Investigating Feedback on Practice Among Teachers: Coherence of Observed and Perceived Feedback

Document status and date:
Published: 24/09/2012

Document Version:
Peer reviewed version

Document license:
CC BY-SA

Open Universiteit
www.ou.nl

Investigating Feedback on Practice Among Teachers: Coherence of Observed and Perceived Feedback

Marieke Thurlings , Marjan Vermeulen , Theo Bastiaens & Sjef Stijnen

Published online: 24 Sep 2012.

Submit your article to this journal

Article views: 591

View related articles

Investigating Feedback on Practice Among Teachers: Coherence of Observed and Perceived Feedback

Marieke Thurlings
LOOK (Scientific Centre for Teacher Research), Open Universiteit, Heerlen, The Netherlands

Marjan Vermeulen
KPC-group, 's-Hertogenbosch, The Netherlands;
Teacher University, Open Universiteit, Heerlen, The Netherlands

Theo Bastiaens
LOOK (Scientific Centre for Teacher Research), Open Universiteit, Heerlen, The Netherlands; FernUniversität, Hagen, Germany

Sjef Stijnen
LOOK (Scientific Centre for Teacher Research), Open Universiteit, Heerlen, The Netherlands

Despite that benefits of feedback in student learning are reported in much research, little has been reported regarding the use of feedback from teachers to other teachers—a key tool in professional development. In this study, we triangulated data from videotaped peer coaching sessions, questionnaires, and interviews regarding 12 primary school teachers in four peer groups in the Netherlands. We focused our research on two issues: the interplay of observed feedback dimensions and elements and perceptions of that feedback. Feedback dimensions were generally effective and the influence of the elements on the dimensions mostly aligned with the expectations. Teachers generally perceived feedback as effective. Moreover, effective observed feedback was perceived as effective. Findings indicate that peer coaches should stimulate coached teachers to become goal directed, specific, detailed, and neutral (neither positive nor negative) by using feedback elements so as to optimize feedback processes.

Keywords: feedback, teachers, peer coaching, TFOS

Benefits of feedback in supporting students' learning are confirmed in much research (Hattie, 2009; Hattie & Timperley, 2007; Mory, 2003; Shute,

Marieke Thurlings, LOOK (Scientific Centre for Teacher Research); Marjan Vermeulen, KPC-group and Teacher University; Theo Bastiaens, LOOK (Scientific Centre for Teacher Research), and FernUniversität; Sjef Stijnen, LOOK (Scientific Centre for Teacher Research).
 Correspondence concerning this article should be addressed to Marieke Thurlings, LOOK (Scientific Centre for Teacher Research), Valkenburgerweg 177, NL 6401 DL, Heerlen, The Netherlands. Email: marieke.thurlings@ou.nl

2008). However, as Scheeler, Ruhl, and McAfee (2004) pointed out, there has been little investigation into feedback among teachers, which can serve as an effective tool for their professional development. As professional development is crucial for the quality of education (OECD, 2002), we aimed to provide insights into (a) what effective feedback among teachers is like and (b) how teachers perceive feedback from their peer coaches.

Feedback Among Teachers

While research on feedback on practice between teachers is scarce (Scheeler et al., 2004), we transferred ideas and definitions from previous studies on feedback from teachers to students to inform our own work and to provide insights on whether feedback provided in hierarchical relationships can be adapted to feedback between peer coaches. In this study, feedback is described as "information that allows for comparison between an actual and a desired outcome" (Mory, 2003, p. 746). Characteristics of feedback can be synthesized along five dimensions (Thurlings, Vermeulen, Kreijns, Bastiaens, & Stijnen, 2012):

(1) Goal directedness vs. person directedness (Black & Wiliam, 1998a; Gibbs & Simpson, 2004);
(2) Specific vs. general (Black & Wiliam, 1998a; Mory, 2003; Scheeler et al., 2004);
(3) Detailed vs. nondetailed (Gibbs & Simpson, 2004; Scheeler et al., 2004; Weaver, 2006);
(4) Positive vs. negative (Scheeler et al., 2004; Schelfhout, Dochy, & Janssens, 2004; Tillema & Smith, 2000; Weaver, 2006);
(5) Immediate vs. delayed (Mory, 2003).

Drawing upon these dimensions, we can define effective and ineffective feedback. First, it is suggested that goal-directed feedback is more effective than person-directed (Black & Wiliam, 1998a; Gibbs & Simpson, 2004) or nongoal directed (Hattie & Timperley, 2007). Second, it is argued that specific feedback is more effective than general (Black & Wiliam, 1998a; Mory, 2003; Scheeler et al., 2004), though general advice on how to improve one's actions in the future is effective too (Black & Wiliam, 1998b; Weaver, 2006). Third, it is suggested that feedback that includes details is more effective than if it lacks details (Gibbs & Simpson, 2004; Scheeler et al., 2004; Weaver, 2006). Fourth, it is unclear whether positive feedback is more effective than negative. Some scholars argued that it should be positive (Scheeler et al., 2004; Tillema & Smith, 2000), whereas others argued that negative feedback can motivate learners in their learning process (Schelfhout et al., 2004). Still others argued that when positive and negative comments are balanced, feedback is more effective (Weaver, 2006). Moreover, Hattie

and Timperley (2007) suggested that feedback on the self-level—irrespective of whether it is positive or negative—is not effective, because it does not provide information on how learners can improve their actions. Fifth, immediate feedback is considered more effective than delayed (Mory, 2003). In addition, delayed feedback is less effective than if it is still relevant for the learner (Black & Wiliam, 1998a; Scheeler et al., 2004).

Based on this previous literature, we can assume that goal-directed, specific, and detailed feedback that is neutral (neither positive nor negative) is more effective than nongoal/person directed, general, vague, and either too positive or too negative feedback. The fifth dimension—timing—was not included in the study, because feedback was always communicated during the peer coaching sessions, and in addition, adequate timing is difficult to observe. When feedback is given in peer groups, such as in this research, the question arises how peer coaches (providers) can stimulate the coached teachers (receivers) to be goal directed, specific, detailed, and neutral, and can simultaneously help them to avoid being person or nongoal directed, general, nondetailed, and either too positive or too negative.

We investigated a number of elements that may lead to effective feedback in the peer group as well as elements that may provoke ineffective feedback. These elements are based upon literature on coaching as well as the peer coaching program (Gallacher, 1997; Jackson & McKergow, 2002; Jeninga, 2003; Smith & Ragan, 1993; see also Thurlings et al., 2012). Feedback elements were made operational as open-ended, closed, guiding, solution-focused, and evocative questions, continuous questioning, summarizing, acknowledging, judging, hinting, finishing sentences, and providing examples from one's own classroom or experience.

Based on the peer coaching program, we formulated expectations of the effectiveness of these elements. Effective elements strengthen the dimensions; that is, they push the feedback dimensions in the desired direction. Ineffective elements weaken the dimensions, pushing them away from the desired direction. We expected that (a) open-ended, closed, solution-focused, and guiding questions, continuous questioning, acknowledging, and summarizing would be effective and (b) hinting, judging, evocative questions, providing an example from one's own experience, and finishing sentences would be ineffective.

Perceived feedback was also included in the study. Perceived feedback is a factor that can support students' learning in formative assessment (Gibbs & Simpson, 2004). Gibbs and Simpson (2004) described that, in quantitative terms, sufficient feedback should be provided frequently and should hold enough details. In qualitative terms, it should be focused at performances that learners can control. Furthermore, it should be appropriate for the tasks learners are performing and should consider learners' understanding of what they should be doing. Moreover, they argued that feedback should be received, attended to, and acted upon such that student learning becomes

optimal. If feedback is perceived in alignment with these conditions, Gibbs and Simpson (2004) suggested that student learning is reinforced.

In summary, we divided feedback into four dimensions and 12 elements. We assume that feedback that is goal directed, specific, detailed, and neutral (neither positive nor negative) is more effective than when it is nongoal/person directed, vague, nondetailed, and either too positive or too negative. We studied the influence of the elements on the dimensions, expecting that (a) open-ended, closed, solution-focused, and guiding questions, continuous questioning, acknowledging, and summarizing are effective and (b) hinting, judging, evocative questions, providing an example from one's own experience, and finishing sentences are ineffective. In addition, perceived feedback was investigated. Furthermore, we attended to the coherence of observed and perceived feedback.

Methods

Research Context and Questions

In this study, we focused on feedback among teachers within a specific professional development activity, namely a peer coaching program, implemented in two Dutch primary schools. Elements of this program were reciprocal peer coaching, videotaped lessons, solution-focused thinking, and a cyclic workflow consisting of two sessions. In the first session, goals were set and actions were formulated. In the second session, the performed actions were evaluated (Jeninga, 2003). The aim of the program was to guide teachers in improving their own teaching behaviors, for instance, being more consistent in applying the classroom rules. Process supervisors participated in peer coaching groups. These persons acted as chairmen and modeled coaching behavior using solution-focused thinking.

In the cyclic workflow, three phases could be distinguished: the observation phase, in which teachers watch and discuss each other's videos; the analysis phase, in which they discuss and formulate goals and actions; and the reflection phase, in which the teachers evaluate their changed teaching behaviors (White, 2009).

We formulated the following research questions:

(1) To what extent do participants that receive effective feedback (as described in the theoretical framework) perceive this feedback as effective; and to what extent do participants that receive ineffective feedback perceive this feedback as ineffective?

    (a) To what extent is effective feedback provided; which feedback elements are effective and which are ineffective; and are there differences imposed by the three phases of the coaching sessions?

    (b) How did participants perceive feedback?

Participants

Twelve teachers from two Dutch primary schools participated in the program. At each of the schools, the peer coaching program had been previously implemented under the guidance of our colleagues. We approached the schools and teachers to seek their involvement in the study. Hence, the study was conducted in a naturalistic setting.

Of the 12 participants, two were men and ten were women. Their mean age was 30.7-years-old (SD = 6.7 years). The teachers were divided into four peer groups. Three groups (nine teachers) were from one primary school; the other group of three teachers was from another primary school. Each group was joined by a different process supervisor (two males and two females). All process supervisors and some of the teachers were experienced in applying the peer coaching program.

Observations

The peer coaching sessions were videotaped and transcribed. Subsequently, each teacher's turn being the coached teacher was divided into the phases of observation, analysis, and reflection (White, 2009). Next, the teachers' turns were analyzed using the Teacher Feedback Observation Scheme (TFOS), which was developed to observe feedback provided within the peer coaching sessions and to determine its quality (Thurlings et al., 2012). TFOS' interrater reliability was shown to be substantial (Thurlings et al., 2012). Cohen's κ—that expresses the interrater reliability (Landis & Koch, 1977)—averaged .765 with a minimum of .410 and a maximum of 1.000. TFOS was based upon the literature described above (section "Feedback Among Teachers"), and distinguishes the four dimensions (goal directedness, specificity, details, and positivity) and the 12 elements (e.g. open-ended and guiding questions, hinting, and summarizing).

The scoring of these dimensions and elements in each utterance of coached teachers and their peer coaches was conducted using Excel. The scoring of the dimensions was executed as follows:

- When an utterance was completely goal directed, we assigned a score of 4.
- When an utterance was completely nongoal/person directed, we assigned a score of −4.
- When an utterance was balanced between goal directedness and nongoal/person directedness, we assigned a score of 0.
- Between these extremes (+4 and −4) and the zero-point, a +2 and a −2 can be assigned.

This method resembled a Likert-type scale and was also applied for the other dimensions.

The feedback elements (e.g. closed questions, acknowledging, and judging) were assigned either a 1 or a -1, when the element, respectively, was expected to be effective (e.g. open-ended questions) and ineffective (e.g. hinting). By assigning either a 1 or a -1, the elements were easily distinguished from the dimensions.

Questionnaires

After each peer coaching session, teachers completed a questionnaire by which we could investigate how they perceived feedback within the session. Three subscales of the Assessment Experience Questionnaire (AEQ) were used to evaluate these perceptions (Gibbs & Simpson, 2003). The AEQ is based upon factors that support student learning. The three subscales were *quantity of feedback* ($\alpha = .693$), *quality of feedback* ($\alpha = .662$), and *what a student does with the feedback* ($\alpha = .614$). The Chronbach's α found in our sample were slightly lower than the ones Gibbs and Simpson found ($\alpha_{quantity} = .87$; $\alpha_{quality} = .77$; $\alpha_{do\ with\ feedback} = .74$), though their sample of almost 800 students was much larger than ours of 21 questionnaires (see section 'Data Collection').

The items were adapted to fit the peer coaching program. One item of the scale *quality of feedback* was eventually deleted because this improved Chronbach's α from .473 to .662 and because the item "The feedback mainly tells me how well I am doing in relation to others" was less relevant to the peer coaching program. Therefore, the respective numbers of the items of the scales were eight, five, and eight. The items were answered on a five-point scale.

Interviews

After the observations and questionnaires were collected, we approached teachers by email to elicit their participation in a semistructured interview. The interview questions were discussed in the research group and were pilot tested. Based on the discussion and the pilot test, we slightly adapted the interview by altering the sequence of questions.

Seven teachers agreed to be interviewed by telephone. Each interview took about 20–30 min. By executing these interviews, we aimed to gather qualitative data on how teachers had experienced feedback. A report was made for each interview and sent to the individual teachers so that they could ensure their opinions were accurately reflected.

Data Collection

Three of the peer groups had two peer coaching sessions in which teachers went through the cyclic workflow once, thus most teachers had two coaching turns. Two teachers were only present during one session and as a consequence did not fill in the questionnaire. Another teacher was present during both sessions, however, she failed to complete the questionnaire after

one of the sessions. The fourth peer group had three sessions, but each teacher was present only twice, which enabled each teacher of this group to experience the cyclic workflow once.

Each peer coaching session was videotaped. The videotapes were then transcribed. After each peer coaching session, the questionnaires were completed. Finally, the interviews took place. It was possible to connect the observations, questionnaires, and interviews to each individual participant.

Because most teachers had more than one turn being the coached teacher (they were present during both sessions), two observational videotapes were recorded for most teachers, and they filled in two questionnaires. There were two teachers who were not present during a session, and they did not complete the questionnaire; another teacher did not complete her questionnaire after one session. Thus, we collected 21 sets of data consisting of observational videotapes and accompanying questionnaires.

Data-Analysis

We performed nonparametric tests to address research questions 1(a) and (b). Nonparametric tests are designed for small samples, such as ours, that do not meet the assumptions for parametric tests (Field, 2005; Moore & McCabe, 2003).

We approached research question 1(a) in two main steps. After coding the observations using TFOS, we performed Kruskall–Wallis tests—the non parametric counterpart of analysis of variance (ANOVA) (Baarda, de Goede, & van Dijkum, 2007; Moore & McCabe, 2003)—in order to test whether the four feedback dimensions differed between the phases of observation, analysis, and reflection. We also performed additional Mann–Whitney tests—the nonparametric counterpart of the *t*-test—because the mean ranks pointed at possible significant differences between the phases.

Second, for each phase of all teachers' turns, we made timeline graphics of the Excel files. By putting the scoring of the feedback dimensions as well as the elements on a timeline, we were able to determine if and how the elements affected the dimensions.

We determined the effectiveness of each feedback element by comparing the position of the dimensions in the utterances on the timeline before and after the element occurred. If the dimensions were higher after the element than before, the element was counted as effective. If the dimensions were equal to or lower than before the element occurred, the element was counted as ineffective. The effectiveness could not be determined for all elements (e.g. those elements in the first utterance of a coaching session), because there was no previous utterance with which we could compare the second utterance.

To approach research question 1(b), we analyzed the questionnaires and interviews. Concerning the questionnaires, we performed nonparametric Wilcoxon signed rank tests—the counterpart of the paired *t*-test (Baarda

Table 1
Overview of the Four Representatives of the Four Groups of Cases

Cross-case matrix	Yonathan (LLL)	Patty (LLH)	Susan (HHL)	Venus (HHH)
Perceived feedback				
Quantity	<u>4.25</u>	<u>4.13</u>	*5.00*	*4.75*
Quality	<u>3.00</u>	*4.20*	*5.00*	*4.80*
Do with feedback	<u>3.50</u>	*4.50*	<u>3.63</u>	*4.38*
Observed feedback				
Expected effective vs. expected ineffective elements	*89–11*	<u>79–21</u>	*94–6*	*87–13*
Effectiveness vs. ineffectiveness of expected effective elements	<u>30–70</u>	*52–48*	*50–50*	*54–46*
Effectiveness vs. ineffectiveness of expected ineffective elements	37–63	<u>13–87</u>	<u>16–84</u>	*43–57*

Notes. The participants' individual scores on the AEQ subscales are shown; the findings of their observed feedback are expressed in percentages.
Underlined numbers are lower than average; italicized numbers are higher than average; non-marked numbers are average.

et al., 2007; Moore & McCabe, 2003). Results of these tests show whether the perceptions of feedback changed from session to session. Concerning the interviews, the participants' answers to the semistructured questions were compared, showing similarities and differences among them.

In order to answer the main research question, we constructed a case-ordered descriptive meta-matrix (Miles & Huberman, 1994). We used the results of the questionnaires to construct four groups of cases within this meta-matrix. A summarized excerpt of this meta-matrix is shown in Table 1. Each case consists of one teacher's individual scores on the three subscales of the AEQ and the findings of their individual observations from one session. There were 21 cases in total (see section 'Data Collection'). The four groups were constructed based on the results of the questionnaires only, and we subsequently investigated whether the observation findings corroborated the findings of the questionnaires.

Results

This section contains four parts: results of the observations (research question 1(a)); results of the questionnaires (research question 1(b)); results of the interviews (research question 1(b)); and, finally, the results of the case-ordered descriptive meta-matrix (main research question).

Observations

In this section, we first present the results of the feedback dimensions (the first main step to answer research question 1(a), see section 'Data-Analysis'). Then, we address the number of feedback elements, and subsequently, we focus on the effectiveness of these feedback elements (the second main step to answer research question 1(a), see section 'Data-Analysis').

Feedback dimensions. Table 2 shows the descriptives of the feedback dimensions.

Regarding goal-directedness, the Kruskall–Wallis test showed that the dimension did not differ significantly among the phases ($\chi^2 = 2.239$, df $= 2$, $p = .326$). The mean ranks for this dimension were 21.17 for observation phases, 23.78 for analysis phases, and 29.00 for reflection phases. These mean ranks differed slightly, which may indicate potential differences between the phases. Therefore, additional Mann–Whitney tests were performed. However, there were no significant differences between the phases (observation–analysis: $U = 143.00$, $p = .547$; observation–reflection: $U = 67.00$, $p = .150$; analysis–reflection: $U = 76.00$, $p = .301$). This illustrated that the degree of goal directedness did not differ between the phases.

Table 2
Descriptives of the Dimensions for Each Phase

Phases	*M*	SD	Minimum	Maximum	*n* (Phases)	*n* (Utterances)
Goal vs. nongoal/person-directed						
Observation	1.77	.98	−.007	4.00	18	335
Analysis	1.89	.82	.34	3.33	18	1,601
Reflection	2.15	.83	.19	3.11	11	479
Specific vs. general						
Observation	.87	.55	.00	2.00	18	335
Analysis	.74	.38	−.003	1.56	18	1,601
Reflection	.83	.46	.17	1.78	11	479
Detailed vs. nondetailed						
Observation	.95	.49	.29	2.00	18	335
Analysis	1.09	.33	.51	1.56	18	1,601
Reflection	1.34	.53	.17	2.13	11	479
Positive vs. negative						
Observation	.06	.11	−.14	.28	18	335
Analysis	.17	.17	.00	.61	18	1,601
Reflection	.39	.37	.00	1.33	11	479

The dimension of specificity did not differ significantly among the phases ($\chi^2 = .084$, df = 2, and $p = .959$). The mean ranks for this dimension were 24.56 for observation phases, 23.28 for analysis phases, and 24.27 for reflection phases. The dimension of details differed significantly among the phases ($\chi^2 = 6.032$, df = 2, and $p = .049$). The mean ranks for this dimension were 18.64 for observation phases, 24.83 for analysis phases, and 31.41 for reflection phases. This implied that in the observation phases, fewer details were provided in the feedback than in the analysis phases and that fewer details were provided during the analysis phases than in the reflection phases. The dimension positivity differed significantly among the phases ($\chi^2 = 12.330$, df = 2, and $p = .002$). The mean ranks for this dimension were 16.36 for observation phases, 25.42 for analysis phases, and 34.18 for reflection phases. This suggested that feedback in the observation phases was more neutral whereas in the analysis phases, more positive feedback was provided and even more positive feedback was provided in the reflection phases.

Feedback elements. Table 3 shows that most elements (e.g. closed questions, summarizing, and hinting) were provided during the analysis phases. This is not surprising, because the analysis phases took more time than the other phases and as a consequence, more utterances were spoken. Generally, more expected effective elements were provided than expected ineffective elements.

Table 3
Number of Expected (In)effective Elements for the Phases

Feedback elements	Observation	Analysis	Reflection
Expected effective elements			
Guiding questions	52	115	78
Open-ended questions	16	142	52
Closed questions	55	217	67
Solution-focused questions	5	67	19
Continuous questioning	28	215	63
Summarizing	14	158	45
Acknowledging	8	66	35
Expected ineffective elements			
Evocative questions	3	13	3
Hinting	2	93	12
Judging	0	17	3
Finishing sentences	2	41	9
Providing own example	0	26	11

Effectiveness of feedback elements. The key questions here were: (a) did expected effective elements lead to more effective dimensions? and (b) did expected ineffective elements lead to more ineffective dimensions?

The second column of Table 4 shows how many times the elements evoked the four dimensions to be more effective (*yes*) or not (*no*). As the table indicates, in most cases the expected effective elements caused the dimensions to be more effective and expected ineffective elements caused the dimensions to be less effective. Closed questions, summarizing, and acknowledging, however, had a different effect than was expected—the dimensions were not affected at all, or they became less effective.

Table 4
The Effectiveness of Elements on Dimensions within the Phases

| Feedback elements	Effectiveness	Observation	Analysis	Reflection
Expected effective elements				
Guiding questions	Yes	27	48	21
	No	12	37	27
Open ended questions	Yes	11	106	44
	No	8	38	10
Closed questions	Yes	24	93	37
	No	28	125	36
Solution-focused questions	Yes	4	49	17
	No	1	20	5
Continuous questioning	Yes	18	147	49
	No	9	104	23
Summarizing	Yes	5	70	17
	No	6	74	24
Acknowledging	Yes	1	31	13
	No	8	45	23
Expected ineffective elements				
Evocative questions	Yes	1	6	2
	No	2	4	1
Hinting	Yes	1	38	5
	No	1	59	12
Judging	Yes	0	4	2
	No	1	12	1
Finishing sentences	Yes	1	14	2
	No	1	38	7
Providing own example	Yes	0	7	6
	No	0	20	5
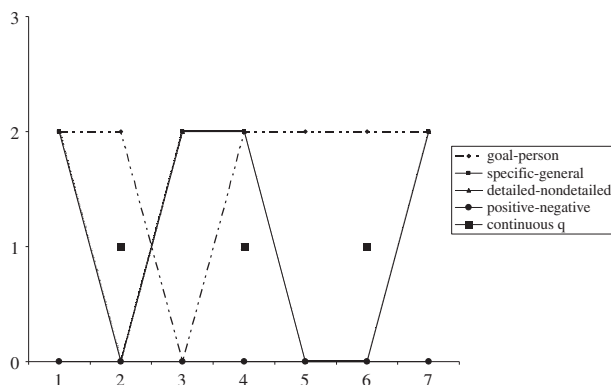
Figure 1. An example of the pattern of ineffective continuous questioning that turned into effective continuous questioning. The dimension specificity has the same line as the dimension details.

Forty percent of the expected effective elements were indeed effective and 60% were not effective. Of the expected ineffective elements, 33% were effective and 67% were indeed ineffective.

A pattern that illustrates the percentages of effectiveness of the expected effective elements (i.e. 40–60%) emerged frequently in all teachers' turns. This pattern showed that two or three expected effective elements were not effective, and then, a third or fourth element was effective. This often occurred with continuous questioning (see Figure 1).

The continuous questioning at point two influenced the dimension of goal directedness to decrease and the other dimensions remained at the same level (comparison of point one and three; see Figure 1). Likewise, the continuous questioning at point four did not influence the dimensions; they remained at the same level, and consequently, this continuous questioning was also not effective. However, the final continuous questioning at point six stimulated the dimensions of specificity and details to rise and did not influence the dimensions of goal directedness and positivity. Therefore, this final continuous questioning was effective.

Questionnaires

Table 5 indicates that the participants' perceptions of feedback only changed slightly from session to session. None of these changes were significant ($Z_{quantity} = -.962$, $p = .336$; $Z_{quality} = -.414$, $p = .679$; and $Z_{what\ to\ do\ with\ feedback} = -.530$, $p = .596$). Quantity was perceived better than quality, and the participants rated what they did with feedback slightly lower than its quantity and quality.

Interviews

Four of the seven interviewed teachers said that they had positive experiences with feedback provided in the sessions. The teachers held different views

Table 5
Descriptives of the AEQ Subscales

AEQ	*M*	SD	Minimum	Maximum	Number of items	*n*
First questionnaire						
Quantity	4.33	.44	3.63	5.00	8	10
Quality	4.12	.57	3.20	4.80	5	10
Do with feedback	4.01	.29	3.63	4.50	8	10
Second questionnaire						
Quantity	4.43	.32	4.00	5.00	8	9
Quality	4.18	.56	3.00	4.80	5	9
Do with feedback	4.07	.50	3.50	4.88	8	9
Third questionnaire						
Quantity	5.00	.00	–	–	8	2
Quality	5.00	.00	–	–	5	2
Do with feedback	4.13	.71	3.63	4.63	8	2

about how to define and describe effective feedback: three teachers described effective feedback as receiving hints on how to solve something; two teachers defined it as receiving compliments, and two teachers described it as receiving questions that aid to find your own solution. In addition, teachers described effective feedback as a reflection, a reaction, and perspectives from colleagues. Teachers had a more similar opinion about ineffective feedback: it contained a hint that is unusable or is too confronting. One teacher never had a negative experience with feedback. Furthermore, three teachers argued that feedback could never be ineffective, because one can always learn something. The teachers learned from the interplay of making video recordings of their teaching behaviors, the peer coaching from their colleagues who asked open-ended, solution-focused questions that guided the coached teachers to find their own solution, and receiving effective feedback.

Case Ordered Meta-Matrix

By applying the meta-matrix, we examined whether effective observed feedback was also perceived as more effective than observed ineffective feedback. In the meta-matrix, four groups were discerned. In the first group, teachers had lower than average scores on the three AEQ subscales. This LLL group[1] held six cases. Teachers in the second group scored lower than average twice and higher than average once on the AEQ subscales. This LLH group consisted of four cases. In the third group, teachers had two higher than average

---

[1]The name of the groups is constructed based upon the scores on the AEQ, where /L/ means lower than average and /H/ means higher than average.

and one lower than average scores on their AEQ. This HHL group held five cases. Teachers in the fourth group had scores higher than average on all subscales of the AEQ. This HHH group consisted of six cases.

In order to discern whether these findings corroborated with the results of the observations, we chose a representative from each group (see Table 1). First, the average scores of these four representatives reflected the scores on the AEQ of their individual groups. Even though Yonathan (representative of the LLL group) received more expected effective elements than Patty (representative of the LLH group), the latter's ratio of effectiveness of the expected effective elements was much higher than the former. Only 30% of expected effective elements were effective in Yonathan's turn whereas the average was 40%; and Patty had a ratio of 52% of effectiveness. However, the ratio of effectiveness of expected ineffective elements was favorable for Yonathan. His ratio was 37%, which was about average, while Patty's ratio was only 13%.

The main difference between Patty (representative of the LLH group) and Susan (representative of the HHL group) lies in the amount of received elements. Of all elements Patty received, 79% were expected to be effective, which was lower than average. Meanwhile, 94% of all elements Susan received were expected to be effective. The ratios of effectiveness of both expected effective and ineffective elements were quite similar for Patty and Susan.

The main difference between Susan (representative of the HHL group) and Venus (representative of the HHH group) lies in the ratio of effectiveness of the expected ineffective elements. Compared to an average of 33%, only 16% of all ineffective elements Susan received were effective, while 43% of all ineffective elements Venus received were effective.

Discussion

In our paper, we focused on observed and perceived feedback on practice among teachers and the coherence between them. We triangulated data from observations, questionnaires, and interviews with 12 Dutch primary school teachers. The teachers participated in a peer coaching program (Jeninga, 2003). We aimed to investigate whether observed effective feedback was also perceived as more effective than observed ineffective feedback; and whether this observed ineffective feedback was perceived as less effective.

Effective feedback was defined as goal directed, specific, detailed, and neutral; ineffective feedback was defined as nongoal/person directed, vague, nondetailed, and either too positive or too negative. Because feedback was communicated in peer coaching sessions, feedback was continuously provided by all participants, including the coached teachers. In order to investigate how participants can stimulate each other to give goal-directed, specific, detailed, and neutral feedback, and avoid nongoal or person-directed, vague, nondetailed, and either too positive or too negative feedback, feedback elements were also studied. These elements (e.g. open-ended and

guiding questions, acknowledging, and hinting) were expected to either heighten or lower the feedback dimensions. In order to observe feedback dimensions and elements, the TFOS (Thurlings et al., 2012) was used. In the observations, the three phases of observation, analysis, and reflection (White, 2009) were also used.

Concerning research question 1(a), the results on the four feedback dimensions showed that the dimensions of details and positivity differ among the phases. Both dimensions were low in the observation phases, moderate in the analysis phases, and high in the reflection phases. Second, the results showed that more expected effective feedback elements were provided than expected ineffective elements. Third, the average effectiveness of expected effective elements was 40% and the average effectiveness of expected ineffective elements 33%.

Contrary to our expectations based on previous literature, closed questions, summarizing, and acknowledging were generally not effective. It therefore may be better for peer coaches to try to formulate open-ended questions, because these questions are usually shown to be effective. Summarizing is probably not effective, because it wraps things up before turning to a new issue. Effective summaries were accompanied by an open-ended question, which provoked coached teachers to elaborate more and probably thereby affected the feedback dimensions. Acknowledgment in itself may not be effective, but might be necessary in terms of relatedness (Ryan & Deci, 2000). If coached teachers felt that their peer coaches acknowledged them in their goals, the coached teachers were more receptive for questions, which helped them to tackle the problem. Most expected ineffective elements (67%) were indeed ineffective.

In general, teachers had positive perceptions of the feedback. The interviews also indicated that teachers were positive about the feedback they received from their peer coaches. They had different views on what effective feedback is, but had similar views on what ineffective feedback is: an unusable hint or too confronting comments. Effective feedback was seen either as useful hints, perspectives from colleagues, compliments, questions, and a reflection or reaction.

The meta-matrix—of which a summarized excerpt was shown in Table 1—used in order to address the main research question shows that the findings of the observations also differed between the representatives of the four groups, corroborating with the results of the questionnaires. In other words, if observed feedback was determined to be effective, the receiver perceived feedback as more effective while if observed feedback was less effective, and that feedback was perceived as less effective.

These results indicated that it is not only important to provide expected effective elements, but also that they really are effective in terms of the dimensions. If these requisites are met, participants are more likely to perceive feedback as more effective, which in turn may lead to better learning outcomes.

Conclusions

The literature study, which mainly contained articles on teacher to student feedback, is confirmed for teacher-to-teacher feedback. This implies that effective feedback is similar for any kind of learner.

In addition, the elements of the peer coaching program were proven as an effective professional development activity: watching video excerpts, asking open-ended, solution-focused questions, acknowledging coached teachers, and helping them to tackle their goals were confirmed as parts of an effective feedback environment.

Furthermore, the expected effective feedback elements were overall effective and expected ineffective elements were overall ineffective. Our research indicated that two aspects are crucial when feedback is provided in peer coaching programs. As an implication, we suggest that enough expected effective elements should be provided and their actual effectiveness ratio should be 40% or more. In addition, if feedback providers give expected ineffective elements, it is important that these elements somehow provoke receivers to be goal directed, specific, detailed, and neutral.

Future Research

In the future, the interaction between the actual content of feedback and the effectiveness of the elements can be investigated. In such a study, we may further explain why certain elements are sometimes effective and sometimes not.

In addition, it could be examined whether feedback providers matter. Do receivers react differently to feedback from their process supervisors or their peer coaches?

Finally, teachers preferred hints, though most hints were shown to be not effective. In future research, we plan to focus on how teachers experience receiving hints and whether there are different styles of providing hints: would an authority hint (e.g. "You should …") or a constructive hint (e.g. "Maybe, you could …") prove beneficial, and what would be their effects on feedback processes?

Notes on contributors

Marieke Thurlings has just received her PhD at the Open Universiteit in the Netherlands. This paper reports on one of the studies she conducted. Her PhD focused on peer-to-peer feedback among teachers, both in face-to-face and online settings. Her main research interests are feedback, coaching, professional development, and how learning can be enhanced.

Marjan Vermeulen is an associate professor at the Teachers University, a project of the Open Universiteit in the Netherlands. She also is employed as a knowledge manager and researcher at the KPC group which is an educational consultancy organization. Her expertise lies in supporting educational practice as well as educational research in the field of professional development of teachers, the

teacher as a professional, the school as the teachers' work environment, school leadership, educational development and innovation.

Theo Bastiaens is a full professor at the Institute for Educational Science and Media Research of the Fernuniversität in Hagen, Germany. Next to this, he is part-time professor of Educational Technology at the Open Universiteit in the Netherlands. His specific research interest is in Instructional Design and E-learning. He has published extensively in these areas.

Sjef Stijnen holds a position as a full professor in distance education at LOOK (Scientific Centre for Teacher Research). His publications deal mainly with teacher shortages, teacher professional development, and the practical relevance of research on teaching.

References

Baarda, D.B., de Goede, M.P.M., & van Dijkum, C.J. (2007). *Basisboek statistiek met SPSS* [Basic book on statistics with SPSS]. Groningen: Noordhoff Uitgevers.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–68. doi: 10.1080/0969595980050102

Black, P., & Wiliam, D. (1998b). Inside the black box. *Phi Delta Kappan, 80*(2), 139–147.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage.

Gallacher, K. (1997). Supervision, mentoring and coaching: Methods for supporting personnel development. In P.J. Winton, J.A. McCollum, & C. Catlett (Eds.), *Reforming personnel preparation in early intervention: Issues, models, and practical strategies* (pp. 191–214). Baltimore, MD: Brooks.

Gibbs, G., & Simpson, C. (2003, September). *Measuring the response of students to assessment: The Assessment Experience Questionnaire*. Paper presented at the the International Improving Student Learning Symposium, Hinckley.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education, 1*, 3–31.

Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. doi: 10.3102/003465430298487

Jackson, P., & McKergow, M. (2002). *Oplossingsgericht* [Solution-focused thinking]. Zaltbommel: Thema.

Jeninga, J. (2003). Peer coaching: "Van en met elkaar leren" als krachtig leermiddel ter bevordering van integrale leerlingbegeleiding en schoolontwikkeling [Peer coaching: Learning from and with each other as a powerful learning tool to stimulate integrated student counseling and school development]. In J. Fanchamps & J. van de Sanden (Eds.), *Integraal ondersteunen van een verni-euwd VMBO* [Integrated support of a renewed secondary vocational education] (pp. 25–32). Antwerpen: Garant.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Moore, D.S., & McCabe, G.P. (2003). *Introduction to the practice of statistics* (4th ed.). New York, NY: W.H. Freeman.

Mory, E.H. (2003). Feedback research revisited. In D.H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 745–783). New York, NY: MacMillan Library Reference.

OECD. (2002). *Education at a glance 2002. OECD indicators*. Paris: OECD.

Ryan, R.M., & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78. doi: 10.1037110003-066X.55.1.68

Scheeler, M.C., Ruhl, K.L., & McAfee, M.K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education, 27*(4), 396–407. doi: 10.1177/088840640402700407

Schelfhout, W., Dochy, F., & Janssens, S. (2004). The use of self, peer, and teacher assessment as a feedback system in a learning environment aimed at fostering skills of cooperation in a entrepreneurial context. *Assessment and Evaluation in Higher Education, 29*(2), 177–201. doi: 10.1080/0260293042000188465

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. doi: 10.3102/0034654307313795

Smith, P.L., & Ragan, T.J. (1993). Designing instructional feedback for different learning outcomes. In J.V. Dempsey & G.C. Sales (Eds.), *Interactive instruction and feedback* (pp. 75–103). Englewoods Cliffs, NJ: Educational Technology.

Thurlings, M., Vermeulen, M., Kreijns, C., Bastiaens, T.J., & Stijnen, P. (2012). Development of the Teacher Feedback Observation Scheme: Evaluating the quality of feedback in peer groups. *Journal of Education for Teaching, 38*(2), 193–208. doi: 10.1080/02607476.2012.656444

Tillema, H.H., & Smith, K. (2000). Learning from portfolios: Differential use of feedback in portfolio construction. *Studies in Educational Evaluation, 26*(3), 193–210. doi: 10.1016/S0191-491X(00)00015-8

Weaver, M.R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment and Evaluation in Higher Education, 31*(3), 379–394. doi: 10.1080/02602930500353061

White, S. (2009). Articulation and re-articulation: Development of a model for providing quality feedback to pre-service teachers on practicum. *Journal of Education for Teaching, 35*(2), 123–132. doi: 10.1080/02607470902770914