

Investigating feedback among teachers : focusing on observed and perceived feedback

Citation for published version (APA):

Thurlings, M. C. G., Vermeulen, M., Bastiaens, T. J., & Stijnen, P. J. J. (2011). *Investigating feedback among teachers : focusing on observed and perceived feedback*. 2011 Annual Meeting American Educational Research Association, AERA 2011, New Orleans, Louisiana, United States.

Document status and date:

Published: 01/01/2011

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.



From the

AERA Online Paper Repository

<http://www.aera.net/repository>

Paper Title Investigating Feedback Among Teachers: Focusing on Observed and Perceived Feedback

Author(s) Marieke Thurlings, Open University of the Netherlands; Theo Bastiaens, Fern University in Hagen; Sjef Stijnen, Open University of the Netherlands; Marjan Vermeulen, Open University of the Netherlands

Session Title School Climate, Organizational Structures, and Contextual Factors Impacting Mentoring

Session Type Poster Presentation

Presentation Date 4/11/2011

Presentation Location New Orleans, Louisiana, USA

Descriptors Elementary Schools, Peer Interaction/Friendship, Teacher Education - In-Service/Professional Development

Methodology Mixed Method

Unit Division K - Teaching and Teacher Education

Each presenter retains copyright on the full-text paper. Repository users should follow legal and ethical practices in their use of repository material; permission to reuse material must be sought from the presenter, who owns copyright. Users should be aware of the [Ethical Standards of the American Educational Research Association](#).

Citation of a paper in the repository should take the following form:
[Authors.] ([Year, Date of Presentation]). [Paper Title.] Paper presented at the [Year] annual meeting of the American Educational Research Association. Retrieved [Retrieval Date], from the AERA Online Paper Repository.

Investigating Feedback among Teachers: Focusing on Observed and Perceived Feedback

Marieke Thurlings^{a*},
Marjan Vermeulen^a,
Theo Bastiaens^{ab}, &
Sjef Stijnen^a

^a Ruud de Moor Centre,
Open University of the Netherlands
^b FernUniversität Hagen, Germany

Paper presented at American Educational Research Association (AERA)
April 8th – April 12th, 2011
New Orleans, LA

*Corresponding author: Marieke Thurlings, MSc. Ruud de Moor Centre, Open University of the Netherlands. Post-box 2960; NL-6401 DL Heerlen, the Netherlands.
E-mail: Marieke.Thurlings@ou.nl; Telephone: 0031455762397; Fax: 0031455762782.

Abstract

This paper focuses on feedback among teachers. The research triangulates data from videotaped peer coaching sessions, questionnaires and interviews with four groups of teachers. All teachers used the VIP procedure, which emphasizes reciprocal peer coaching using video recordings of teaching behaviors and solution-focused thinking. The study provides insights into how feedback elements influenced feedback dimensions, and which feedback elements were effective in doing so, and which were not. The study also presents how teachers experienced the feedback processes. In addition, four teachers are used as cases in order to link the observed and perceived feedback, revealing that effective observed feedback was also perceived as effective feedback; and that ineffective observed feedback was also perceived as ineffective. These conclusions can be used to guide teachers in feedback processes.

Objectives

Feedback can be very effective in supporting students' learning (Hattie, 2009; Hattie & Timperley, 2007). Most research on feedback focuses on feedback from teachers to their students (Hattie & Timperley, 2007; Mory, 2003). However, little research has investigated feedback among teachers (Scheeler, Ruhl, & McAfee, 2004), even though feedback among teachers can serve as an effective tool within their professional development.

Therefore, this study focuses on feedback among teachers within a specific professional development activity, i.e. peer coaching. Specifically, this study addresses the following research questions: 1) Which and to what extent do effective and ineffective feedback dimensions and feedback elements occur in the peer coaching activities, and how do they relate to one another? 2) How did the teachers perceive feedback? 3) To what extent does the observed feedback relate to the perceived feedback?

Theoretical framework

Feedback in this study is regarded as “information that allows for comparison between an actual and a desired outcome” (Mory, 2003, p. 746). Characteristics of feedback can be synthesized in five dimensions (Thurlings, Vermeulen, Kreijns, Bastiaens, & Stijnen, in press):

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. non-detailed (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).

Based upon these dimensions synthesized from literature, effective and ineffective feedback can be described. First, it is assumed that goal-directed feedback is more effective than person-directed feedback (Black & Wiliam, 1998a; Gibbs & Simpson, 2004). Second, it is argued that specific feedback is more effective than general feedback (Black & Wiliam, 1998a; Mory, 2003; Scheeler et al., 2004), although general advice on how to improve one's actions in the future is effective too (Black & Wiliam, 1998b; Weaver, 2006). Third, it is

assumed that feedback that focuses on details is more effective than feedback lacking details (Gibbs & Simpson, 2004; Scheeler et al., 2004; Weaver, 2006). Fourth, it is unclear whether positive feedback is more effective than negative feedback or not. Some argue that feedback should be positive (Scheeler et al., 2004; Tillema & Smith, 2000), others argue that negative feedback can motivate learners in their learning process (Schelfhout et al., 2004), or argue that feedback that is balanced between positive and negative comments is more effective (Weaver, 2006). Fifth, immediate feedback is considered more effective than delayed feedback (Mory, 2003). In addition, delayed feedback is less effective than feedback that is still relevant for the learner (Black & Wiliam, 1998a; Scheeler et al., 2004). Therefore, based on feedback literature, it is argued that goal-directed, specific, and detailed feedback that is balanced between positive and negative comments is more effective than feedback that is person-directed, general, vague, and either too positive or too negative. Dimension 5 (i.e., timing) will not be included in this study, because the feedback is always communicated during the peer coaching sessions, and in addition, adequate timing of feedback is difficult to observe.

Note that the literature described stems from research on feedback from teachers to students. The research on feedback between teachers is in its infancy (Scheeler et al., 2004), and therefore, research on feedback from teachers to students needs to be transferred to research on feedback between teachers.

Context

Twelve Dutch primary school teachers participated in the study. They were divided into four groups. Nine teachers (i.e., three groups) were from the same primary school, whereas the other group was from another primary school. Two men and ten women participated. Their mean age was 30.7 years old ($sd = 6.7$ years). One group of teachers held three peer coaching sessions, and the other groups held two peer coaching sessions.

All groups used the VIP procedure (Video Intervision Peer coaching; Jeninga, 2003). The VIP procedure combines reciprocal peer coaching (B. Showers & Joyce, 1996), video recordings (Brophy, 2004), and solution-focused thinking (Jackson & McKergow, 2002), and has a cyclic workflow. In the first step of this workflow, teachers decide which teaching behavior they want to improve and videotape a (part of a) lesson, in which the behavior occurs (e.g. being more consistent in applying classroom rules). In the second step, the teachers meet and coach each other (i.e., they all get their turn being the coached teacher, and the two other teachers are peer coaches). During the session, the teachers use the concept of

solution-focused thinking (Jackson & McKergow, 2002) to assist the coached teachers to define their goal and actions that could lead to improved behavior. The concept of solution-focused thinking suggests that asking open-ended and solution-focused questions, confirming what the teachers already do well, and focusing on what teachers can do will aid in formulating and achieving those actions. In the third step, teachers try out the formulated actions and make another video recording showing the changed behavior. In the final step, the teachers meet again and view the video recording of the changed behavior. The coached teachers provide themselves feedback, and next, the coaching teachers provide feedback as well. After the feedback, the coached teachers evaluate their changed behavior by providing themselves with a mark out of 10. The mark indicates how satisfied the teachers are with their changed behavior. They are also asked to explain why they gave themselves this mark, and elaborate on what they can do to heighten one mark (e.g. to get an 8, instead of a 7). Then, the teachers decide whether they are satisfied or not. If teachers are satisfied, they can select another teaching behavior that they want to improve, or in other words start a new workflow. If teachers are not satisfied, their goal and actions are readdressed and can be reformulated. They will then try to change their behavior over and over again, until they are satisfied.

Peer coaching groups are joined by a process supervisor (Jeninga, 2003). This person acts as a chairman, models coaching behavior using solution-focused thinking, and explicitly reflects on the coaching behavior of the teachers. In this study, each group is joined by a different process supervisor (two males; two females).

Methods and data sources

The peer coaching sessions were videotaped. They were transcribed and analyzed using the Teacher Feedback Observation Scheme (TFOS; Thurlings, et al., in press), which was developed to observe feedback provided within the peer coaching sessions. TFOS is based upon the feedback literature as described above, and measures four dimensions of feedback as well as feedback elements. These elements are several types of questions and coaching skills. The types of questions are open-ended, closed, guiding, solution-focused, and continuous questions. The coaching skills are summarizing and acknowledging. These questions and coaching skills are expected to be effective, i.e. are expected to heighten the feedback dimensions. Furthermore, judging, hinting, finishing sentences, evocative questions, and providing examples from one's own classroom or experience are scored. These elements are expected to lower the feedback dimensions.

After each peer coaching session, teachers filled in a questionnaire, which consisted of four parts. The first part aimed to investigate how the teachers perceived the feedback within the session. To measure the perceived feedback, three subscales of the Assessment Experience Questionnaire (AEQ; Gibbs & Simpson, 2003) were used, namely 'quantity of feedback' ($\alpha = 0.693$), 'quality of feedback' ($\alpha = 0.662$), and 'what a student does with the feedback' ($\alpha = 0.614$). The items were adapted to fit the VIP procedure. One item of the scale 'quality of feedback' was deleted, because this improved Chronbach's α from 0.473 to 0.662, and because the item (The feedback mainly tells me how well I am doing in relation to others) is irrelevant to the VIP procedure. Therefore, the numbers of the items of the scales respectively were eight, five, and eight. These items were answered on a five-point scale.

The second part of the questionnaire consisted of six questions, which focus on how the teachers experienced the session. For instance: "I learned a lot when I was the coached teacher" and "The session gave me new insights". These questions could be answered by providing a mark out of 10. The third part measured teachers' self-efficacy, using the short form of the Teacher Self Efficacy Scale (TSES, Tschannen-Moran & Woolfolk-Hoy, 2001), that contains three subscales: classroom management, instructional strategies, and student engagement. These items are answered on a nine-point scale. The final part consisted of questions on gender, birth date, years of experience as a teacher, years of experience with the VIP procedure etc. This final part was only included in the questionnaire to be filled in after the first peer session. In the questionnaires to be filled in after the second and third sessions, teachers only filled in their birth date. Using this data, the two questionnaires could be connected to the individual teachers.

After the completion of the questionnaires, the teachers were approached by email to participate in a semi-structured interview. Seven teachers agreed to be interviewed. Five teachers indicated that they did not have time, because the interviews were scheduled during the week before the summer holidays. The interview was held by the telephone and took about 20 to 30 minutes. The interview aimed to gather qualitative data on how the teachers had experienced feedback. The questions were focused at general impressions of the VIP procedure, learning outcomes, received feedback, and linking the learning outcomes with the received feedback. A report was made for each interview, which was sent to the individual teachers, so that they could check whether the report reflected their opinions. It was possible to link the observations, questionnaires, and interviews of all individual teachers.

Table 1 provides an overview of all teachers within their groups, and the data that was collected for each teacher.

		Session 1				Session 2				Session 3					
		Questionnaire				Questionnaire				Questionnaire				Interview	
Group Nick	Olivia	O *	A	R	Yes **	O	A		Yes	n.a.				n.a.	Yes
	Patty Quinta	O A	RA O		Yes Yes	O R	R A		Yes Yes	n.a. n.a.				n.a. n.a.	Yes Yes
Group Norbert	Ricky	O	A		Yes	O	R	A	Yes	n.a.				n.a.	No
	Susan			n.a.	n.a.	O	A		Yes	R				Yes	No
	Tiffany	O	A		Yes		n.a.		n.a.	O	R	O	R	Yes	Yes
Group Natasha	Ursella	O	A		No	O	R		Yes	n.a.				n.a.	Yes
	Venus	O	A		Yes		n.a.		n.a.	n.a.				n.a.	No
	Wanda	O	A		Yes	O	R		Yes	n.a.				n.a.	No
Group Nicole	Xavier	A	O	A	Yes	RA			Yes	n.a.				n.a.	Yes
	Yonathan	O	A		Yes	R	A		Yes	n.a.				n.a.	No
	Zoey	O	A		Yes		n.a.		n.a.	n.a.				n.a.	Yes

Table 1. Overview of collected observations and the sequence of phases of each teachers' turns and overview of collected questionnaire and interview data.

* O= Observation phase, A = Analysis phase, R = Reflection phase

** yes = this type of data was collected for this teacher; no = this type of data was not collected for this teacher; n.a. = this type of data was not applicable.

Data-analysis

After the observations were transcribed, each teacher's turn being the coached teacher was divided into the phases of White's model of process feedback; Observation, Analysis, and Reflection (White, 2009). White developed a quality feedback process model in the context of teacher education, consisting of three stages. The Observational stage "is derived from lecturers observing students while they are teaching on practicum" (2009, p. 128). In the VIP procedure, this stage is addressed in the second and fourth step, when the peer groups watch and discuss the video excerpt. In White's Analysis stage, lecturers coach students in formulating goals and actions to improve their practices. In the VIP procedure, this stage is mainly addressed in the second step, when teachers are coached to formulate their goals and actions. White's Reflection stage consists of a debriefing session, in which lecturers provide written and oral feedback to the students regarding the actions that they initiated. The fourth step of the VIP procedure is similar to this stage, in which the changed behavior is evaluated. After the turns were divided according to these phases, all phases of teachers' turns were summarized in data-matrices (Miles & Huberman, 1994).

The scoring of TFOS-dimensions and elements was done using Excel-files, for each phase in all teachers' turns. The scoring of the dimensions was done as follows: when an utterance was completely goal-directed, a 4 was assigned; when an utterance was goal-directed for 75% a 2 was assigned; when a utterance was goal-directed for 50% a 0 was assigned; when an utterance was person- or student-directed for 75% a -2 was assigned; and when an utterance was person- or student-directed for 100% a -4 was assigned. This method was also used for the other dimensions. To test whether the feedback dimensions differ between the phases, a Kruskal-Wallis test was performed, and an additional Mann-Whitney test was performed (Baarda, de Goede, & van Dijkum, 2007; Moore & McCabe, 2003). The questions, coaching skills, and other elements were assigned either a 1 or a -1, when the element respectively was expected to be effective and ineffective. Subsequently, for each phase of all teachers' turns, a timeline graphic was made. By putting the scoring of the dimensions of feedback on a time line and by indicating the types of questions on the same time line, it can be investigated if and how the feedback elements affect the feedback dimensions.

The effectiveness of each feedback element was determined by comparing the position of the feedback dimensions on the time lines before and after the feedback element. If the feedback dimensions were higher after the element than before, the feedback element was counted as effective. If the feedback dimensions were equal like before or lower than before the feedback element occurred, the element was counted as ineffective. Two premises were formulated:

1. Effective feedback elements (i.e. open-ended, closed, solution-focused, and guiding questions, continuous questioning, acknowledging, and summarizing) lead to more effective feedback dimensions
2. Ineffective feedback elements (hinting, judging, evocative questions, providing an example from one's own experience, and finishing sentences) lead to less ineffective feedback dimensions

The questionnaires were analyzed in SPSS using non parametric tests. To test whether the scales of the AEQ relate to the experience of the sessions, a Friedman test was performed (Baarda, et al., 2007; Moore & McCabe, 2003).

Finally, based upon the questionnaire, four outliers were chosen. These outliers were two participants that had an above-average perspective of the quantity and quality of the

feedback, and two participants that had a below-average perspective of the quantity and quality of the feedback. To investigate the relation between the observed feedback and the perceived feedback, the results of the observations and questionnaires were examined.

Results

This section comprises four parts. First, we describe the results of the observations, making an inventory of the dimensions and elements, and focusing on the effectiveness of the elements. Second, the results of the questionnaires are described. Third, the results of the interviews are addressed. Finally, four participants being outliers on the questionnaires are examined into more detail, and the relation between their observed and perceived feedback will be clarified.

Observations

Table 2 shows the means, standard deviations, minimum, and maximum of the feedback dimensions. The Kruskal-Wallis test showed that the dimension of goal-directedness did not differ significantly among the phases ($Chi^2 = 2.239$, $df = 2$, $p = 0.326$). The mean ranks for this dimension were 21.17 for Observation phase; 23.78 for Analysis phase; and 29.00 for Reflection phase. This means that the degree of goal-directedness did not differ between the phases. The Kruskal-Wallis test showed that the dimension of specificity did not differ significantly among the phases ($Chi^2 = 0.084$, $df = 2$, $p = 0.959$). The mean ranks for this dimension were 24.56 for Observation phase; 23.28 for Analysis phase; and 24.27 for Reflection phase. This suggests that the degree of specificity did not differ between the phases. The Kruskal-Wallis test showed that the dimension of details differs significantly among the phases ($Chi^2 = 6.032$, $df = 2$, $p = 0.049$). The mean ranks for this dimension were 18.64 for Observation phase; 24.83 for Analysis phase; and 31.41 for Reflection phase. This implies that in the Observation phase fewer details were provided in the feedback than in the Analysis phase; and that fewer details were provided in the feedback during the Analysis phase than in the Reflection phase. The Kruskal-Wallis test showed that the dimension positivity differs significantly among the phases ($Chi^2 = 12.330$, $df = 2$, $p = 0.002$). The mean ranks for this dimension were 16.36 for Observation phase; 25.42 for Analysis phase; and 34.18 for Reflection phase. This suggest that feedback in the Observation phase was more neutral, whereas in the Analysis phase more positive feedback was provided, and even more positive feedback was provided in the Reflection phase.

	Goal- vs. person/student-directed			Specific vs. general			Detailed vs. non-detailed			Positive vs. negative		
	O*	A	R	O	A	R	O	A	R	O	A	R
Mean	1.77	1.89	2.15	0.87	0.74	0.83	0.95	1.09	1.34	0.06	0.17	0.39
Sd	0.98	0.82	0.83	0.55	0.38	0.46	0.49	0.33	0.53	0.11	0.17	0.37
Minimum	-0.07	0.34	0.19	0.00	-0.03	0.17	0.29	0.51	0.17	-0.14	0.00	0.00
Maximum	4.00	3.33	3.11	2.00	1.56	1.78	2.00	1.56	2.13	0.28	0.61	1.33

Table 2. Means, standard deviations, minimum, and maximum of the feedback dimensions for each phase.

* O= Observation phase, A = Analysis phase, R = Reflection phase

Additional Mann-Whitney tests were performed to examine the dimension goal-directedness, because the mean ranks pointed at differences between the individual phases. However, there were no significant differences between the phases (Observation – Analysis: $U = 143.00, p = 0.547$; Observation – Reflection: $U = 67.00, p = 0.150$; Analysis – Reflection: $U = 76.00, p = 0.301$). This implies that the degree of goal-directedness did not differ between the phases.

	Guiding question	Open ended question	Closed question	Solution focused question	Continuous questioning	Summarizing	Acknowledging
OBSERVATION PHASE							
Nick (n = 5*)	20	2	5	2	4	3	1
Norbert (n = 6)	17	7	21	3	13	4	5
Natasha (n = 5)	11	5	22	0	6	2	1
Nicole (n = 3)	4	2	7	0	5	5	1
Total (n = 19)	52	16	55	5	28	14	8
ANALYSIS PHASE							
Nick (n = 6)	26	19	32	5	39	28	18
Norbert (n = 4)	54	47	82	31	82	66	23
Natasha (n = 3)	19	33	35	20	40	18	3
Nicole (n = 6)	16	43	68	11	54	46	22
Total (n = 19)	115	142	217	67	215	158	66
RELFECTION PHASE							
Nick (n = 4)	16	5	6	3	7	4	7
Norbert (n = 4)	45	25	39	8	32	26	16
Natasha (n = 2)	16	17	14	8	16	8	8
Nicole (n = 2)	1	5	8	0	8	7	4
Total (n = 12)	78	52	67	19	63	45	35

Table 3. The number of effective feedback elements, for each phase per group and in total.

* Between brackets, it is depicted which and how many phases occurred in the groups.

Tables 3 and 4 show the number of respectively effective and ineffective feedback elements, separate for each group and each phase, as well as for all groups. The phases seem to make a difference in what effective feedback elements emerge. In the Analysis phase, many more questions were posed, compared to the Reflection phase and to the Observation phase (see Table 3). Note that the Analysis phase took more time than the other phases, and therefore, these numbers are relative to this fact.

In Norbert's group, more effective feedback elements emerged, in comparison to the other groups. The amount of effective feedback elements in the other groups were usually equal to one another, and in some cases only a few effective feedback elements emerged, for instance in Nick's group only five solution-focused questions were posed in the Analysis phases (see Table 3).

	Evocative question	Providing own experience	Finishing sentences	Judging	Hinting
OBSERVATION PHASE					
Nick (n = 5*)	0	0	0	0	0
Norbert (n = 6)	0	0	1	0	2
Natasha (n = 5)	3	0	1	0	0
Nicole (n = 3)	0	0	0	0	0
Total (n = 19)	3	0	2	0	2
ANALYSIS PHASE					
Nick (n = 6)	4	17	15	9	25
Norbert (n = 4)	2	6	16	2	26
Natasha (n = 3)	5	1	5	3	8
Nicole (n = 6)	2	2	15	3	34
Total (n = 19)	13	26	41	17	93
REFLECTION PHASE					
Nick (n = 4)	0	1	1	0	11
Norbert (n = 4)	0	6	6	0	11
Natasha (n = 2)	3	4	1	3	1
Nicole (n = 2)	0	0	1	0	0
Total (n = 12)	3	11	9	3	12

Table 4. The number of ineffective feedback elements, for each phase per group and in total.

* Between brackets, it is depicted which and how many phases occurred in the groups.

A quick look at Table 4 indicates that less ineffective feedback than effective feedback elements occurred. Hinting was done frequently and finishing sentences too. Most ineffective feedback elements emerged in the Analysis phases. Evocative questions, judging, and finishing sentences hardly occurred in the Observation and Reflection phases, however, they

emerged relatively more in the Analysis phases. Compared to the other groups, only in Natasha's group, ineffective feedback elements occurred less in all phases.

The formulated premises focus on the influence of the (in)effective feedback elements on the feedback dimensions. Do expected effective feedback elements lead to more effective feedback dimensions, and do expected ineffective feedback elements lead to more ineffective feedback dimensions? The real effectiveness of each feedback element was determined by comparing the scores of the feedback dimensions in the utterances before and after the feedback element. If the feedback dimensions were higher after the element than before, the feedback element was counted as effective. If the feedback dimensions were equal like before or lower than before the feedback element occurred, the element was counted as ineffective.

Table 5 shows how many times the feedback elements evoked the feedback dimensions to be more effective or not. The table clearly indicates that in most cases, effective feedback elements did cause the feedback dimensions to be more effective, and ineffective feedback elements did cause the feedback dimensions to be less effective. Closed questions, summarizing and acknowledging, however, seemed to have had a different effect than was expected. The dimensions were not affected at all, or they became less effective.

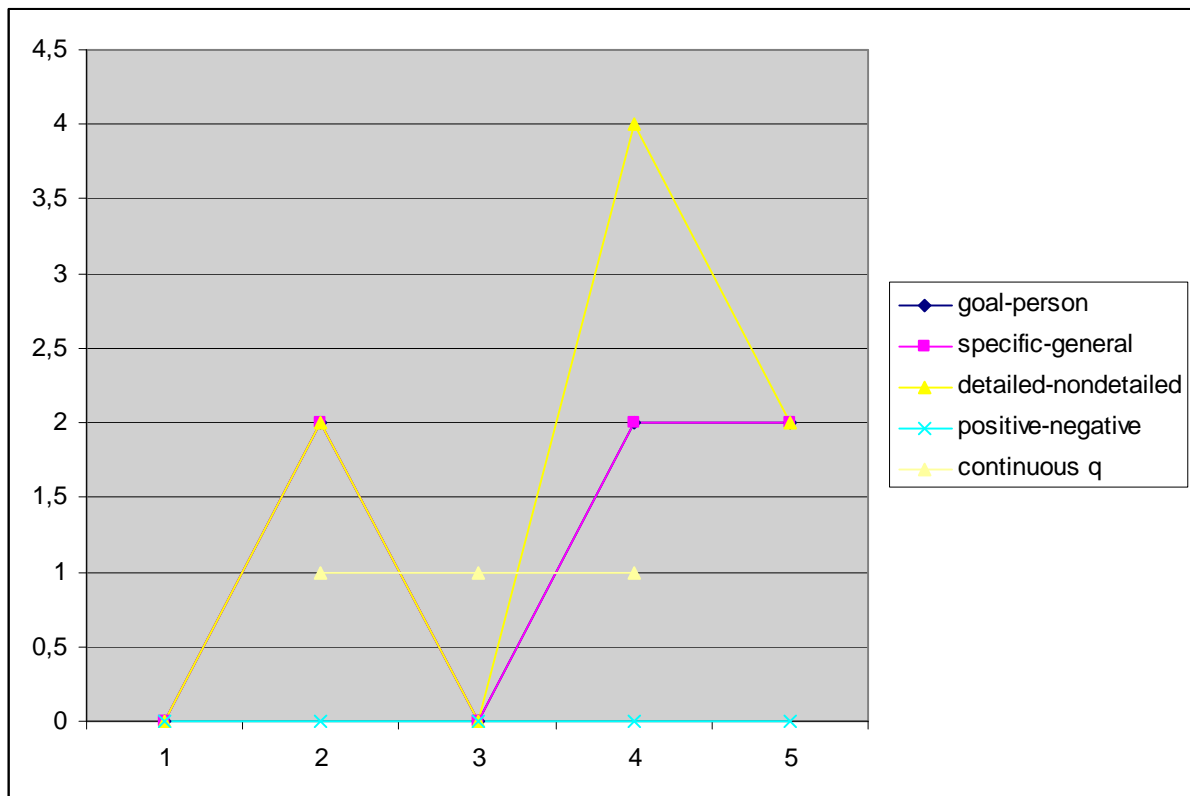
			OBSERVATION	ANALYSIS	REFLECTION
EFFECTIVE FEEDBACK ELEMENTS	Guiding question	Yes	27	48	21
		No	12	37	27
	Open ended question	Yes	11	106	44
		No	8	38	10
	Closed question	Yes	24	93	37
		No	28	125	36
	Solution-focused question	Yes	4	49	17
		No	1	20	5
	Continuous questioning	Yes	18	147	49
		No	9	104	23
	Summarizing	Yes	5	70	17
		No	6	74	24
Acknowledging	Yes	1	31	13	
	No	8	45	23	
INEFFECTIVE FEEDBACK ELEMENTS	Evocative question	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

Table 5. Do effective feedback elements lead to more effective feedback dimensions?

A pattern that occurred frequently was that about two to three expected effective feedback elements were not effective, and then, a third or fourth and identical feedback element was effective. This often happened with continuous questioning (see Figure 1). The first continuous questioning at point 2 did not affect the dimensions, the second continuous question at point 3, however, affected the dimensions of specificity and details to raise (comparison of point 2 and 4), and another continuous questioning occurred. The continuous questioning at point 4 only affected the dimension of details, however, in an ineffective way.

Figure 1. An example of the pattern of ineffective continuous questioning that turned into effective continuous questions, from Tiffany's turns.

NB: the dimension goal-directedness has the same line as the dimension specificity



Questionnaires

Table 6 indicates that the perceptions of the participants on quantity and quality of feedback, and what they did with the feedback only changed slightly from session to session. Quantity of feedback was perceived better than quality, and the participants rated what they do with feedback slightly lower than quantity and quality. The teachers rated the quantity and quality of feedback slightly higher after the second session, than after the first session. There was no difference on the scale what to do with feedback. It is not necessary to interpret the finding of

the third questionnaire at this stage, because only two teachers of Norbert's group had a third meeting.

		Mean	SD	Minimum	Maximum	Number of items	N
First questionnaire	Quantity	4.2917	0.45501	3.63	5.00	8	9
	Quality	4.1200	0.56725	3.20	4.80	5	10
	What to do with feedback	4.0125	0.29137	3.63	4.50	8	10
Second questionnaire	Quantity	4.3929	0.31810	4.13	5.00	8	7
	Quality	4.1778	0.56075	3.00	4.80	5	9
	What to do with feedback	4.0156	0.050195	3.63	4.88	8	8
Third questionnaire	Quantity	5.00	0.00	-	5.00	8	2
	Quality	5.00	0.00	-	5.00	5	2
	What to do with feedback	4.125	0.70711	3.63	4.63	8	2

Table 6. Means, standard deviations, minimum and maximum on the three AEQ subscales.

Table 7 reports the means for the questions on the experiences of the participants for each session. The Friedman test showed that the AEQ scales and the marks differed significantly, based on all sessions ($Friedman\ Chi^2 = 82.874, df = 8, p = 0.000$).

	First session	Second session	Third session
I learned a lot when I was the coached teacher	7.8571 (7*)	7.4286 (7)	8.0000 (2)
I learned a lot when I was the coaching teacher	7.7778 (9)	7.8889 (9)	7.5000 (2)
This session lived up to my expectations	7.6000 (10)	7.4444 (9)	7.5000 (2)
During this session, there was a good atmosphere	8.4000 (10)	8.4444 (9)	8.0000 (2)
The session gave me new insights	7.8000 (10)	7.5556 (9)	8.0000 (2)
The session gave me new learning goals	7.3750 (8)	6.8750 (8)	8.0000 (2)

Table 7. The means of the marks on the experience-questions for each session.

* Between brackets, the number of respondents to this item is depicted.

Interviews

Results of the interviews show that four out of the seven interviewed teachers told that they had positive experiences with feedback provided in the sessions. The teachers had different views of what effective feedback is. Three teachers described effective feedback as receiving hints on how to solve something, two teachers described effective feedback as receiving compliments, and two teachers described effective feedback 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 consistent opinion on what ineffective feedback is. Ineffective feedback contains a hint that is unusable or is too confronting. One teacher did not have a negative experience with feedback, and three teachers argued that feedback is never ineffective, because one can always learn something. The teachers learned from the interplay of making videotapes of their teaching behaviors, the peer coaching from their colleagues who asked open-ended, solution-focused questions that guided the coached teacher to find his own solution, and receiving effective feedback.

Cases: four outliers

This final part of the Results section examines four outliers into more detail. These outliers were selected based on their scores on the AEQ scales: two outliers with the lowest scores (i.e., Yonathan and Quinta) and two outliers with the highest scores (i.e., Tiffany and Wanda). Table 8 shows an overview of these outliers. The lower part of the table shows their data on the questionnaire and the upper part shows their results of the observations. This upper part is divided into three sections: feedback dimensions, the amount of feedback elements, and effectiveness of feedback elements.

The first section of the upper part of Table 8 reviews the *four feedback dimensions* (i.e., goal-directed vs. person-directed; specific vs. general; detailed vs. non-detailed; and positive vs. negative) as they appeared during the turns of the coached teachers. The section shows how many and with how much variation the feedback dimensions are lower, average, and higher compared to the means as shown in Table 2. The description of the individual outliers provides more detail on each dimension. The second section shows the *amount* of the given (in)effective *feedback elements*, for the Observation, Analysis, and Reflection phases. The third section focuses at the *real effectiveness* of the *feedback elements* on the feedback dimensions. The real effectiveness of each feedback element was determined by comparing the scores of the feedback dimensions in the utterances before and after the feedback element. This third section of the upper part of the table shows how many expected effective feedback elements were actually effective, or not; and how many expected ineffective feedback elements were indeed ineffective, or effective, during this teacher's turn.

We first describe the outliers' characteristics, and then examine their results of the observations and the questionnaire as shown in Table 8.

		Yonathan	Quinta	Tiffany	Wanda	
Observation	1. Feedback dimensions*	N°. lower 4 6 (¼ to 1 sd)	6 (¼ to 2 sd) 6 8 (¼ to 3 sd)	8 (1/3 to 1 sd) 2 10 (½ to 2 sd)	3 (¼ to ¾ sd) 7 6 (½ to 1½ sd)	
		O A R	O A R	O A R	O A R	
	2. Amount feedback elements*	Effective Non effective Total	13 112 11 1 13 14 125 11	15 77 38 15 115	11 130 65 16 15 11 146 80	7 68 45 8 6 7 76 51
	3. Effectiveness feedback elements**	Actual: EFF NOT	Actual: EFF NOT	Actual: EFF NOT	Actual: EFF NOT	
		Expected: EFF	Expected: EFF	Expected: EFF	Expected: EFF	
		Expected: NOT	Expected: NOT	Expected: NOT	Expected: NOT	
	AEQ session 1	Quantity Quality Do with	4.00 3.40 3.63	3.63 4.40 3.88	5.00 4.60 4.13	4.75 4.80 4.50
	AEQ session 2	Quantity Quality Do with	4.13 3.00 3.50	4.25 4.00 3.63	5.00 5.00 4.63	- 4.80 -
	Mark average session 1		7.4	9.0	8.0	8
	Mark average session 2		7.4	8.83	8.17	8
Questionnaire****	TSES session 1	Student engagement Instructional strategies Classroom management	6.75 7.25 7.75	6.50 6.50 6.50	9.00 8.25 9.00	7.25 7.50 7.00
	TSES session 2	Student engagement Instructional strategies Classroom management	6.75 7.25 7.75	6.50 6.50 6.50	8.50 7.75 8.25	7.50 7.50 7.25

Table 8. Overview of the individual results of the four outliers.

* How many feedback dimensions are lower, average, and higher compared to the means in Table 2; between brackets are the smallest and greatest standard deviation, expressing to what extent the dimensions varied from the averages in Table 2.

** The amount of the (in)effective feedback elements, for the Observation (O), Analysis (A), and Reflection (R) phases.

*** How much expected effective feedback elements are effective on the feedback dimensions, or not; and how much expected ineffective feedback elements are ineffective on the feedback dimensions, or effective

**** The mean scores on the scales of the AEQ (Perceived quantity and quality and what someone does with the feedback), the marks, and the TSES (Self-efficacy for student engagement, instructional strategies, and classroom management), separate for both sessions.

Yonathan: low AEQ scores

The first outlier is Yonathan (male, 28 years old). He had been a teacher for four years, and was teaching the fifth grade (his students are 10-11 years old). He was a new participant in the

VIP procedure. In his first session, being in the second step of the VIP procedure (i.e., setting goals and actions), Yonathan had an Observation phase (5 minutes; 30 utterances) and an Analysis phase (9 minutes; 68 utterances). During the second session (the fourth step of the VIP procedure, i.e., evaluating), a Reflection phase (5 minutes; 17 utterances) and an Analysis phase (20 minutes; 85 utterances) emerged. While Yonathan participated in the VIP procedure, he was mentoring an intern¹. The intern was teaching most of the lessons, and therefore Yonathan's goal was directed at his mentoring skills: how to help my intern with her classroom management skills.

The four feedback dimensions were scored separately for each phase in both sessions. From our literature perspective, the dimensions of goal-directedness, specificity, and details should be as high as possible in order to be effective. The dimension of positivity should be average in order to be effective. In Yonathan's turns, the dimension goal-directedness was once lower than average; once average; and twice higher than average. The dimension specificity was twice lower; once average; and once higher. A remarkable difference between the sessions is that this dimension is lower during the first and higher during the second session. In other words, during the first session, the feedback was more general, while during the second session, the feedback was more specific. The dimension of details was once lower; and three times higher. The dimension positivity was twice lower; and twice average. Even though this seems balanced, the variation in the dimensions was larger on the low scores and smaller on the high scores (see first section of upper part of the table). This indicates that the feedback dimensions during Yonathan's turns are not favorable.

Most feedback elements are given during the Analysis phases (see second section of the upper part of the table). This is not surprising, because the Analysis phases took the most time. The amount of feedback elements during the Observation and Reflection phases are equal. In all phases, much more effective feedback elements are provided than ineffective feedback elements. This implies that Yonathan received effective feedback.

However, the ratio of the real effectiveness of these elements is not advantageous (see third section of the upper part of the table). About 33% of the expected effective feedback elements were effective, and 67% were not effective. Five out of six solution-focused questions Yonathan received, that did not elicit the feedback dimensions to be more effective are a remarkable example of this ratio; Table 5 shows that solution-focused questions usually

¹ In the Netherlands, students from teacher education institutions follow internships at several schools. They are obliged to teach lessons themselves, during which the regular teacher observes the student.

Figure 2a. An example of a solution-focused question in Yonathan's turn, that did not affect the feedback dimensions.

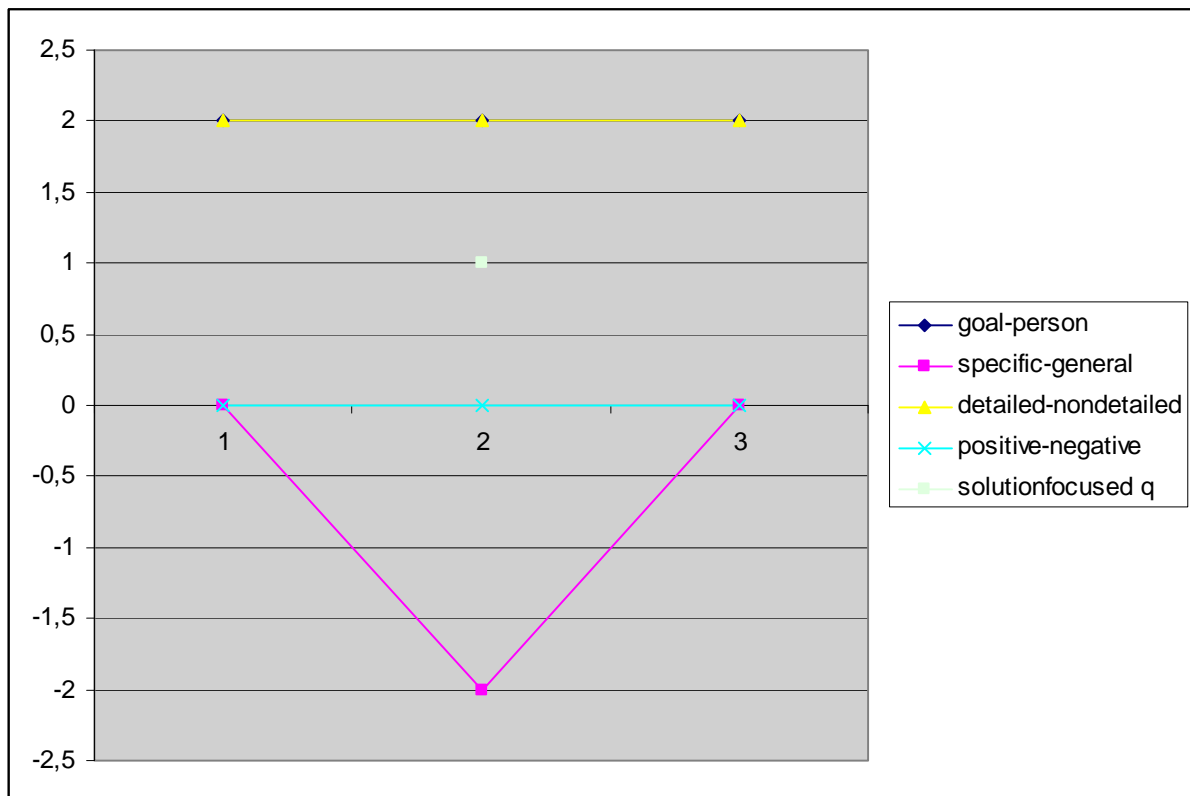
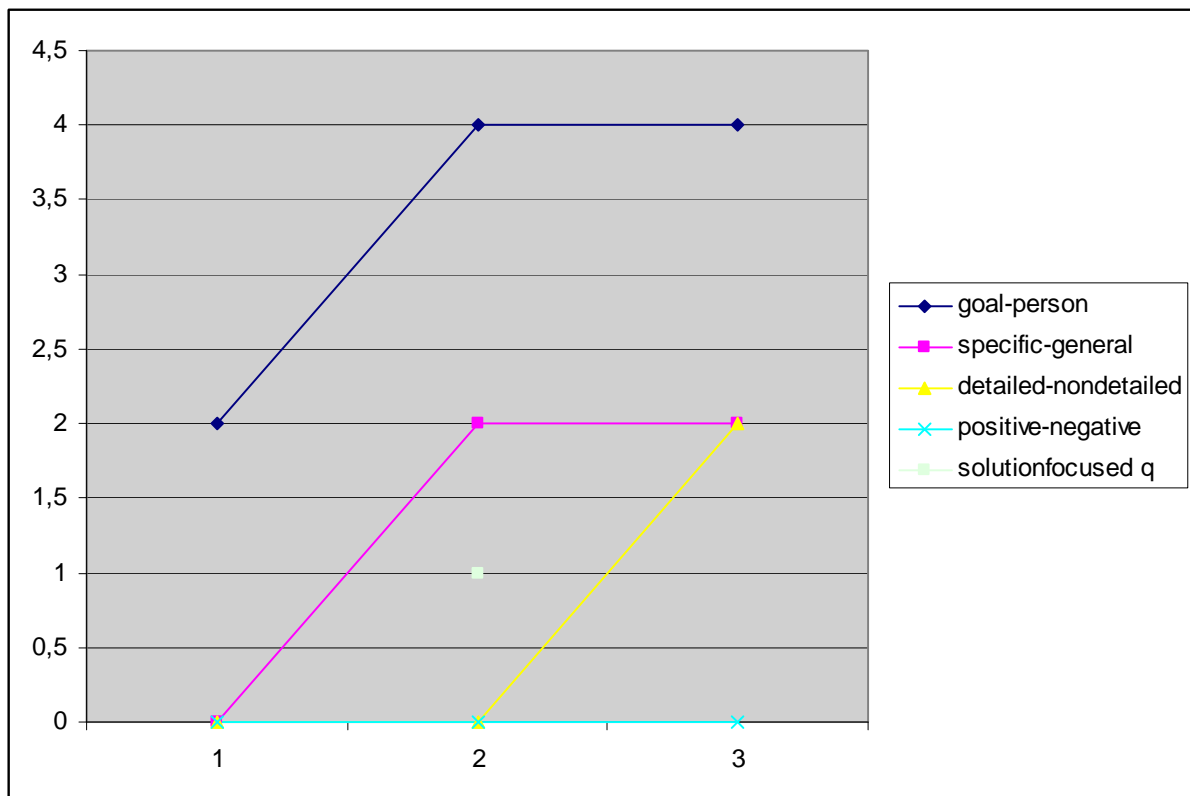


Figure 2b. An example of a solution-focused question that did affect the feedback dimensions, as emerged in Olivia's turns.



have a positive effect on the feedback dimensions (see also Figures 2a and 2b). Figure 2a clearly shows that the feedback dimensions are not affected, whereas in Figure 2b, all feedback dimensions are positively affected. The proportion between ineffective feedback elements that are indeed ineffective and are unexpectedly effective is about fifty-fifty. This indicates that the feedback Yonathan received did not stimulate him in being more goal-directed, more specific, and more detailed.

The AEQ scales provide insights into how Yonathan perceived feedback. He perceived the quantity of feedback the best in both sessions ($M = 4.00$ and 4.13). This collaborates with the amount of feedback elements he received. Yonathan perceived the quality of feedback not as high ($M = 3.40$ and 3.00): his average was just above the middle of the scale (i.e., 2.5). This seems to collaborate with the other findings of the observation. Yonathan was intended to do something with the feedback he received, yet this score was higher during the first session ($M = 3.63$) than in the second session ($M = 3.50$). This could be explained by the workflow of the VIP procedure. That is, during the first session, Yonathan was in the second step of the VIP procedure, and during the second session, he was the fourth step.

Yonathan's marks averaged 7.4 . In both sessions, he judged the atmosphere (8), learning when being the coached teacher (8), and living up to his expectations (8) higher, whereas he was less positive about getting new insights (7) and new learning goals (6).

Yonathan's self-efficacy did not change between the sessions. He indicated his self-efficacy on the three subscales of the TSES quite high. His self-efficacy on classroom management was the highest ($M = 7.75$).

Quinta: low AEQ scores

Quinta is the second outlier (female, 36 years old). She had been a teacher for three years and was teaching the third grade (her students are 8-9 years old). She was a new participant to the VIP procedure; however, she had experience using video. In her first session, Quinta went through an Analysis phase (1 minute, 9 utterances), an Observation phase (1 minute, 2 utterances), and a second Analysis stage (14 minutes, 87 utterances). In the second session, an Observation phase (9 minutes, 25 utterances) and an Analysis phase (13 minutes, 105 utterances) occurred. In the VIP procedure, Quinta was working on a goal which was focused at having a positive classroom climate and how she could approach specific students, which she felt she did not approach in a positive way.

In Quinta's turns, the dimension of goal-directedness was three times lower; and two times higher than average. The dimension of specificity was twice lower; and three times higher. The dimension of details was once lower; once average; and three times higher. The dimension of positivity was always average. Within the first session, all feedback dimensions were higher than average in the first phases (the short Analysis phase and the Observation phase), however, they were all lower than average in the second Analysis phase ($1/2$ to 2 *sd*). The variation was greater in the dimensions that were higher than average than in the dimensions that were lower than average. This indicates that the feedback dimensions in Quinta's turns were not pronounced effective or ineffective.

During the Observation phases, Quinta only received effective feedback elements (15). During the Analysis phases, she received about twice as much effective feedback elements than ineffective. About 42% of the expected effective elements were indeed effective, and about 58% were ineffective. The ratio of the expected ineffective feedback elements was about fifty-fifty. This indicates that the feedback during Quinta's turns was rather effective.

The AEQ scales provide insights into how Quinta perceived the feedback. She perceived the quantity of feedback lower in the first session ($M = 3.63$) than in the second session ($M = 4.25$). This is in line with the results, which showed that she received more feedback elements in the second session (78 elements) than in the first (50 elements). Furthermore, Quinta perceived the quality of feedback higher during the first session than in the second session. Correspondingly, she received 39 effective feedback elements and 11 ineffective feedback elements in the first session; and in the second session she got 53 effective and 25 ineffective feedback elements. Finally, Quinta was more inclined to do something with the feedback she received in the first session than in the second. This might be explained by the fact that the second session was scheduled a few weeks before the summer holidays.

Quinta's marks averaged 9.0 in the first session, and 8.83 in the second session. After the first session, she judged all items with the score 9, and after the second session she judged all items again with the score 9, except for getting new insights (8).

Finally, Quinta's self-efficacy did not change between the sessions. On all scales, she indicated her self-efficacy moderately ($M = 6.50$).

Tiffany: high AEQ scores

Tiffany is the third outlier (female, 29 years old). She had been a teacher for four years and was teaching in Kindergarten (her students are 4-6 years old). She was a new participant in

the VIP procedure. During her first session being in the second step of the VIP procedure, an Observation phase (2 minutes, no utterances) and an Analysis phase (26 minutes, 224 utterances) emerged. In the second session, Tiffany went through the fourth step of the VIP procedure. In this session, an Observation phase (4 minutes, 15 utterances), a Reflection phase (2 minutes, 9 utterances), another Observation phase (2 minutes, 3 utterances), and another Reflection phase (16 minutes, 109 utterances) emerged. During both sessions, Tiffany's goal was focused at making a queue of students for walking to the playground. Tiffany wanted to be clearer to the students on what they should do, and as a consequence, the queue would be made quicker and more efficient.

During Tiffany's turns, the dimension of goal-directedness was twice lower than average; once average; and twice higher than average. The dimension of specificity was twice lower; and three times higher than average. The dimension of details was three times lower; and two times higher. The dimension of positivity was once lower; once average; and three times higher than average. The variation was smaller on the dimensions that were lower than average and greater on the dimensions that were higher than average. Only two dimensions were average. This indicates that the feedback dimensions during Tiffany's turns were slightly more effective than ineffective.

In the Observation phases, Tiffany only received effective feedback elements (11). In both the Analysis and Reflection phases, the amount of effective feedback elements was much greater than the amount of ineffective feedback. This indicates that Tiffany received a lot of effective feedback. About 42% of the expected effective feedback elements were indeed effective, and 58% was not effective. The ratio between the expected ineffective feedback elements being ineffective or effective was about fifty-fifty. This indicates that the feedback during Tiffany's turns was rather effective.

The AEQ scales provide insights into how Tiffany perceived the feedback. After both sessions, she perceived the quantity of feedback utmost high ($M = 5.00$). However, she perceived the quality of feedback higher after the second session ($M = 5.00$) than in the first session ($M = 4.60$). However, during the first session, Tiffany received 130 effective feedback elements and 16 ineffective; in the second session, she received 76 effective feedback elements and 15 ineffective. Controversially, the feedback dimensions were more often higher than average in the second session than in the first. Finally, Tiffany was more intended to do something with the feedback after the second session ($M = 4.63$) than after the first session ($M = 4.13$). This might be explained by the fact that during the second session another idea emerged that might have helped Tiffany to achieve her goal even better.

Tiffany's marks averaged 8.0 after the first and 8.17 after the second session. She judged all items with the score 8 after the first session, after the second session, she judged all items with the score 8, except for the item learning as being the peer coach (9).

Finally, Tiffany's self-efficacy is utmost high, even though her self-efficacy is slightly lower after the second session. After the first session, she averaged 9 on student engagement and classroom management, and 8.25 on instructional strategies. After the second session, her self-efficacy scored highest on student engagement, than on classroom management, and again lowest on instructional strategies.

Wanda: high AEQ scores

Wanda is the final outlier (female, 38 years old). She had been a teacher for 14 years and was teaching the first grade (students are about 6-7 years old). She was a new participant in the VIP procedure. During her first session (second step of the VIP procedure), Wanda had an Observation phase (8 minutes, 13 utterances) and an Analysis phase (16 minutes, 81 utterances). During her second session (fourth step of the VIP procedure), Wanda went through an Observation phase (11 minutes, 8 utterances) and a Reflection phase (13 minutes, 50 utterances). During the VIP procedure, Wanda was working on a goal concerning her instructional strategies during writing lessons. She wanted to be more efficient and wanted her students to learn to write specific letters faster than before.

During Wanda's turns, the dimension goal-directedness was once lower than average; once average; and twice higher than average. The dimension of specificity was twice average; and twice higher. The dimension of details was twice average; and twice higher. The dimension of positivity was twice lower; and twice average. Moreover, the variation on the dimensions that were lower than average was smaller than the variation on the dimensions that were higher than average. This indicates that the feedback dimensions in Wanda's turns were advantageous.

During the Observation phases, Wanda only received effective feedback elements (7). During both the Analysis phase and the Reflection phase, Wanda received much more effective feedback elements than ineffective feedback elements. About 41% of the expected effective feedback elements were indeed effective and about 59% were ineffective. The ratio between ineffective feedback elements being ineffective and effective was about fifty-fifty. This indicates that the feedback during Wanda's turns was rather effective. Some remarkable points in Wanda's turns are that closed questions never affected the feedback dimensions (see

Figure 3. An example of closed questions in Wanda's turns that did not affect the feedback dimensions.

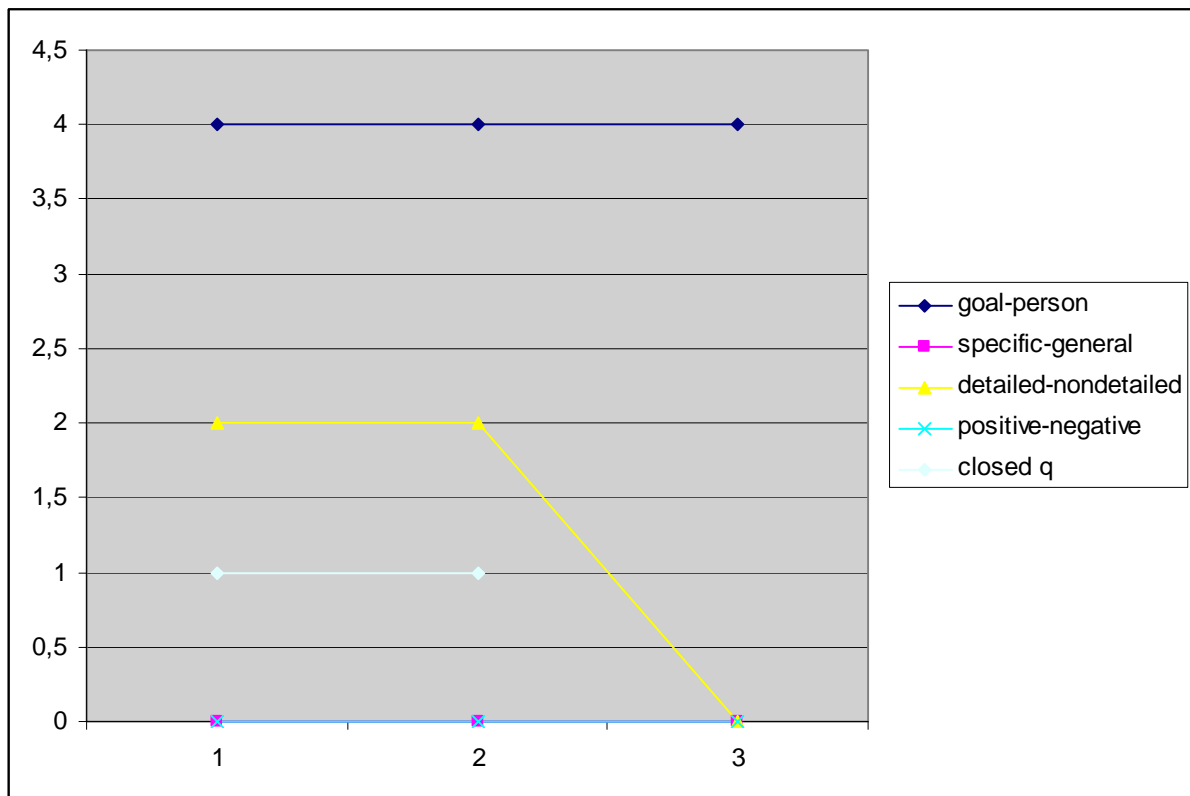


Figure 4. An example of providing an example in Wanda's turns that did affect the feedback dimensions.

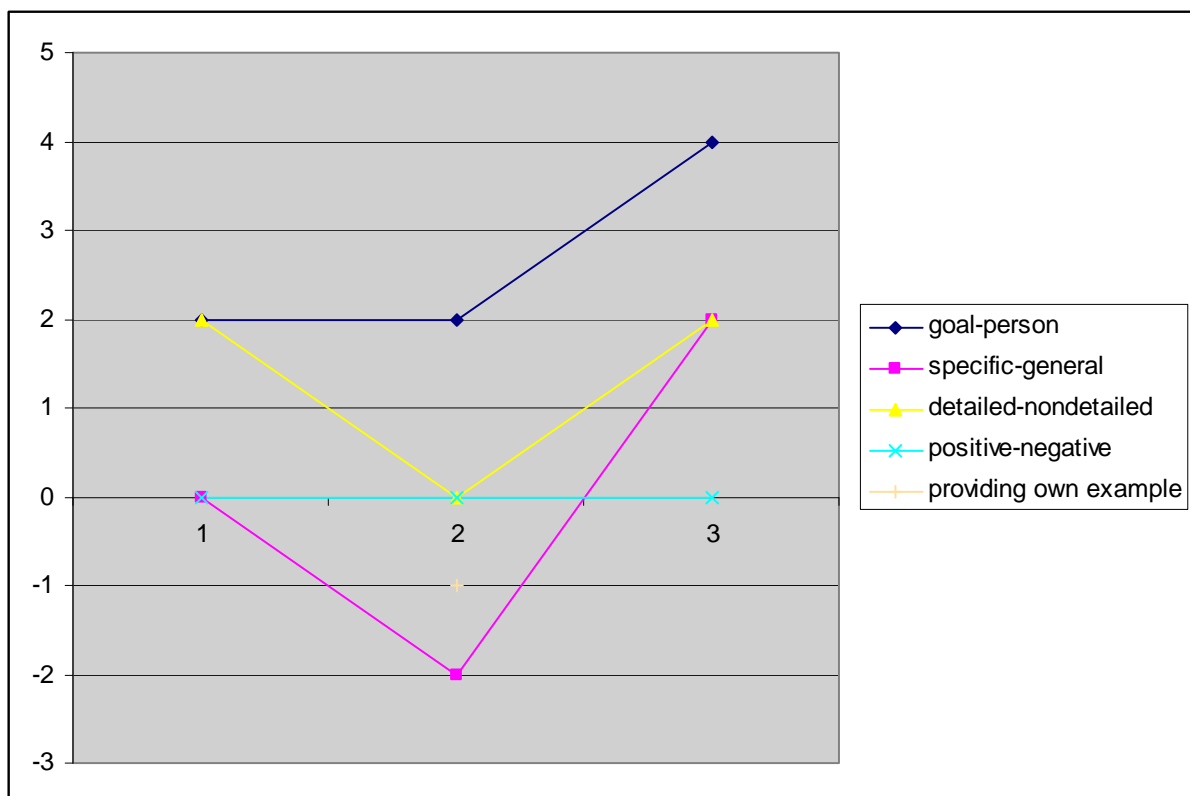


Figure 3) and that providing an example affected the feedback dimensions (see Figure 4), while all other ineffective feedback elements were not effective.

The AEQ scales provide insights into how Wanda perceived feedback. However, there are some missing data in her second questionnaire on the quantity-scale and what-to-do-scale. In the first session, Wanda perceived the quantity of feedback fairly high ($M = 4.75$), as well as the quality ($M = 4.80$). Wanda was intended to do something with the feedback ($M = 4.50$). After the second session, Wanda again perceived the quality of feedback high ($M = 4.80$).

Wanda's marks averaged 8.0. After both sessions, she judged learning as being the coached teacher (8), learning as being the peer coach (8), and living up to her expectations (7) equally. After the first session, she judged the atmosphere with a 9, and after the second session with a perfect 10. After the first session, she judged getting new insights with an 8, and after the second session with a 7.

Finally, Wanda's self-efficacy was fairly high and changed on two subscales. Her self-efficacy on student engagement rose from 7.25 after the first session to 7.50 after the second session. Her self-efficacy on classroom management rose from 7.00 after the first session to 7.25 after the second session. Her self-efficacy on instructional strategies remained 7.50 after both sessions.

Conclusions

This paper focuses at observed and perceived feedback among teachers. The research triangulates data from observations, questionnaires, and interviews with 12 primary school teachers. The teachers participated in a method for peer coaching, the Video Intervention Peer coaching procedure (Jeninga, 2003). The research questions were formulated as follows 1) Which and to what extent do effective and ineffective feedback dimensions and feedback elements occur in the peer coaching activities, and how do they relate to one another? 2) How did the teachers perceive feedback? 3) To what extent does the observed feedback relate to the perceived feedback?

In order to address the first research question, the Teacher Feedback Observation Scheme (TFOS, Thurlings, et al., in press) was used. In the observations, the three phases of White's process feedback model (2009) were used. First, the results on the four feedback dimensions (i.e., goal-directedness vs. person-directedness; specific vs. general; detailed vs. non-detailed; and positive vs. neutral vs. negative) show that the dimensions of details and positivity differ among the phases. Both dimensions are lowest in the Observation phase, than in the Analysis phase, and highest in the Reflection phase. Second, the results show that in all

cases more expected effective feedback elements were provided than expected ineffective feedback elements. Third, the real effectiveness of the feedback elements was determined by comparing the position of the feedback elements before and after the feedback element occurred. Most expected effective feedback elements were indeed effective, with closed questions, summarizing and acknowledging as exceptions. It may be better for teachers to try to formulate questions in an open-ended way instead of closed, because open-ended questions usually are effective. Summarizing is probably not effective, because it wraps things up, before turning to a new issue. Those summaries that were effective were accompanied by a question, which elicited the coached teacher to elaborate more, and probably thereby affecting the feedback dimensions. Acknowledging in itself may not be effective, but might be necessary in terms of relatedness (Ryan & Deci, 2000). If a coached teacher feels that his coaches acknowledge him in his goals, the coached teacher might be more receptive for questions, which help him to tackle the problem. Most expected ineffective feedback elements were indeed ineffective.

In conclusion to the first research question, one can argue that the literature study, which mainly contained articles on teacher to student feedback, is confirmed for teacher to teacher feedback. This indicates that effective feedback is similar for any kind of learner. Furthermore, the expected effective feedback elements are overall effective and expected ineffective feedback elements are overall ineffective. Moreover, the VIP procedure is confirmed as an effective professional development activity: the principles of the VIP procedure of watching video excerpts, asking open-ended, solution-focused questions, acknowledging the coached teachers and helping them to tackle their goals are all confirmed in the observations.

A shortcoming of the observations is that the content of feedback was not taken into account. Future research could address this issue by combining feedback dimensions and feedback elements with the content of the utterances. Another shortcoming is the small population that consisted only of primary school teachers, and therefore, the results and conclusions cannot be readily generalized to a larger population.

In order to address the second research question, the teachers filled in a questionnaire after each session that was based on the Assessment Experience Questionnaire (AEQ, Gibbs & Simpson, 2003) and they were interviewed. In general, teachers perceived the feedback quantity and quality in a positive way. The marks that expressed their satisfaction reflected this. The teachers were most satisfied with the atmosphere in both sessions. Getting new insights and new learning goals was judged lower after the second session, which might be

explained by the fact that the final sessions were planned a few weeks before the summer holidays. A non-parametric test (Friedman test) showed that the scores on the AEQ scales and the marks relate to one another. The interviews suggested that teachers were positive about the VIP procedure, as well as the feedback. They had different views on what effective feedback is, but had similar views on what ineffective feedback is. Ineffective feedback is a hint that is unusable or feedback that is too confronting, whereas effective feedback was seen as useful hints, perspectives from colleagues, compliments, questions, and a reflection or reaction.

A shortcoming regarding this research question is that between the first and second sessions, only small differences in the AEQ scales were found. It might be better to replace the five-point scale with for instance a seven-point scale. Furthermore, it is unclear how the AEQ scales relate to the marks. The Friedman test does not provide this information. If the study would be repeated with more participants, such that correlations can be calculated, more insights might be provided. Third, the questionnaires were filled in on the basis of the whole session, whereas the observation data was divided into the three phases. Therefore, the results of the observations and the questionnaires cannot be linked by means of statistics. The questionnaire or interview might be adapted such that the participants have to answer questions on the part of the session when they were watching the video excerpt (i.e., the Observation phase); when they were the coached teacher to formulate their goal and actions (i.e., the Analysis phase); and when they were looking back at how they tried out the actions (i.e., the Reflection phase).

In order to address the third research question, four teachers were examined into more detail. Two teachers were selected based on their lower than average scores on the AEQ scales (i.e., Yonathan and Quinta); and two teachers were selected based on their higher than average scores (i.e., Tiffany and Wanda). The amount of *feedback dimensions* that was lower than average, average, or higher than average did not differ between Yonathan and Quinta. Tiffany's dimensions scored average twice, which implies that her feedback dimensions were either effective or not. Wanda only had three lower than average dimensions; the amount of dimensions that were average or higher than average was equal. Furthermore, the variation on the lower than average dimensions differed: the variation on these dimensions was greater in Yonathan's and Quinta's turns than in Tiffany's and Wanda's. In other words, if the feedback dimensions were lower in Tiffany's and Wanda's turns, they were only in a small amount. In contrast, the variation on the higher than average dimensions showed a different pattern: in Yonathan and Wanda's turns, the variation was small (about 1 *sd*); in Tiffany's turns, the

variation was larger (about 2 *sd*); and in Quinta's turns, the variation was the greatest (about 3 *sd*).

All teachers only received effective *feedback elements* during the Observation phase. In all cases, the amount of effective feedback elements was larger than the amount of ineffective feedback elements. An interesting difference lies in the Reflection phase: Quinta did not even go through one, whereas the Reflection phase in Yonathan's sessions was short in itself. Tiffany and Wanda had longer Reflection phases.

The ratio of the real *effectiveness* of the ineffective *feedback elements* was about fifty-fifty for these four teachers. The ratio of the real effectiveness of the expected effective elements was about the same for Quinta, Tiffany, and Wanda: 41-2% of expected effective elements were indeed effective, and 58-9% were ineffective. The ratio on the expected effective feedback elements was least advantageous for Yonathan: 33% of the expected effective elements were indeed effective, and 67% was ineffective.

Combining the results from the observations and questionnaires of these teachers, one could conclude that the observed effectiveness of feedback relates to the perceived effectiveness of feedback. Yonathan, who received inadequate feedback, also perceived the feedback to be inadequate. Quinta, who received average feedback, also perceived the feedback to be average. Tiffany and Wanda, who received adequate feedback, also perceived the feedback to be adequate. Even though they were not aware of it, they judged the quantity of feedback higher, when they received more feedback; even though they were not aware of it, they judged the quality of feedback higher, when they received feedback that was indeed effective. Future research could investigate how these implicit thoughts on the quality and quantity of feedback are linked to the observed quality and quantity of feedback.

Scientific significance

Feedback is considered relevant for any kind of learning, whether the learner is a student or a teacher (Hattie & Timperley, 2007). Feedback among teachers can be regarded as crucial for their professional development, because it stimulates and guides teachers' learning and development. However, little is known about feedback among teachers (Scheeler, et al., 2004). This paper focuses on feedback among teachers by observing feedback, investigating how teachers perceived feedback in terms of effectiveness, and by combining this, the study resulted in the conclusion that effective observed feedback was also perceived as effective; and that ineffective observed feedback was also perceived as ineffective. In addition, the paper demonstrated that the beliefs about effective feedback for students also hold true for

teachers. Moreover, the paper contributes to the scientific and practical knowledge on feedback among teachers, which can guide the development of teacher professional development programs.

References

- Baarda, D. B., de Goede, M. P. M., & van Dijkum, C. J. (2007). *Basisboek statistiek met SPSS [Basic statistics with SPSS]*. Groningen, the Netherlands: Noordhoff Uitgevers.
- Black, P., & William, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-68.
- Black, P., & William, D. (1998b). Inside the black box *Phi Delta Kappan*, 80, 139-147.
- Brophy, J. (2004). *Advances in research on teaching. Volume 10. Using video in teacher education*. Amsterdam: Elsevier.
- Gibbs, G., & Simpson, C. (2003). *Measuring the response of students to assessment: The Assessment Experience Questionnaire*. Paper presented at the the International Improving Student Learning Symposium, Hinckley, UK.
- 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.
- Jackson, P., & McKergow, M. (2002). *Oplossingsgericht denken [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. In J. Fanchamps & J. v. d. Sanden (Eds.), *Integraal ondersteunen van een vernieuwd VMBO* (pp. 25-32). Antwerpen/ Apeldoorn: Garant.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis. Second edition*. Thousand Oakes: Sage.
- Moore, D. S., & McCabe, G. P. (2003). *Introduction to the practice of statistics. Fourth edition*. New York: W. H. Freeman and Company.
- 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: MacMillan Library Reference.
- 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.
- Scheeler, M. C., Ruhl, K. L., & McAfee, M. K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education*, 27(4), 59-70.
- 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.
- Showers, B. (1985). Teachers coaching teachers. *Educational Leadership*, 42(7), 43-48.
- Showers, B., & Joyce, B. (1996). The evolution of peer coaching. *Educational Leadership*, 54(3), 12-16.
- Thurlings, M., Vermeulen, M., Kreijns, C., Bastiaens, T. J., & Stijnen, P. (in press). Development of the Teacher Feedback Observation Scheme: Evaluating the quality of feedback in peer groups. *Journal of Education for Teaching*.
- Tillema, H. H., & Smith, K. (2000). Learning from portfolios: Differential use of feedback in portfolio construction. *Studies in Educational Evaluation*, 26, 193-210.
- Tschannen-Moran, M., & Woolfolk-Hoy, A. (2001). Teacher efficacy: An elusive construct. *Teaching and Teacher Education*, 17, 783-805.
- Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment and Evaluation in Higher Education*, 31(3), 379-394.
- 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.