In the training group, the percentage of Recall questions in the dialogues was less than in the pre-test, while the opposite was true for the percentages of Hybrid prompt, Deep reasoning, and Reasoning prompt questions, and Prompt responses. The control group hardly used Deep reasoning questions or Prompt questions and responses in either the pre-test or the post-test. Therefore, we conclude that the changes in teacher questioning and responding were due to and in accordance with the aims of the training programme.

Sequences in Dialogues

Of all possible T-S sequences in both groups, the combination Recall question-Accept response was used by far the most, with, in the second place, the combination Surface reasoning-Neutral response. In the control group, the number of different T-S-T sequences that reached the threshold in the pre-test remained approximately the same in the post-test, while teachers of the training group extended their repertoire from 18 to 38 different T-S-T sequences. Of the new sequences, 90% contained at least a Deep reasoning question, a Reasoning prompt, or a Prompt response.

Our conclusion is that the teachers in the training group became accustomed to a broader repertoire because their dialogues contained many more utterances and sequences, which included more questions and prompts that potentially stimulate learning strategies for self-regulation and stimulate students to elaborate and reflect (Berthold et al., 2007; Davis, 2003; Ifenthaler, 2012). Therefore, we assume that the training programme did contribute to the teachers' competence development in supporting students' reflection skills.

Study Limitations

This study is subject to some limitations. First, a limitation of the effect study was that the random selection of schools or teachers and matching of groups for the quasi-experimental design was not feasible due to turbulent circumstances during the study in Dutch senior secondary nursing education. Moreover, in the study, one selection criterion was teacher's interest in professional development. Second, a limitation concerns the coding of Hybrid and Hybrid prompt questions. Questions that are hybrid can be coded both as Recall questions and as one of the Reasoning question categories. We only coded the Hybrid question itself but not the question category to which the Hybrid question referred. Therefore, we do not know exactly how many questions were Recall questions and how many questions concerning Reasoning question categories were posed by the teachers in both groups in the pre-test and the post-test. Third, the percentages agreement received in coding questions and responses of teachers and students during the feedback dialogues were not perfect. This should be taken into account when interpreting the results of the study. Further, a small number of teachers and students participated in the study during a specific period. However, the results of the study provide insight into which question-and-response categories teachers might use to stimulate students to reflect.

Practical Implications

The results of this study might be used to develop guidelines for teaching reflection skills in nursing education and in pre-service nursing teacher education. The findings seem broadly applicable and useful in other contexts, for example other health disciplines, pre-service, and general teacher education, and secondary and tertiary vocational education for social work professions. In these contexts the findings can be disseminated in a training programme for teachers as well as for tutors. Next, they can use question-and-response categories in feedback dialogues to promote their students' reflection.

In a revised training programme, teachers might be instructed explicitly about Reasoning question categories and their possible effects on reflection. Moreover, teachers can be trained in teaching students to reflect by giving students explicit instruction about how they can pose questions themselves and how they can deal with questions in written reflection tasks. These

instruction and tasks should be consistent across courses in school and internships, which is possible when theory is linked to practice.

Suggestions for Further Research

It would be interesting and might be relevant to constructively replicate the study. In future research, different ways of coding could be compared, for example by coding hybrid questions and both categories to which these hybrid questions refers. Also, teachers' questions in feedback dialogues with a small group of students could be explored. Moreover, the effects of coaching teachers in action by means of ear-bud headphones could be studied. Further research might focus on whether the changes in teachers' feedback have the desired effect on their students. Research could, for example, focus on how teachers' questions influence students' thinking processes and reflection skills, and how students appreciate teachers' questions. Finally, research could be done to establish what students learn from teachers' feedback by giving them a reflection task before and after the feedback.

Funding

The study was funded by the Dutch Organisation for Scientific Research NWO-PROO 411-06-311.

References

- Banerjee, M., Capozzoli, M., McSweeney, L., & Sinha, D. (1999). Beyond kappa: A review of inter-rater agreement measures. *Canadian Journal of Statistics*, 27, 3–23. doi:10.2307/3315487
- Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, 17, 564–577. doi:10.1016/j.learninstruc.2007.09.007
- Blatt, B., Confessore, S., Kallenberg, G., & Greenberg, L. (2008). Verbal interaction analysis: Viewing feedback through a different lens. *Teaching and Learning in Medicine*, 20(4), 329–333. doi:10.1080/10401330802384789
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (Eds.). (1956). *Taxonomy of educational objectives – The classification of educational goals – handbook 1: Cognitive domain*. New York: David Mc Kay.
- Branch, W. T., & Paranjape, A. (2002). Feedback and reflection: Teaching methods for clinical settings. *Academic Medicine*, 77(12), 1185–1188. Retrieved from http://ovidsp.tx.ovid.com/sp-3.12.0b/ovidweb.cgi?&S=JELCFPFOHEDDGDDINCMKDELBFDOPAA00&Link+Set=S.sh.65%7c1%7csl_10
- Butler, D. L. (1998). The strategic content learning approach to promoting self-regulated learning: A report of three studies. *Journal of Educational Psychology*, 90, 682–697. doi:10.1037/0022-0663.90.4.682
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education: Principles, Policy and Practice*, 14, 171–184. doi:10.1080/03075070600572132
- Chi, M. T. H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology*, 10, 33–49. doi:10.1002/(SICI)1099-0720(199611)10:7<33::AID-ACP436>3.0.CO;2-E
- Chi, M. T. H. (1997). Quantifying qualitative analysis of verbal data: A practical guide. *Journal of the Learning Sciences*, 6, 271–315. doi:10.1207/s15327809jls0603_1
- Chin, C. (2006). Classroom interaction in Science: Teacher questioning and feedback to students' responses. *International Journal of Science Education*, 11, 1315–1346. doi:10.1080/09500690600621100

- Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *Journal of the Learning Sciences*, 12(1), 91–142. doi:10.1207/S15327809JLS1201_4
- Dekker-Groen, A. M., Van der Schaaf, M. F., & Stokking, K. M. (2013). A teacher competence development programme for supporting students' reflection skills. *Teachers and Teaching: Theory and Practice*, 19(2), 150–170. doi:10.1080/13540602.2013.741837
- Erkens, G. (2005). *Multi Episode Protocol Analysis* (MEPA). Version 4.10. Retrieved from <http://edugate.fss.uu.nl/mepa>
- Field, A. (2009). *Discovering statistics using SPSS* (3th ed.). London: Sage.
- Flammer, A. (1981). Towards a theory of question asking. *Psychological Research*, 43(4), 407–420. doi:10.1007/BF00309225
- Gé, X., & Land, S. M. (2004). A conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development*, 52(2), 5–22. doi:10.1007/BF02504836
- Good, T. L., & Brophy, J. E. (2000). *Looking in classrooms*. New York: Longman.
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31, 104–137. doi:10.3102/00028312031001104
- Huck, S. W., & McLean, R. A. (1975). Using a repeated measure ANOVA to analyze the data from a pre-test-post-test design: A potentially confusing task. *Psychological Bulletin*, 82(4), 511–518. doi:10.1037/h0076767
- Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Educational Technology & Society*, 15(1), 38–52.
- Jeong, A. (2005). A guide to analyzing message-response sequences and group interaction patterns in computer-mediated communication. *Distance Education*, 26(3), 367–383. doi:10.1080/01587910500291470
- Kearsley, G. P. (1976). Questions and question asking in verbal discourse: A cross-disciplinary review. *Journal of Psycholinguistic Research*, 5, 355–375. doi:10.1007/BF01079934
- Koedinger, K. R., & Alevan, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19, 239–264. doi:10.1007/s10648-007-9049-0
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212–218. doi:10.1207/s15430421tip4104_2
- Lee, H-J. (2005). Understanding and assessing preservice teachers' reflective thinking. *Teaching and Teacher Education*, 21, 699–715. doi:10.1016/j.tate.2005.05.007
- Lizzio, A., & Wilson, K. L. (2004). Action learning in higher education: An investigation of its potential to develop capability. *Studies in Higher Education*, 29, 469–488. doi:10.1080/0307507042000236371
- Mercer, N. (2004). Sociocultural discourse analysis: Analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics*, 1(2), 137–168.
- Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction*, 19, 259–271. doi:10.1016/j.learninstruc.2008.05.002
- Roter, D. L., Larson, S., Shinitzky, H., Chernoff, R., Serwint, J. R., Adamo, G., & Wissow, L. (2004). Use of an innovative video feedback technique to enhance communication skills training. *Medical Education*, 38, 145–157. doi:10.1111/j.1365-2923.2004.01754.x
- Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education*, 35(5), 535–550. doi:10.1080/02602930903541015
- Shuell, T. J. (1996). Teaching and learning in a classroom context. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 726–763). Mahwah, NJ: Lawrence Erlbaum Associates.
- Strijbos, J. W., & Stahl, G. (2007). Methodological issues in developing a multi-dimensional coding procedure for small-group chat communication. *Learning and Instruction*, 17, 394–404. doi:10.1016/j.learninstruc.2007.03.005
- Thillmann, H., Küsting, J., Wirth, J., & Leutner, D. (2009). Is it merely a question of 'What' to prompt or also 'When' to prompt? *Zeitschrift für Pädagogische Psychologie*, 23(2), 105–115. doi:10.1024/1010-0652.23.2.105

Thomas Farrar, M. T. (1986). Teacher questions: The complexity of the cognitively simple. *Instructional Science*, 15(1), 89–107.

Thompson, I. (2009). Scaffolding in the writing centre: A microanalysis of an experienced tutor’s verbal and nonverbal tutoring strategies. *Written Communication*, 26(4), 417–453. doi:10.1177/0741088309342364

Van den Boom, G., Paas, F., & Van Merriënboer, J. J. G. (2007). Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction*, 17, 532–548. doi:10.1016/j.learninstruc.2007.09.003

Van der Schaaf, M., Baartman, L., Prins, F., Oosterbaan, A., & Schaap, H. (2011). Feedback dialogues that stimulate students’ reflective thinking. *Scandinavian Journal of Educational Research*, 57(3), 227–245. doi:10.1080/00313831.2011.628693

Van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers’ ‘learning to notice’ in the context of a video club. *Teaching and Teacher Education*, 24, 244–276. doi:10.1016/j.tate.2006.11.005

Van Zee, E., & Minstrell, J. (1997). Using questioning to guide student thinking. *Journal of the Learning Sciences*, 6(2), 227–269. doi:10.1207/s15327809jls0602_3

Wampold, B. E., & Margolin, G. (1982). Non-parametric strategies to test independence of behavioural states in sequential data. *Psychological Bulletin*, 92, 755–765.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70. doi:10.1207/s15430421tip4102_2

Appendix A

Table A1
Coding Scheme Question and Response Categories

| Question category | Short description | Example | Reference |
|---------------------------|---|---|--|
| 1. Recall (TS) | Question to which a reply is expected that requests few, e.g., a verification, property or specification, complementation of a concept, quantitative data | T: Where is your internship? S: Is it correct now? | Chin (2006); Graesser & Person (1994) |
| 2. Hybrid (T) | Question that the student may perceive as a short-answer question, but which is intended to obtain a long answer | T: Do you know what the cause is from that? | Graesser & Person (1994) |
| 3. Hybrid prompt (T) | Question that may encourage productive learning strategies, which the student in principle is capable of, but does not spontaneously show or shows on a (too) low level | T: Has your changed attitude ensured that the contact with your colleagues is better now? | Berthold et al. (2007); Ifenthaler (2012); Nückles et al. (2009) |
| 4. Request (TS) | Question to prompt a particular action | T: Tell me. S: Explain just once more, please. | Graesser & Person (1994) |
| 5. Surface reasoning (TS) | Question to which an answer longer than one sentence is expected but not an argument, e.g., demand for a definition, example, comparison, interpretation or judgment | T: What else did go well? S: What do you mean by explain? | Graesser & Person (1994) |

(Continued)

Table A1
Continued

| Question category | Short description | Example | Reference |
|-------------------------|--|---|--|
| 6. Deep reasoning (TS) | Question to which an answer longer than one sentence is expected and that exposes a clear way of reasoning, e.g., about what led to an event, cause and effect, purposefulness, expectations, a means which enables to do something or an instrumental-procedural question | T: Why do you think that would be of influence? S: How could I do it the next time? | Graesser & Person (1994) |
| 7. Reasoning prompt (T) | Question that may encourage productive learning strategies, which the student in principle is capable of, but does not spontaneously show or shows on a (too) low level | T: You mention 'panting in your neck' of people who want to hurry. How does that influence you? | Berthold et al. (2007); Graesser & Person (1994); Ifenthaler (2012); Nückles et al. (2009) |
| Response category | | | |
| 1. Neutral (TS) | A reaction that is not acceptance or negation of what the other said before | T: Yes, but I ask this myself. S: I am doing an internship in a hospital. | Blatt et al. (2008); Chin (2006) |
| 2. Deny (TS) | A negation of what the other said | T: No, that is not right. S: No, that is not what I mean. | Blatt et al. (2008); Chin (2006) |
| 3. Accept (TS) | Acceptation and/or confirmation of what the other said; a positive utterance | T: Yes, indeed. S: Yes, I can do it that way. | Blatt et al. (2008); Chin (2006) |
| 4. Prompt (T) | A tip or hint that can stimulate the student to use learning strategies that the student is capable of in principal but does not spontaneously show or shows on too low a level | T: A next time you could use the alarm button. | Berthold et al. (2007); Ifenthaler (2012); Nückles et al. (2009) |

Note: T = teacher, S = student.

Appendix B

Example of Coding and Sequences

The example is a part of a feedback dialogue between a teacher and student in the training group in the post-test. Codes of teacher questions and responses and of student responses are included. Teacher questions concern Recall and Hybrid prompt and teacher responses concern Accept, Neutral, and Prompt. Student responses concern Accept and Neutral. The digits show how different sequences were distinguished for analyses; digit 1 shows which utterances together form sequence 1, digit 2 shows which utterances together form sequence 2, and so on.

Table B1

Example of Coding and Sequences

| T or S | Utterances* | Category | T-S sequence | T-S-T sequence |
|--------|---|---------------|--------------|----------------|
| T | And could you think of something, uhm, an example or name, of which you say that you have sometimes experienced? That you talk about a critical situation? | Hybrid prompt | 1 | 1 |
| S | Now, uhm, critical, acute, uhm, yes acute, a gentleman who has fallen and then everywhere lies blood of which you think yes, you don't see immediately where it does come from. And then it is ohjee, what should I... | Neutral | 1 | 1 |
| T | Yes. | Neutral | 2 | 1-2 |
| S | And then nevertheless behave in a right way by taking the man inside, if he can still walk the distance, cleaning up everything and then quietly looking what it is. | Neutral | 2 | 2 |
| T | Mhmm, yes. | Accept | 3 | 2-3 |
| S | And then, based on this, phone a resident or possibly a specialist. | Neutral | 3 | 3 |
| T | That you take action? | Recall | 4 | 3-4 |
| S | Yes, then take action yes. Uhm, yes I then always, I always keep quiet in view of such things. I always first look carefully and talk with this man and then, uhm, reassure him and that. Yes such situations or a situation of, uhm, a man who got a bad result that he had a tumour. | Accept | 4 | 4 |
| T | Yes. | Accept | 5 | 4-5 |
| S | Yes, in that case it is first something of ohjee, because you are the first to hear it, because I was the first person who came there when they just had received the result. Then it is for a moment sort of, yes, how do I react on the spot. So first I will quietly listen for a while to the story. And, uhm, you know, what the people have to say themselves, and even then I do not react much, I simply let them talk and then I ask if they have something to ask or want something from me or if I can do something. | Accept | 5 | 5 |
| T | Yes, and it is to your satisfaction when you are going home in the evening? | Hybrid prompt | 6 | 5-6 |
| S | Yes, I actually feel quite good about this. | Accept | 6 | 6 |
| T | Yes. | Accept | 7 | 6-7 |
| S | Because I think I can fairly well imagine myself in other people. | Neutral | 7 | 7 |
| T | You do also write it here. | Neutral | 8 | 7-8 |
| S | Yes, I added it there. | Accept | 8 | 8 |
| T | That is one of your qualities. | Prompt | 9 | 8-9 |

Notes: T = teacher, S = student, *Utterances as part of a teacher-student feedback dialogue in the post-test of the training group.