Adv in Health Sci Educ (2015) 20:803–816
DOI 10.1007/s10459-014-9567-8

Framing of feedback impacts student's satisfaction, self-efficacy and performance

J. M. Monica van de Ridder · Claudia M. M. Peters · Karel M. Stokking · J. Alexander de Ru · Olle Th. J. ten Cate

Received: 9 May 2013 / Accepted: 12 November 2014 / Published online: 27 December 2014

Abstract Feedback is considered important to acquire clinical skills. Research evidence shows that feedback does not always improve learning and its effects may be small. In many studies, a variety of variables involved in feedback provision may mask either one of their effects. E.g., there is reason to believe that the way oral feedback is framed may affect its effect if other variables are held constant. In a randomised controlled trial we investigated the effect of positively and negatively framed feedback messages on satisfaction, self-efficacy, and performance. A single blind randomised controlled between-subject design was used, with framing of the feedback message (positively–negatively) as independent variable and examination of hearing abilities as the task. First year medical students' (n = 59) satisfaction, self-efficacy, and performance were the dependent variables and were measured both directly after the intervention and after a 2 weeks delay. Students in the positively framed feedback condition were significantly more satisfied and showed significantly higher self-efficacy measured directly after the performance. Effect sizes found were large, i.e., partial $\eta^2 = 0.43$ and $\eta^2 = 0.32$ respectively. They showed a better performance throughout the whole study. Significant performance differences were found both at the initial performance and when measured 2 weeks after the intervention: effects

J. M. M. van de Ridder · O. T. J. ten Cate
Center for Research and Development of Education, University Medical Center Utrecht, Utrecht, The Netherlands

J. M. M. van de Ridder (✉)
Department of Education, Albert Schweitzer Hospital, P.O.Box 444, 3300 AK Dordrecht, The Netherlands
e-mail: m.van.de.ridder@asz.nl

C. M. M. Peters
Council of Water Education, Nieuwegein, The Netherlands

K. M. Stokking
Department of Educational Sciences, Utrecht University, Utrecht, The Netherlands

J. A. de Ru
Department of Otolaryngology, Central Military Hospital Utrecht, Utrecht, The Netherlands

were of medium size, respectively $r = -.31$ and $r = -.32$. Over time in both conditions performance and self-efficacy decreased. Framing the feedback message in either a positive or negative manner affects students' satisfaction and self-efficacy directly after the intervention be it that these effects seem to fade out over time. Performance may be enhanced by positive framing, but additional studies need to confirm this. We recommend using a positive frame when giving feedback on clinical skills.

Keywords Assessment · Clinical skills · Feedback · Framing · Performance · Satisfaction · Selfefficacy

Introduction

Feedback is considered to be important in clinical training (Bowen and Irby 2002; Dornan 2006; Irby 1995; Liberman et al. 2005; McIlwrick et al. 2006). In clinical education feedback has been defined as 'Specific information about the comparison between trainee's performance and a standard given with the intent to improve trainee's performance' (van de Ridder et al. 2008).

Students view feedback as supportive when participating in the clinical context (Dornan 2006), and they rate "giving constructive feedback" as the second most important behavioral characteristic of a preceptor (Schultz et al. 2004). Feedback enhances learning and may lead to more responsibility and autonomy (Dornan 2006). Most importantly, effective feedback has the potential to reduce trial-and-error behavior by trainees in patient care.

However, from meta-analyses and reviews we know that feedback does not always improve performance but can also decrease performance (Kluger and DeNisi 1996), and that effect sizes of feedback interventions are small (Ivers et al. 2012; Veloski et al. 2006). Sometimes feedback is perceived as effective, but performance outcomes after the feedback is received do not reflect a change in behavior (Boehler et al. 2006).

A possible explanation for the diverse feedback effects are the different ways in which it is provided (Branch and Paranjape 2002; Daelmans et al. 2005; van de Ridder et al. 2006 unpublished). Valence framing -framing the feedback message positively or negatively- is one variety (Dunegan 1996; Levin et al. 1998). Differences in supervisors' feeling of time constraint (Kilminster and Jolly 2000), feelings towards the role of teacher (Bowen and Irby 2002), differences in training in providing feedback (Salerno et al. 2002), or awareness of the importance of feedback (McIlwrick, et al. 2006) have been suggested to underlie the variety in such framing.

Valence framing has defined as casting 'the same critical information in either a *positive* or a *negative* light' (p.150) (Levin et al. 1998). So, framing refers to the *packaging* of the message and it is independent of the message content, which can consist of positive -about good points- or negative feedback -about points for improvement-, or both. "You did this quite well, but there are some points for improvement…" would be considered a positively framed feedback message. Negatively framed feedback is: "You did not do this correctly. You should…" In both examples the *content of the feedback message* is negative.

Message framing is an important aspect of communication. In studies outside the medical field it has been shown to influence the outcomes of punishment (Dunegan 1996), the process of decisions taking, especially in the presence of risk (Dunegan 1993;

Kahneman and Tversky 1984; Levin et al. 1998; Tanner and Medin 2004), evaluations of objects (Levin et al. 1998; Schul and Ganzach 1995), persons' attitude certainty (Rucker et al. 2008), and persuasiveness in communication (Levin et al. 1998).

We identified two studies specifically focusing on positive and negative framing of a feedback message, but the task and the context in these studies are non-medical (Dunegan 1995; Waung and Jones 2005). Since framing is an important aspect of communication, and it also is applied in giving feedback we would like to measure the impact on a clinical task performance in a medical setting. In this study we evaluated the effect of positively and negatively framed feedback on outcome measures such as students' satisfaction, self-efficacy, and performance. These outcome measures relate to two levels of Kirkpatrick's hierarchy of evaluation outcomes (Kirkpatrick and Kirkpatrick 2006): reaction (satisfaction and self-efficacy) and learning (performance).

Satisfaction

Evaluating students' perception of provided feedback is important. Positive reactions of students (e.g., 'I liked to receive this feedback') can motivate, stimulate their learning and lead to loyalty (Kirkpatrick and Kirkpatrick 2006; Morgan and Rego 2006), even if their perceptions are not accurate (Jussim 1991). Research shows that positive feedback on performance positively influences students' satisfaction (Nesbit and Burton 2006), receiving process feedback in a virtual team leads to an increase of satisfaction by team members (Geister et al. 2006) and a comparison of students' satisfaction after receiving praise or feedback shows higher satisfaction for receiving praise (Boehler et al. 2006). We expect students' satisfaction to be higher in a 'positively framed feedback' than in a 'negatively framed feedback' condition.

Self-efficacy

Bandura (1986) defines self-efficacy as: 'people's judgments of their capabilities to organize and execute courses or action required to attain designated types of performances' (p.391). It can be classified as self-perception. This self-perception is based on enactive attainment, vicarious experiences, verbal persuasion, and one's physiological state. Self-efficacy influences for example choice behavior, effort expenditure and persistence, thought patterns, and emotional reactions. Self-beliefs contribute in various ways to psychosocial functioning (Bandura 1986).

Self-enhancement theory suggested that everyone wants to improve their own functioning and therefore only positive feedback -about good points- is effective (Swann et al. 1987). Results in line with this theory show that particularly negative feedback reduces self-efficacy and positive feedback increases self-efficacy (Reynolds 2006). Negative feedback has more influence on subjects with low self-efficacy compared to subjects showing high self-efficacy (Baker 2001).

Self-verification theory states that when the feedback is in line with a person's self-concept they will endorse the feedback as valid, even when it is negative (Swann et al. 1987). Empirical research in line with self-verification theory shows that negative feedback is better accepted by people with low self-efficacy about a specific competency compared to subjects with high self-efficacy (Nease et al. 1999).

Performance

Based on research evidence it is hard to formulate how feedback framing will influence performance. Generally feedback is considered to have a small or moderate impact on performance (Ivers et al. 2012) effect sizes are often small (Kluger and DeNisi 1996), or show an enormous variation (Hattie and Timperley 2007). When feedback is given systematically, by a credible source, combined with other interventions such as education or guidelines, it appears to be more effective (Ivers et al. 2012; Veloski et al. 2006).

Most studies on the effect of feedback only report direct effects, not many collect data after a time delay. To detect the sustained influence of feedback over time we decided to repeat measures of performance and self-efficacy after 2 weeks.

The following research question was formulated: What is the effect of feedback framing on students' satisfaction, self-efficacy, and performance?

Methods

Design

A single-blind randomised controlled between-subject design was used, with feedback framing (positive–negative) as the independent variable. Dependent variables used were students' satisfaction, self-efficacy, and performance outcomes.

Task

The task given to students was to practice the Weber and Rinne (WR) tuning fork procedures to test hearing deficiencies on a standardized patient (SP) according to guidelines provided by an otorhinolaryngologist on an instructional video. This test is used in the clinical setting in both the otolaryngology and neurology (Boatman et al. 2007; Bagai et al. 2006). The task is complex, in the sense that it consists of several elements: explaining the task to the patient, placing the tuning fork at patient's forehead and mastoid bone and asking the right questions at the right time. The students need to combine the retrieved information and diagnose whether the patient has a hearing impairment and if so, if it is a sensorineural or a conductive impairment (Bagai et al. 2006). The student concludes the task with informing the patient about the findings. For a detailed description we refer to a short video-instruction (Ear, Nose and Throat Examination—Medi-Vision Films 12 2010).

This task is suitable for our experiment because the task is not time consuming to perform, it is observable from a video recording, the pathology is easy to simulate by a SP, and it has obvious relevance for medical training. The curricular approach for all first year medical students had been the same prior to our study. The first year students were not acquainted with the WR task, as it is only taught in the second year of this medical curriculum.

Framing

In both conditions students received a feedback message with negative content, provided by a final year medical student with stage-play experience, acting as an experienced physician familiar with the WR task. In the positive framing condition, feedback was

voiced as: "You did this well; some tips are…" Feedback framed negatively started with: "This is not well done; you should change…"

To ensure that the interaction was natural, the feedback provider was allowed to engage in an ordinary dialogue with the student. To keep the feedback message focused, the feedback provider choose, after observing students' performance, from a list of four pre-selected feedback points one item which feedback point was most suitable for this student. The four pre-selected points consisted of students' most frequent mistakes in performing the WR task as determined in a pilot study (Kruisheer et al. 2006). Besides the pre-selected points, no other feedback was given to optimally keep the conditions under control.

The feedback provider was trained in using the experimental protocol, selecting the feedback points, framing the feedback in a similar manner for both conditions, and in engaging in a natural dialogue that would not negatively affect the conditions in the study.

Participants

All 210 first year medical students from the University Medical Center Utrecht (UMCU) potentially available in the study period were approached orally and by email 3 weeks before the study and asked to participate. The medical student population in the Netherlands consists of a stable 65–70 % woman since early 1990s (Ten Cate 2007). By quota sampling we established a group, representative for gender of the medical student population. It was explained that "volunteers were sought for research on clinical skills teaching" and that data would treated confidentially. Participants were asked to sign an informed consent form. Students were informed that they would receive a 7.50 Euros reward after completing the experiment and the follow-up.

Procedure

The experiment was carried out in June 2007, in the skills lab of the UMCU. Other instructional events were unlikely to interfere with the study, as at this time of the year no classes take place and students are preparing for exams.

To collect the outcome measures for this study, each individual went through the route according to Fig. 1. By drawing lots, students were randomized by one investigator (CP) over the two conditions (at time T0). Males and females were almost equally divided over the two conditions. The proportion in the sub groups were representative for the medical student population in the Netherlands (Fig. 1). Each student was asked to rate his/her self-efficacy in performing the WR task (SE1) at time T1. Next, the student watched an instructional video about the WR task procedure. At time T2 each student carried out the WR task [performance 1 (P1)] on a SP while being observed by the feedback provider. The SP was instructed to silently signal to the feedback provider when a procedure was not correctly done, e.g., when a tuning fork was not firmly placed upon the mastoid bone or the forehead. Subsequently, the student would rate his or her self-efficacy again (SE2). Directly after this, the student would receive negatively or positively framed feedback from the supervisor. Then, the student was to perform the tests for the second time on the same SP (P2), and would rate their self-efficacy for the third time (SE3) and fill out the satisfaction (SAT) scale (T3). Two weeks later a follow-up took place (T4). The student would rate self-efficacy for the fourth time (SE4), and perform the WR task for the third time on the same SP (P3) All performances were video taped.

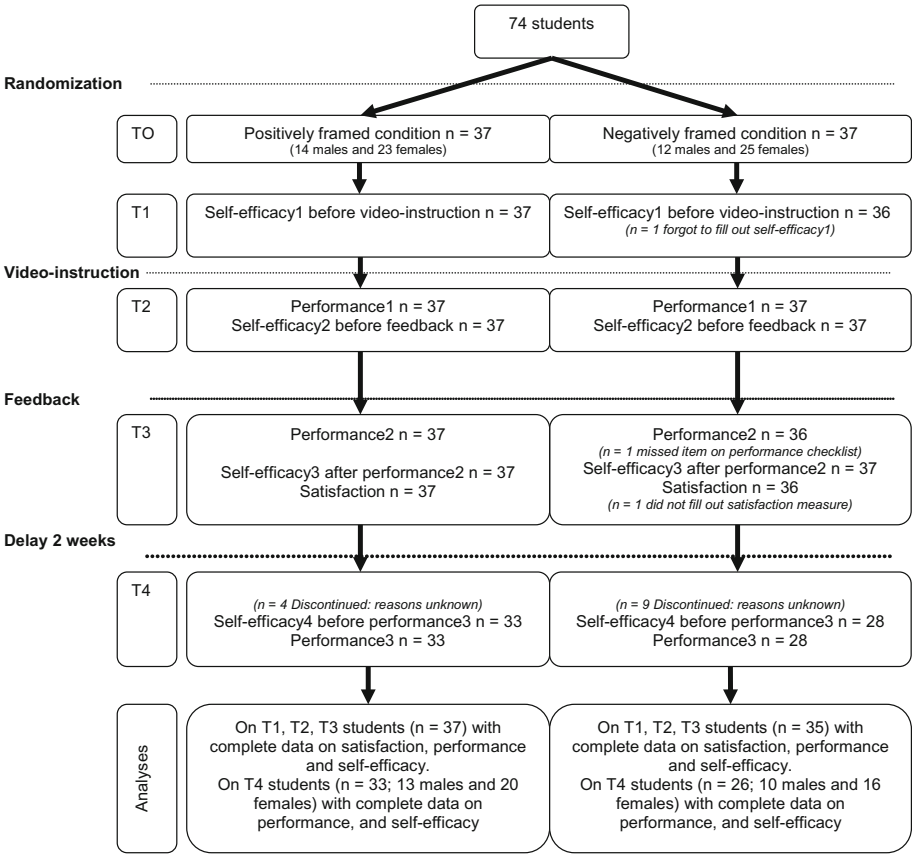

Fig. 1 CONSORT flow-chart (Moher et al. 2001) of the chronological procedure of the trial and the amount of participants who participated in each stage and were included in the analyses

Instruments

Students' satisfaction with the video instruction, training opportunities, and the received feedback is measured with a five-item scale.

A visual analogue scale (VAS) was used to measure self-efficacy regarding the detection of hearing loss by using the WR task (0 mm = 'I am extremely confident', 100 mm = 'I am extremely unconfident'). This procedure for measuring students' self-efficacy towards medical skills has been used in a study by Turner (Turner et al. 2008). This study provides evidence for the validity of the instrument for a similar purpose.

Video recordings of the students' performances for each element of the WR task were scored by two independent raters (CP and MvdR) using a 13-item dichotomous observational checklist (0 = not performed; 1 = performed).

Data analysis

Correlations were determined to get insight in the relationship between the variables satisfaction (SAT), performance (P1, P2, P3) and self-efficacy (SE1, SE2, SE3 and SE4). Whenever assumptions of normality, linearity, homogeneity of variances, and/or homogeneity of regressions slopes were violated, a Mann–Whitney U test was performed, otherwise a mixed model analysis or T test was used to compare framing effects on satisfaction (SAT), performance (P1, P2, P3) and self-efficacy (SE1, SE2, SE3 and SE4).

A balance between type I and type II error was sought by setting $\alpha = 0.10$, because of the relatively small group. Applying a Bonferroni correction for multiple statistical testing on the five measurements of the dependent variables resulted in a significance level of (0.1/5 = 0.02) (Stevens 1996).

Effect sizes (ES) were reported using partial η^2 and r. Small, medium, and large effect sizes are respectively 0.01, 0.06, and 0.14 for partial η^2 (for those unfamiliar with partial η^2, this is equal to $d = 0.2$, 0.5. and 0.8) and 0.1, 0.3, and 0.5 for r (Cohen 1988).

All analyses were done with subjects having a complete data set (n = 59). Subjects with incomplete data sets (n = 15) were removed (Fig. 1).

Results

Participants

Seventy-four students participated in the study. Sixty-one (82 %) students completed all WR tasks of the study. In the positive and negative frame condition 4 (10 %) and 9 (24 %) students respectively dropped out at T4 (Fig. 1). We had complete data sets of 59 students (Fig. 1). Their mean age was 19.5 (SD = 2.3) No significant differences in age and sex distribution between conditions were found.

A comparison of gender, age distribution and performance at baseline, pretest and posttest between the drop-outs and students with a complete data set, did not show significant differences.

Quality of the instruments

In this study, the rather complex task to be carried out appeared to have never done perfectly by any of the first year students. In all cases, it appeared theoretically possible to provide positive as well as negative feedback.

The satisfaction scale showed an internal consistency of Cronbach's $\alpha = 0.85$.

The VAS for measuring self-efficacy is a one-item instrument and cannot yield reliability estimates.

The inter-rater agreement between the raters of the observational checklist for scoring the video recordings was taken as a reliability indicator for the performance measurement (Downing 2004). Four items showed low inter-rater agreement (Cohen's $\kappa \leq 0.45$) and were removed. The nine remaining items of the checklist used as performance measurement had high average inter-rater agreement (Cohen's $\kappa = 0.78$). The sum scores on the checklist were used as performance outcome (0 = minimum; 9 = maximum).

Assumptions

SE3 showed a significant Levene's test outcome ($F(1,57) = 13.7$, $p < 0.01$), indicating no homogeneity of regression slopes. The performance scores (P1, P2, and P3) were not normally distributed and therefore Mann–Whitney U tests were performed (Field 2005).

Correlations

Correlations were determined with Spearman's rank correlation coefficient. There was a positive relationship between SE1 and the first performance (P1) $r_s = .26$, $p < 0.05$ and with SE2, $r_s = .34$, $p < 0.01$. SE2 correlates high with SE3 ($r_s = .53$, $p < 0.01$). SE3 has a positive relationship with satisfaction ($r_s = .56$, $p < 0.01$).

The first performance (P1) positively correlated with P2 ($r_s = .34$, $p < 0.01$), P3 ($r_s = .28$, $p < 0.05$), satisfaction ($r_s = .26$, $p < 0.05$) and SE4 ($r_s = .43$, $p < 0.01$).

Effect of feedback framing on satisfaction

An independent samples t test revealed a significant difference in satisfaction between the positively ($M = 3.8$, $SD = 0.7$, $CI = 3.6$–4.0) and negatively ($M = 2.6$, $SD = 0.7$, $CI = 2.4$–2.8) framed feedback condition; t (57) = 6.7, $p < 0.001$, ES partial $\eta^2 = 0.43$ (equal to $d = 1.73$) (Table 1).

Effect of feedback framing on self-efficacy

The self-efficacy scores increased after the first task performance (SE2). They further increased in the positively framed feedback condition and decrease in the negatively framed feedback condition at SE3, and at SE4 in both conditions the score was lower than the initial baseline score. A linear mixed model analysis was employed to evaluate the effect of framing on the improvement concerning self-efficacy. The interaction between time and the framing conditions was tested to examine changes of the effect over time. The differences in self-efficacy scores between the two conditions were not equal over time. The linear mixed model showed an interaction effect between self-efficacy and time [$F(3,57) = 10.7$, $p < 0.01$]. Consequently, we tested the differences between the groups at the specific time points. Directly after the feedback intervention (T3) the group in the positive framing condition had a significantly higher self-efficacy score than the group in the negative framing condition t (57) = 5.11, $p < 0.00$, ES partial $\eta^2 = 0.32$. We did not find significant differences at T1, T2 and T4 (Table 1; Fig. 2).

Effect of feedback framing on performance

Between T2 and T3 performance scores stayed the same, and they decreased at T3. A Mann–Whitney U test revealed a significant difference between P1 and P3. During P1 students in the positively framed feedback condition ($Mdn = 8.00$) performed better than students in the negatively framed condition ($Mdn = 7.00$) ($U = 277$, $Z = -2.40$, $p < 0.02$, $r = -.31$). No significant difference was found during P2 ($U = 308$, $Z = -1.91$, $p < 0.06$. $r = -.25$). However after two weeks, the students in the positively framed feedback condition ($Mdn = 7.00$), performed better than the students in the negatively framed condition ($Mdn = 6.00$) $U = 272$, $Z = -2.43$, $p < 0.02$, $r = -.32$ (Table 1; Fig. 3).

Table 1 Means (M), standard deviations (SD), medians (Mdn) and 90 % credibility intervals (CI) of satisfaction, self-efficacy and performance by feedback framing condition and time

	N	Satisfaction			Self-efficacy[b]			Performance			
		M	SD	90 % CI	M	SD	90 % CI	M	SD	Mdn	90 % CI
Positive frame											
T1 baseline	33				53.3	17.2	48.3–58.4				
T2 pre-test	33				63.3	15.4	58.8–67.8	7.5	1.2	8.0[a]	7.1–7.8
T3 post-test	33	3.8[a]	0.7	3.6–4.0	73.5[a]	11.5	70.1–76.9	7.5	1.3	8.0	7.2–7.9
T4 2 weeks	33				48.3	19.2	42.6–54.0	6.7	1.7	7.0[a]	6.2–7.2
Negative frame											
T1 baseline	26				53.8	22.0	46.4–61.1				
T2 pre-test	26				61.0	17.5	55.1–66.9	6.8	1.2	7.0[a]	6.4–7.2
T3 post-test	26	2.6[a]	0.7	2.4–2.8	50.9[a]	21.8	43.6–58.2	6.9	1.6	7.0	6.3–7.4
T4 2 weeks	26				45.5	23.1	37.7–53.2	5.8	1.5	6.0[a]	5.3–6.3

[a] These values significantly differ in the positively and negatively framed feedback condition

[b] As most readers will associate a high VAS scores with a high self-efficacy, these results are reported after reversal of the values of the VAS as used in our study

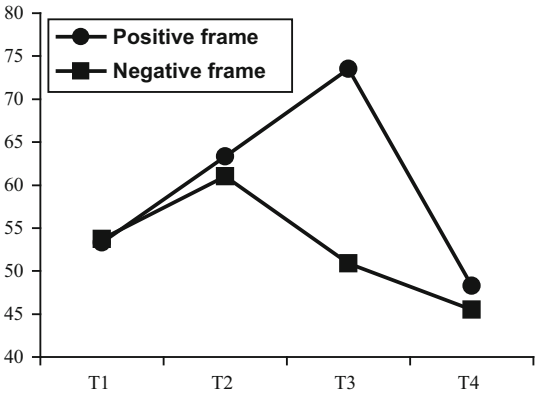

Fig. 2 Self-efficacy (M) by feedback condition over time

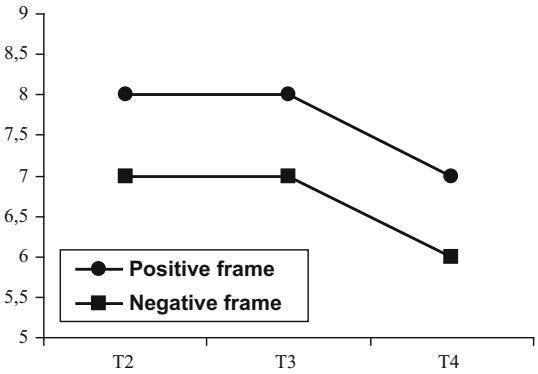

Fig. 3 Performance score (Mdn) by feedback condition over time

In contrast, analyses of covariance using P2 and P3 as dependent measures with P1 performance as a covariate indicated that there were no significant differences between positively and negatively framed conditions [P2: $F(1, 56) = 1.06$, $p = 0.31$; P3: $F(1, 56) = 2.45$, $p = 0.12$].

Discussion

When medical students receive a feedback message with a negative content -about points for improvement- the framing of the message, either positively or negatively, affects their satisfaction with the feedback, their self-efficacy regarding task performance and their task performance. We conclude that the positively framed feedback group was more satisfied, had higher self-efficacy immediately after receiving feedback, and performed better 2 weeks *after* receiving feedback than the group in the negatively framed condition. In addition, we found that in the negatively framed feedback condition, students showed significantly lower performance scores at T2, i.e., even before feedback was received.

How do we explain our findings? The results on satisfaction show a pattern which is in line with other studies: positive feedback messages, such as compliments, praise (Boehler et al. 2006) or high performance feedback (Nesbit and Burton 2006) increase students' satisfaction.

The changes over time in the SE scores can be explained by both the influence of the WR performance and the feedback. Before T1 students had neither seen the instructional video nor performed the task. The explanation on the video about the WR performance and its actual performance gave students an impression of the task difficulty in relation to their own capability. This might explain why SE scores for both conditions are higher at SE2. Between SE2 and SE3 students had either received positively or negatively framed feedback and performed the task again. It appears that the feedback affects their feeling of competence regarding the WR task, which can explain further increase of self-efficacy in the positively framed feedback condition and decrease in the negatively framed feedback condition. How do we know self-efficacy was not affected by performance instead of by the feedback? No significant correlation between P2 and SE3 was found, which we interpret as SE3 being more influenced by the feedback than by P2. The time between SE3 and SE4 is 2 weeks, in which students have not practiced the WR task. The lack of practice of this new, complex task might explain why SE4 dropped so much. At SE4 students were no longer ignorant and they knew what to expect regarding task difficulty. This may explain why self-efficacy at T4 was even lower than at T1.

The findings on performance are somewhat puzzling. At P1 students in the positively framed feedback condition performed significantly better than in the negatively framed condition. Is this coincidence or an experimental effect? A possible explanation is that students in both conditions unconsciously 'perceived' signals from feedback provider's non-verbal behavior and tone of voice about the message type, and that this influenced their first performance, as the actor was aware of the condition and may have unconsciously disclosed this awareness while briefly instructing the candidate. We see a similar phenomenon in situations in which bad news is transmitted: feedback recipients often have a 'feeling' of what comes (Maynard 1996). It is very unlikely that the SP could have been instrumental in any unconscious revelation of the condition prior to T2 as she was not pre-informed about the experimental condition of the students.

We observed a non-significant P2 difference suggesting a benefit in the positively framed condition with a close to medium effect size ($r = -.25$). A larger cohort of students might have generated a significant effect. The absence of a significant effect might also be a consequence of task type. Kluger and DeNisi (1996) address this possibility, suggesting that feedback on complex tasks is less effective than on easy tasks. A possible explanation is that feedback about a complex psychomotor task might need more time to be processed and to sink in, compared to feedback on single task, or feedback on a cognitive task. The effect found at P3 seems to illustrate this.

Another explanation is that feedback given on one specific aspect of the task is applied in the next performance but at the expense of the good performance of other task aspects, and this does not lead to an increase of overall performance.

The students were randomly placed in a feedback condition; however, the two conditions were significantly different on initial performance and differences in the learning processes are retained. Despite the random allocation of students to conditions, we cannot exclude the possibility that by chance the students in de positively framed condition were somewhat better from the outset.

We do not think that drop-out of subjects has affected our results, given the fact that no differences in outcomes were found at baseline (T1), pretest (T2) and posttest (T3) when the analyses were repeated with inclusion of the data from the drop-outs.

All students received feedback from the same person, and they all performed the task on the same SP, so this does not offer an alternative explanation of our findings.

A limitation of this study is that we were not able to control for all possible influences. For reasons of ecological validity the feedback dialogue had to sound natural. Therefore, the supervisor had received only global guidelines on what had to be said during the encounter, but he was specifically trained in similar framing of the feedback in the two conditions. Thus, small variations in the feedback dialogues might have been present. No indications are present that the specific feedback *content* caused the effects we found. All students received feedback on task aspects they could improve. Difficulties with Weber and Rinne tuning fork test in medical students had previously been explored, revealing two topics on which feedback is most necessary: (a) doctor-patient communication and (b) how to use the tuning fork (Kruisheer et al. 2006).

This study was only performed in the content area of the Weber and Rinne tuning fork test, and only in one institution. Additional studies in other contexts and content areas are needed to know if our findings can be generalized. The fact that all students received feedback from one and the same person, and they all performed the task on the same SP eliminates the possibility of confounding of framing condition and SP, but it is a limitation from the perspective of the generalizability of the results.

Framing and communicating an oral feedback message in daily life cannot be disentangled from the supervisor's tone of voice, facial expressions, and body posture. We conclude that the results of this study are caused by the positive versus negative framing of the feedback, plus that non-verbal cues as tone of voice might have strengthened the impact of framing on the student's self-efficacy and satisfaction.

Implications

What is the meaning of these results for daily practice in the clinical learning environment? The positive formulation led to more satisfied students with higher self-efficacy immediate after the feedback. Although satisfaction and self-efficacy are measured on the perceptual level, we know perceptions affect behavior (Jussim 1991; Sitzman et al. 2008). Further, we see that in the positively framed feedback condition the pre-existing differences still exist after 2 weeks, which adds to the recommendation to give feedback using a positive framing. We acknowledge however that further studies need to confirm this recommendation.

Acknowledgments The authors are grateful to the Skills lab of the UMCU, especially to Jaap Buis, MD, and Jantien de Loor, MD, who offered us the possibility to use part of their teaching moments to carry out this study, and to Carlos Kok and Niels Morgenstern for the technical assistance during the experiment. Further we acknowledge Fred Tromp for his useful comments on developing the observational checklists. We would like to thank Dawi van der Stap who provided the feedback, Marijke Sterman, who fulfilled the role of the Standardized Patient. We would also like to thank Bill McGaghie for his comments on the manuscript.

Ethical standard This study was deemed exempt of ethical approval by the Medical Ethics Review Committee of the UMCU.

References

Bagai, A., Thavendiranathan, P., & Detsky, A. S. (2006). Does this patient have hearing impairment? *JAMA, 295*(4), 416–428.

Baker, D. F. (2001). The development of collective efficacy in small task groups. *Small Group Research, 32*, 451–474.

Bandura, A. (1986). Self-efficacy. In A. Bandura (Ed.), *Social foundations of thought and action. A social cognitive theory* (pp. 390–453). Englewood Cliffs, NJ: Prentice Hall.
Boatman, D. F., Miglioretti, D. L., Eberwein, C., Alidoost, M., & Reich, S. G. (2007). How accurate are bedside hearing tests? *Neurology, 68*, 1311–1314.
Boehler, M. L., Rogers, D. A., Schwind, C. J., Mayforth, R., Quin, J., Williams, R. G., & Dunnington, G. (2006). An investigation of medical students reactions to feedback: A randomised controlled trial. *Medical Education, 40*, 746–749.
Bowen, J. L., & Irby, D. M. (2002). Assessing quality and costs of education in the ambulatory setting: A review of the literature. *Academic Medicine, 77*, 621–680.
Branch, W. T., & Paranjape, A. (2002). Feedback and reflection: Teaching methods for clinical settings. *Academic Medicine, 77*, 1185–1188.
Cohen, J. (1988). *Statistical power analysis for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Publishers.
Daelmans, H. E. M., Hoogenboom, R. J. I., Scherpbier, A. J. J. A., Stehouwer, C. D., & van der Vleuten, C. P. M. (2005). Effects of an in-training assessment programme on supervision of and feedback on competencies in an undergraduate Internal medicine clerkship. *Medical Teacher, 27*, 158–163.
Dornan, T. (2006). Experienced based learning: Learning clinical medicine in workplaces. Dissertation, University of Maastricht.
Downing, S. M. (2004). On the reproducibility of assessment data. *Medical Education, 38*, 1006–1012.
Dunegan, K. J. (1993). Framing, cognitive modes and image theory: Toward an understanding of a glass half full. *Journal of Applied Psychology, 78*, 491–503.
Dunegan, K. J. (1995). Image theory: Testing the role of image compatibility in progress decisions. *Organizational Behavior and Human Decision Processes, 62*, 79–86.
Dunegan, K. J. (1996). Fines, frames, and images: Examining formulation effects on punishment decisions. *Organizational Behavior and Human Decision Processes, 68*, 58–67.
Ear, Nose & Throat Examination—Medi-Vision Films 12. (2010). Retrieved January 5, 2013, from http://www.youtube.com/watch?v=dsS7d6_k1F8
Field, A. (2005). *Discovering statistics. Using SPSS*. London: Sage.
Geister, S., Konradt, U., & Hertel, G. (2006). Effects of process on motivation, satisfaction, and performance in virtual teams. *Small Group Research, 37*, 459–489.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112.
Irby, D. (1995). Teaching and learning in ambulatory care settings: A thematic review of the literature. *Academic Medicine, 70*, 898–931.
Ivers, N., Jamtvedt, G., Flottorp, S., Young, J. M., Odgaard-Jensen, J., French, S. D., et al. (2012). Audit and feedback: Effects on professional practice and health care outcomes. *Cochrane Database Systematic Review, 6*, CD000259. doi:10.1002/14651858.
Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review, 98*, 54–73.
Kahneman, D., & Tversky, A. (1984). Choices, values and frames. *American Psychologist, 39*, 341–350.
Kilminster, S. M., & Jolly, B. C. (2000). Effective supervision in clinical practice settings: A literature review. *Medical Education, 34*, 827–840.
Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). The four levels: An overview. In D. L. Kirkpatrick & J. D. Kirkpatrick (Eds.), *Evaluating training programs. The four levels* (pp. 21–65). San Fransisco: Berett-Koehler Publishers.
Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284.
Kruisheer, E. M., van de Ridder, J. M. M., & Meeuwsen, A. (2006). *Evaluatie van knelpunten bij het vaardigheidonderwijs van het Otologisch onderzoek.* [*Students' perception of difficulties in learning tests for detecting hearing loss.*] Unpublished report, University Medical Center Utrecht, Utrecht, Netherlands.
Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76*, 149–188.
Liberman, A. S., Liberman, M., Steinert, Y., McLeod, P., & Meterissian, S. (2005). Surgery residents and attending surgeons have different perceptions of feedback. *Medical Teacher, 27*, 470–472.
Maynard, D. W. (1996). On "realization" in everyday life: The forecasting of bad news as a social relation. *American Sociological Review, 61*, 109–131.
McIlwrick, J., Nair, B., & Montgomery, G. (2006). "How Am I doing?" Many problems but few solutions related to feedback delivery in undergraduate psychiatry education. *Academic Psychiatry, 30*, 130–135.
Moher, D., Schulz, K. F., & Altman, D. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA, 285*, 1987–1991.

Morgan, N. A., & Rego, L. L. (2006). The value of different customers satisfaction and loyalty metrics in predicting business performance. *Marketing Science, 25*, 426–439.

Nease, A. A., Mudgett, B. O., & Quiñones, M. A. (1999). Relationships among feedback sign, self-efficacy, and acceptance of performance feedback. *Journal of Applied Psychology, 84*, 806–814.

Nesbit, P. L., & Burton, S. (2006). Students justice perceptions following assignment feedback. *Assessment and Evaluation in Higher Education, 31*, 655–670.

Reynolds, D. (2006). To what extent does performance-related feedback affect managers' self-efficacy? *International Journal of Hospitality Management, 25*, 54–68.

Rucker, D. D., Petty, R. E., & Briñol, P. (2008). What's in a frame anyway? A meta-cognitive analysis of the impact of one versus two sided messages on attitude certainty. *Journal of Consumer Psychology, 18*, 137–149.

Salerno, S. M., O'Malley, P. G., Pangaro, L. N., Wheeler, G. A., Moores, K. D., & Jackson, J. L. (2002). Faculty development seminars based on the one-minute preceptor improve feedback in the ambulatory setting. *Journal of General Internal Medicine, 17*, 779–787.

Schul, Y., & Ganzach, Y. (1995). The effects of accessibility of standards and decision framing on product evaluations. *Journal of Consumer Psychology, 4*, 61–83.

Schultz, K. W., Kirby, J., Delva, D., Godwin, M., Verma, S., Birtwhistle, R., et al. (2004). Medical students' and residents' preferred site characteristics and preceptor behaviours for learning in the ambulatory setting: A cross-sectional survey. *BMC Medical Education, 4*, 12. doi:10.1186/1472-6920-4-12.

Sitzman, T., Brown, K. G., Casper, W. J., Ely, K., & Zimmerman, R. D. (2008). A review and meta-analysis of the nomological network of trainee reactions. *Journal of Applied Psychology, 93*, 280–295.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Swann, W. B., Griffin, J. J., Predmore, S. C., & Gaines, B. (1987). The cognitive-affective crossfire: When self-consistency confronts self-enhancement. *Journal of Personality and Social Psychology, 52*, 881–889.

Tanner, C., & Medin, D. L. (2004). Protected values: No omission bias and no framing effects. *Psychonomic Bulletin & Review, 11*, 185–191.

ten Cate, O. (2007). Medical education in the Netherlands. *Medical Teacher, 28*, 752–757.

Turner, N. B., van de Leemput, A. J., Draaisma, A. J., Oosterveld, P., & ten Cate, Th J. (2008). Validity of the visual analogue scale as an instrument to measure self-efficacy in resuscitation skills. *Medical Education, 42*, 503–511.

van de Ridder, J. M. M., Stokking, K. M., & ten Cate, Th. J. (2006, September). Relevant variables in feedback processes in clinical education. Paper presented at the Association for Medical Education in Europe 2006 Conference; September, Genoa, Italy.

van de Ridder, J. M. M., Stokking, K. M., McGaghie, W. C., & ten Cate, Th J. (2008). What is feedback in clinical education? *Medical Education, 42*, 189–197.

Veloski, J., Boex, J. R., Grasberger, M. J., Evans, A., & Wolfson, D. B. (2006). Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical Teacher, 28*, 117–128.

Waung, M., & Jones, D. R. (2005). The effect of feedback packaging on rate reactions. *Journal of Applied Social Psychology, 35*, 1630–1655.