

RESEARCH ARTICLE

Open Access

The benefit of repetitive skills training and frequency of expert feedback in the early acquisition of procedural skills

Hans Martin Bosse¹, Jonathan Mohr², Beate Buss², Markus Krautter³, Peter Weyrich⁴, Wolfgang Herzog², Jana Jünger² and Christoph Nikendei^{2*}

Abstract

Background: Redundant training and feedback are crucial for successful acquisition of skills in simulation trainings. It is still unclear how or how much feedback should best be delivered to maximize its effect, and how learners' activity and feedback are optimally blended. To determine the influence of *high-* versus *low-frequency* expert feedback on the learning curve of students' clinical procedural skill acquisition in a prospective randomized study.

Methods: N = 47 medical students were trained to insert a nasogastric tube in a mannequin, including structured feedback in the initial instruction phase at the beginning of the training (T₁), and either additional repetitive feedback after each of their five subsequent repetitions (high-frequency feedback group, HFF group; N = 23) or additional feedback on just one occasion, after the fifth repetition only (low-frequency feedback group, LFF group; N = 24). We assessed a) *task-specific clinical skill performance* and b) *global procedural performance* (five items of the *Integrated Procedural Performance Instrument* (IPPI)); on the basis of expert-rated videotapes at the beginning of the training (T₁) and during the final, sixth trial (T₂).

Results: The two study groups did not differ regarding their baseline data. The calculated ANOVA for *task-specific clinical skill performance* with the between-subject factor 'Group' (HFF vs. LFF) and within-subject factors 'Time' (T₁ vs. T₂) turned out not to be significant ($p < .147$). An exploratory post-hoc analyses revealed a trend towards a superior performance of HFF compared to LFF after the training (T₂; $p < .093$), whereas both groups did not differ at the beginning (T₁; $p < .851$). The smoothness of the procedure assessed as *global procedural performance*, was superior in HFF compared to LFF after the training (T₂; $p < .004$), whereas groups did not differ at the beginning (T₁; $p < .941$).

Conclusion: Deliberate practice with both *high-* and *low-frequency* intermittent feedback results in a strong improvement of students' early procedural skill acquisition. *High-frequency* intermittent feedback, however, results in even better and smoother performance. We discuss the potential role of the cognitive workload on the results. We advocate a thoughtful allocation of tutor resources to future skills training.

Keywords: Cognitive workload, Procedural skills acquisition, Expert feedback, Learning curve, Nasogastric tube, Proficiency level, Skills training, Undergraduate medical education, Simulation

* Correspondence: christoph.nikendei@med.uni-heidelberg.de

²Department of General Internal and Psychosomatic Medicine, University of Heidelberg Medical Hospital, Im Neuenheimer Feld 410, 69120 Heidelberg, Germany

Full list of author information is available at the end of the article

Background

Skills lab facilities provide an effective and safe learning environment for undergraduate medical students to acquire clinical technical skills. Skills lab training leads to improved knowledge, skills, and behaviors when compared to standard clinical training or no training, with a moderate general effect for patient-related outcomes [1]. Skills lab training enables trainees to perform procedural skills faster, more accurately and more professionally on patients in terms of both technical and communicational aspects as compared to standard clinical training [2]. Furthermore, skills lab training leads to superior objective structured clinical examination (OSCE) results, both in longitudinal [3] and prospective controlled designs [4] even for long-time follow-up [5]. With regard to potential transfer, skills lab training sessions provide a better preparation for clinical clerkships [6] and result in a higher number of procedural skills being performed at bedside on wards [7]. A prerequisite for such transfer is to exercise great care in designing training models and scenarios in order to prove their validity with regard to the real clinical setting.

Four factors are described to enhance the learning of motor skills: observational practice, the learner's focus of attention, feedback, and self-controlled practice [8]. Accordingly, in their Best Evidence Medical Education (BEME) guide, Issenberg et al. [9] identified similar factors leading to a maximum benefit of simulation-based medical education. The most relevant factors were a) *repetitive, active and standardized educational experiences*, to prevent trainees from being passive bystanders, b) *educational feedback*, and c) *embedding the training in the standard curriculum*. The majority of relevant articles found in their review agree on these three factors, and a plethora of data is available on the latter two. The fact that repetition and trainees' activity per se are important factors to promote long-term retention is also unquestioned, both for low-complexity skills [10] and high-complexity skills [11].

Regarding learning curves in simulation-based education and motor-skills training, a dose-response relationship is assumed to exist, with a rising number of repetitions resulting in an increasingly superior performance until a performance plateau is reached [12,13]. So far, learning curves in skills lab training have been examined in different study populations for differing skills in various settings. For example, undergraduate students were shown to reach a plateau in endoscopic sinus surgery simulation or simulated peripheral venous cannulation after 5 to 10 trials [14,15], with fewer trials required as subjects progressed in their medical training. In a postgraduate clinical setting of anesthesia first-year residents, a rapid improvement of success for anesthesia procedures was observed during the first 20 attempts,

leading to a success rate between 65 and 85% [12]. The model of Peyton seems to add an additional advantage in very early skill acquisition [5,16], serving as a springboard which enhances the benefit from subsequent repetitive practice. However, it remains unclear how subsequent practice should be optimally timed and designed.

Although most studies show feedback to be crucial for learning, many issues of how best to deliver feedback remain a matter of debate: *how much* feedback is required to attain a maximum benefit from repetitive skills training or to reach a proficiency level in the early acquisition of procedural skills; the ideal frequency, or mode of delivery of repetitive feedback; and how repetitions and feedback are optimally blended. Moreover, it is unclear whether repeating the feedback necessarily substantially improves performance at all [17-21].

Therefore, our randomized prospective study was designed to determine the influence of *high-* versus *low-frequency* expert feedback on the learning curve of students' clinical procedural skill acquisition. Our hypothesis was that repetitive practice is beneficial for reaching a proficiency level, with an additional benefit when intermittent feedback is given at a higher frequency as opposed to intermittent feedback at a lower frequency.

Methods

Study design

The presented randomized prospective study investigated the influence of repetitive expert feedback in skills training on the learning curve of students in the early acquisition of procedural skills. Nasogastric tube placement was selected as the clinical task as this skill represents a pivotal routine procedure in internal medicine. If it is not performed accurately, severe complications may occur, resulting in considerable costs [22,23], and it is therefore an integral part of undergraduate skills training curricula [24,25]. The study was conducted over a period of two and a half weeks alongside the regular curriculum at our faculty.

Participants

Trainees were recruited via advertisements among medical students in their first or second year of medical training. A total of N = 50 participants volunteered to participate. Only right-handed individuals were eligible for inclusion in order to standardize the camera position and facilitate raters' assessment by not needing to change perspectives. Written consent was provided by all participants and data from all participants were treated anonymously. The trainees were informed that the purpose of the study was to assess their skill performance, but no further details were provided. All participants received a minor financial compensation for their study participation. Ethical principles according to

the World Medical Association Declaration of Helsinki *Ethical Principles for Medical Research Involving Human Subjects* of 2008 were adhered to. Ethics approval was granted by the ethic committee of the University of Heidelberg (Nr. S-211/2009). Students with previous experience in inserting nasogastric tubes were excluded from the study. Refusal to participate had no impact on the subsequent evaluations or other assessments in the curriculum.

Assessment prior to the training (T₀)

To control for potential confounding variables, each study participant provided data on their age, sex, handedness, previous clinical experience regarding clerkships, qualification as a paramedic or nurse, as well as a general self-efficacy rating using the *General Self-Efficacy Scale* and individual learning styles using the *Kolb Learning Style Inventory (KLSI)* in order to prove that conditions in the two groups were equal.

General self-efficacy scale

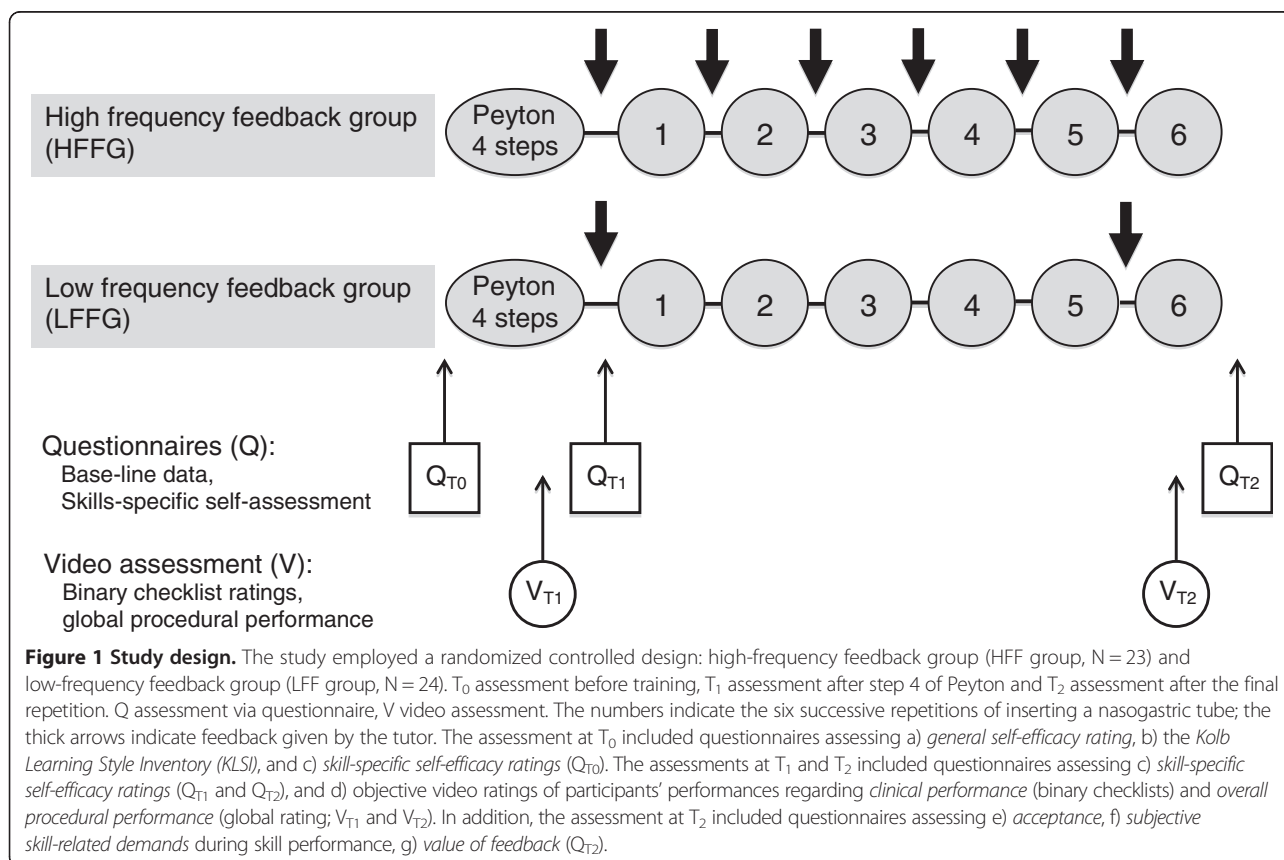
This questionnaire consists of ten positively worded items rated on a four-point Likert scale (4 = “I agree” to 1 = “I disagree”). It assesses perceived self-efficacy in the event of adversity and stressful life events [26].

Kolb Learning Style Inventory (KLSI)

The *Kolb Learning Style Inventory (KLSI)* of 2005 [27] showed an even distribution of learning styles as a potential confounder in skill acquisition although it's immediate effect is unclear [28]. On the other hand, well-organized and strategic learning styles assessed with other inventories (which may be compared to the steps *reflective observation* and *abstract conceptualization* of the KLSI) have been shown to provide a benefit for students' later performance [29,30]. Additionally, learning settings should accommodate individual learning styles to maximize individual learning achievement [31] as learning styles not only differ but may shift across cognitive and motor settings [32].

Assignment to study groups

N = 50 participants were randomly assigned to one of the two study groups, one receiving high-frequency feedback (high-frequency feedback group, HFF group; N = 25) after each of the five repetitive practice trials before the final, sixth repetition, and one receiving low-frequency feedback only twice, i.e. after the first independent skill performance and just before the final, sixth repetition (low-frequency feedback group, LFF group; N = 25; for details see “skill training session” below; see Figure 1). After participants had been included



in the study, three students opted not to participate without stating reasons or due to illness, resulting in a dropout of $N = 2$ in the HFF group and $N = 1$ in the LFF group. Thus, the final sample consisted of $N = 23$ in the HFF group and $N = 24$ in the LFF group.

Skill training session

The skill training was conducted with a student-teacher ratio of 1:1 in analogy to previous studies [24,33] and under consideration of previously published checklists for nasogastric tube placement [24,34]. Training sessions were structured as follows: Both groups received a short case history and role-play directives (i.e. including talking to the mannequin as if it were a real patient), both of which have been previously shown to enhance perceived realism in the training as well as the patient-physician communication [24,33]. Both groups were then instructed to insert a nasogastric tube in a mannequin using the four steps of *Peyton's Four-Step Approach* [16,35] and subsequently performed six further repetitions of inserting a nasogastric tube. Both groups received structured feedback from the skills lab trainers [9,36] after the first independent performance of nasogastric tube placement (step 4 of Peyton's four-step approach; T_1). However, the HFF group received performance-related feedback after each of their following five repetitions, whereas the LFF group received further feedback once, after the fifth repetition only. Finally, both groups performed a final, sixth repetition of nasogastric tube placement (T_2 ; see Figure 1). The feedback was always given immediately after the respective repetition of the skill, as proximate feedback enhances its effectiveness [37]. Feedback was positively worded and aimed at inducing an external focus, i.e. aiming at the movement effect to facilitate automaticity in motor control and promote movement efficiency (for detailed reviews, see [8,38]).

Skills lab teachers

Both the HFF and LFF group were both trained by four certified skills lab peer teachers, all of whom had at least one year of experience as skills lab trainers. Two tutors were male (both aged 22 years), and two were female (aged 21 and 22 years). The four tutors were randomly assigned to the students in the study groups. As previously shown, trained medical students as peer teachers deliver training and feedback on a par with that of faculty staff in skills training [39-41]. All trainers were blinded to the study design and received an introductory course including training in nasogastric tube placement and delivering feedback prior to the study.

Assessment of training

Assessment of the skills lab training encompassed a) acceptance ratings including value of tutor's feedback, b) subjective skill-related demands during skill performance

(cognitive workload), c) skill-specific self-efficacy ratings related to nasogastric tube placement competencies, and d) objective video expert ratings of participants' performances by blinded independent assessors ($N = 2$).

Acceptance ratings and value of trainer's feedback

For the evaluation of acceptance of the training session and the tutor's feedback, participants completed a questionnaire with five positively worded items rated on a six-point Likert scale (6 = fully agree; 1 = completely disagree) after completing the training (after T_2 , see Figure 1). For the specific pre- and post-evaluation of the value of the trainer's feedback, the participants completed an additional questionnaire with ratings on a six-point Likert scale (6 = fully agree; 1 = completely disagree) after step 4 of Peyton's four-step approach but before beginning the repetitions (T_1 , see Figure 1; 5 items) and after the final, sixth repetition (T_2 ; 12 items).

Cognitive workload/skill-related demands

We assessed the perceived overall cognitive workload using the *National Aeronautics and Space Administration Task Load Index* (NASA-TLX) [42] as a score of six subscales: *mental*, *physical*, and *temporal demands*, as well as *own performance*, *effort* and *frustration*. Assessment took place after step 4 of Peyton's four-step approach (T_1) and after the final, sixth repetition (T_2), with ratings on Likert scales from 5 (very low demands) to 100 (very high demands), resulting in a sum score between 0 (very low demands) and 100 (very high demands).

Skill-specific self-efficacy ratings

Skill-specific self-efficacy ratings related to nasogastric tube placement competencies were assessed as in previous studies [33], with five items referring to a) knowledge of the anatomical structures required to insert a nasogastric tube, b) knowledge of the materials required to insert a nasogastric tube, c) knowledge of the steps involved in inserting a nasogastric tube, d) competence in inserting a nasogastric tube in a mannequin, and e) competence in inserting a nasogastric tube in a patient (6 = fully agree; 1 = completely disagree). Skill-specific self-efficacy ratings were assessed prior to the training (T_0) as well as after step 4 of Peyton's four-step approach (T_1) and after the final, sixth repetition (T_2).

Independent video assessment of performance

Participants' performance in step 4 of Peyton's four-step approach (T_1) and in the final, sixth repetition (T_2) were videotaped in both the HFF group and the LFF group using high-resolution digital cameras with optical zoom to capture all of the details required for a precise evaluation. The videos were digitally processed and were independently rated by two clinically experienced and

trained video assessors who were blinded to both the aim and design of the study as well as the assignment of participants to the study groups. Raters were provided with binary checklists and global rating forms. The binary checklist consisted of 16 items reflecting the procedural steps of inserting a nasogastric tube [24,34,43]. Regarding binary checklists, video raters were asked to indicate whether single procedural steps were performed correctly or incorrectly. A global rating form, which was based on the *Integrated Procedural Performance Instrument* (IPPI) proposed by Kneebone et al. for the assessment of procedural skills in a clinical context, was also used [44]. The IPPI was designed to evaluate global professional and confident performance of clinical technical skills. Items of the IPPI considered relevant for the procedure were selected (items 4, 5, 9, 10, 11; six-point Likert scale; 6 = very good to 1 = unsatisfactory).

Statistical analysis

Primary endpoint was the global procedural performance. Secondary endpoints were task-specific clinical skill performance, skill-specific self-efficacy ratings, and pre- and post-assessment of trainer’s feedback (T₁ and T₂). Data are presented as means and standard deviation. Continuous data serving sample description were compared using a Student’s t-test (assuming equal variances), whereas ordinal data were assessed using Mann-Whitney U test (M-W-U-Test)). Differences in group characteristics pertaining to sex, previous education in a medical profession, and completed medical electives were compared using chi-square tests. For repeated

measures, ordinal data (acceptance ratings, cognitive workload assessed with NASA-LTX, skills-specific self-efficacy ratings, and global skills performance assessed with the IPPI) were calculated using Wilcoxon signed-rank test or Friedman test where appropriate. Group comparisons at T₁ and at T₂ were calculated using M-W-U-Tests. For interval data (task-specific clinical skills performance reflected in binary checklist ratings), an ANOVA with the between-subject factor ‘Group’ (HFF vs. LFF) and the within-subject factor ‘Time’ (T₁ vs. T₂) was conducted. LSD-post-hoc tests were used where appropriate. A p-value < .05 was considered statistically significant. Inter-rater reliability for the two video assessors was calculated using a Pearson’s-Correlation. The software package STATISTICA 8, 2007 (Statsoft, Inc., Tulsa, OK) was used for statistical analysis.

Results

Participants

There were no statistically significant differences between the two groups with regard to age, sex, completed education in a medical profession, or completed medical electives and general self-efficacy rating prior to skills training, as well as *learning styles* (see Table 1), with a distribution of subscales of learning styles as described previously [28].

Acceptance of training (rated at T₂) and assessment of trainer’s feedback (T₁ and T₂)

Participants of both study groups confirmed a high training acceptance after the skills training session.

Table 1 Group characteristics of the study groups

Group characteristics	High-frequency feedback group (HFF group) N = 23	Low-frequency feedback group (LFF group) N = 24	p-value
Age (years)	21.00 ± 2.94	20.62 ± 1.74	.596 ¹
Sex (m/f)	12 (52.17%)/11 (47.82%)	12 (50.00%)/12 (50.00%)	.882 ²
General self-efficacy rating	30.83 ± 3.42	30.42 ± 4.15	.842 ¹
Education in a medical profession	3 (13.04%)	1 (4.16%)	.276 ²
Medical electives	20 (86.95%)	21 (87.50%)	.955 ²
Kolb learning style inventory			
Concrete experience (<i>feeling</i>)	24.61 ± 7.65	24.79 ± 5.85	.927 ¹
Reflective observation (<i>watching</i>)	30.30 ± 6.88	29.88 ± 5.54	.815 ¹
Abstract conceptualization (<i>thinking</i>)	31.34 ± 8.61	32.37 ± 7.43	.663 ¹
Active experimentation (<i>doing</i>)	33.30 ± 7.55	32.95 ± 5.28	.856 ¹

¹t-test.

²χ² test.

Group characteristics of the high-frequency feedback group (HFF group, N = 23) and the low-frequency feedback group (LFF group, N = 24) are depicted regarding:

- age (age; mean ± standard deviation; Student’s t-test, p-values).
- sex (male/female; N and %, chi-square test, p-values).
- general self-efficacy rating prior to skills training (score of 10 items using Likert-scale ratings; 4 = I fully agree; 1 = I completely disagree; mean ± standard deviation; M-W-U-Test, p-values).
- completed education as paramedic, medical secretary, nurse, or occupational therapist.
- (N, % and chi-square test p-values).
- completed electives in surgery, internal medicine, pediatrics, or psychiatry (N, % and chi-square test, p-values).

Participants rated the training session as realistic (HFF group 4.66 ± .98, LFF group 4.34 ± 1.13, p = .285) and the tutor’s feedback as objective (HFF group 5.96 ± .21, LFF group 5.96 ± .20, p = .992), motivating (HFF group 5.74 ± .45, LFF group 5.67 ± .76, p = .858), supportive (HFF group 5.96 ± .21, LFF group 5.92 ± .28, p = .825), and courteous (HFF group 6.0 ± .0, LFF group 6.0 ± .0, p = 0.992), with a positive effect on learning success (HFF group 6.0 ± .0, LFF group 5.92 ± .28, p = 0.635) all ratings are given as mean of Likert scale ratings from 6 = fully agree; 1 = completely disagree).

In the pre- and post-assessment, both groups assessed trainers’ feedback as very valuable after step 4 of Peyton’s four-step approach (T₁) as well as after the final, sixth repetition (T₂), with no significant difference between the study groups (see Table 2). Both groups assessed trainers’ feedback as more valuable after the final, sixth repetition (T₂) compared to after step 4 of Peyton’s four-step approach (T₁) (HFF p = .004; LFF p = .018).

Workload/skill-related demands perceived at T₁ and T₂

Both groups rated the overall skill-related demands as high, with no differences between the HFF and LFF group after step 4 of Peyton’s four-step approach (T₁; HFF group 50.78 ± 12.0%; LFF group 47.15 ± 17.3%; p = .395; mean of scores from 0 = very low to 100 = very high using the *National Aeronautics and Space Administration Task Load Index, NASA-TLX*). After the final, sixth repetition (T₂), both groups rated the

task as less demanding compared to T₁ (HFF group T₁ 50.78 ± 12.02, T₂ 40.51 ± 18.74, p.003; LFF group T₁ 47.15 ± 17.35, T₂ 35.97 ± 16.58, p = <.001), with no differences emerging between the HFF group and the LFF group at T₂ (HFF 40.51 ± 18.74; LFF 35.97 ± 16.58; p = .407).

Skill-specific self-efficacy ratings (T₀, T₁ and T₂)

Self-efficacy ratings related to competencies regarding nasogastric tube placement improved substantially over the course of the training (HFF p < .001; LFF p < .001), but there was no significant difference between the study groups prior to the training (T₀), after step 4 of Peyton’s four-step approach (T₁) or after the final, sixth repetition (T₂; see Table 3).

Independent video assessment: task-specific clinical skill performance by expert binary checklist rating (T₁, T₂)

For expert binary checklist ratings, an ANOVA with the between-subject factor ‘Group’ (HFF vs. LFF) and within-subject factors ‘Time’ (T₁ vs. T₂) was conducted. The calculated ANOVA was not statistically significant for all effects (main effects and interaction; p = .147). Nevertheless, an exploratory post-hoc analysis was performed according to our predefined hypotheses. As expected, during step 4 of Peyton’s four-step approach (T₁), no difference was found between the HFF group and the LFF group in the total score for specific clinical skill performance (p = .851; see Table 4). At T₂, both groups scored higher on *binary checklist ratings* compared

Table 2 Pre- and post-assessment of trainer’s feedback (T1 and T2) High-frequency feedback group

	High-frequency feedback group (HFF group) N = 23	Low-frequency feedback group (LFF group) N = 24	p-value ¹
<i>After Peyton’s step 4 (T1)</i>			
Item 1 <i>The tutor’s feedback was comprehensible.</i>	5.83 ± .38	5.84 ± .38	.931
Item 2 <i>I could easily follow the tutor’s instructions.</i>	5.39 ± .66	5.17 ± .96	.077
Item 3 <i>The feedback was helpful for improving skills.</i>	5.78 ± .42	5.79 ± .51	.382
Item 4 <i>The tutor was attentive and concentrated.</i>	5.83 ± .39	5.83 ± .38	.931
Item 5 <i>The tutor seemed competent during feedback.</i>	5.83 ± .39	5.87 ± .34	.518
Mean Items 1-5	5.73 ± .31	5.70 ± .46	.876
<i>After the final, sixth repetition (T2)</i>			
Item 1 <i>The tutor’s feedback was comprehensible.</i>	5.96 ± .21	6.00 ± .00	1.000
Item 2 <i>I could easily follow the tutor’s instructions.</i>	5.83 ± .39	5.58 ± .50	.224
Item 3 <i>The feedback was helpful for improving skills.</i>	5.91 ± .29	5.96 ± .20	.108
Item 4 <i>The tutor was attentive and concentrated.</i>	5.87 ± .34	5.92 ± .28	.351
Item 5 <i>The tutor seemed competent during feedback.</i>	6.00 ± .00	5.92 ± .28	1.000
Mean Items 1-5	5.91 ± .17	5.88 ± .19	.341
p-value²	.004	.018	

¹Mann Whitney U test.

²Wilcoxon signed rank test.

Perceived value of the feedback between the two groups after Peyton’s step 4 (T₁), and after the final, 6th repetition (T₂); mean and standard deviation of six-point Likert scales from 6 = fully agree to 1 = completely disagree; M-W-U-Test, p-values).

Table 3 Skill-specific self-efficacy ratings

	High-frequency feedback group (HFF group) N = 23	Low-frequency feedback group (LFF group) N = 24	MWU p-value ¹
Prior to the training (T0)	2.35 ± .71	2.33 ± .64	.640
After Peyton's step 4 (T1)	4.70 ± .59	4.86 ± .74	.296
After the final, sixth repetition (T2)	5.27 ± .39	5.37 ± .40	.872
p-value²	<.001	<.001	

¹Mann Whitney U test.

²Friedman test.

Self-efficacy ratings relating to five competencies in inserting a nasogastric tube between the two groups before the training (T₀), after Peyton's step 4 (T₁), and after the final, 6th repetition (T₂; mean and standard deviation of six-point Likert scales from 6 = fully agree to 1 = completely disagree; M-W-U-Test, p-values).

to T₁ (p < .001) with the HFF group showing higher scores at T₂, although not significantly (p < .093; see Table 4).

final, sixth repetition when using global performance ratings.

Independent video assessment: global procedural performance rating (T₁, T₂)

Global ratings of procedural performance assessed with the *Integrated Procedural Performance Instrument* (IPPI) also revealed no difference at the beginning of the training during step 4 of Peyton's four-step approach (T₁; p = .941). At T₂, both groups scored higher in their *global procedural performance compared to T1* (HFF group p < .001; LFF group p = .002). In addition, after the final, sixth repetition (T₂), the HFF group achieving better scores than the LFF group (p = .004; see Table 4).

Discussion

In this randomized prospective study, we assessed the influence of frequency of expert feedback during redundant practice in the early acquisition of students' clinical procedural skills. Both *high-* and *low-frequency* intermittent feedback resulted in a significant improvement of students' clinical procedural skill performance as primary endpoint, in line with earlier findings [1,2,4,45]. *High-frequency* intermittent feedback, however, proved to result in an even better procedural performance compared to *low-frequency* intermittent feedback. Regarding the exploratory analyses of task specific performance (as reflected by our binary checklist ratings), both groups benefited from the training but with no significant difference between the two groups. A limitation of our results is that high scores in skill performance were already found during the very first independent performance, which we attribute to the success of Peyton's four-step framework of deconstruction and learners' comprehension before actually performing the skill.

Inter-rater reliability for independent video raters

Standardized inter-rater reliability for independent video raters was .79 for the assessment of step 4 of Peyton's four-step approach, and .75 for the evaluation of the final, sixth repetition when using binary checklist ratings, and .76 for the assessment of step 4 of Peyton's four-step approach, and .81 for the evaluation of the

Table 4 Task-specific clinical skill performance and global procedural performance

Task-specific clinical skill performance (binary checklists)			
	High-frequency feedback group (HFF group) N = 23	Low-frequency feedback group (LFF group) N = 24	p-value ¹
Peyton's step 4 (T1)	91.06 ± 7.48	91.42 ± 9.14	.851
Final, 6th repetition (T2)	99.22 ± 2.25	96.04 ± 4.96	.093
p value ¹	<.001	<.001	
Global procedural performance (global rating)			
	High-frequency feedback group (HFF group) N = 23	Low-frequency feedback group (LFF group) N = 24	p-value ²
Peyton's step 4 (T1)	5.31 ± 0.50	5.30 ± .64	.941
Final, 6th repetition (T2)	5.95 ± 0.07	5.65 ± .48	<.004
p value ³	<.001	.002	

¹LSD-post-hoc Tests.

²Mann-Whitney-U Test.

³Wilcoxon signed-rank test (T1 vs. T2).

Performance ratings of the two groups in Peyton's step 4 (T₁) and in the final, 6th repetition (T₂) in *task-specific clinical skill performance* (binary checklist rating as mean score in percent of maximum achievable points and standard deviation; checklist of 16 items with a minimum of 0 and a maximum of 16 points; ANOVA, p-values) and *global procedural performance* (global performance rating as mean score of global rating scales ± standard deviation; six-point Likert scale from 6 = very good to 1 = unsatisfactory; M-W-U-Test, p-values).

Regarding the investigated cohort, there was no difference between the two study groups prior to the experiment with respect to age, sex, prior medical education, number and field of chosen electives, general and skill-specific self-efficacy ratings relating to nasogastric tube placement, or individual learning styles. Pre- and post-skill-specific self-efficacy ratings were higher in both groups, with no significant difference. Although the correlation between self-efficacy – reflecting a modification of physicians' perception, motivation, and activity [46] – and superior objective performance measures is called into doubt in the literature [47], previously published research implies that higher self-efficacy in skills training results in more rigorous demands by the students for supervision during the performance of clinical skills at bedside on patient wards [46-48]. Our results strengthen the findings that – irrespective of the frequency of intermittent feedback – repetitive deliberate practice represents one of the most important factors to enhance self-efficacy.

Taking a closer look at the learning curve (dependent on the complexity of the motor skill), both groups' performance started from the same level at beginning of the training with relatively high scores after the first 4-step Peyton training, even before the first feedback (90%, see Table 4) and reaching a proficiency level of 95% after five consecutive repetitions. Thus we assume a medium degree of complexity compared to other settings [49-52].

The feedback of trainers aimed at aspects of trainees' performance with better than average performance and the trainees' very high rating of this feedback associated to their improved performance is in line with previous findings suggesting both informational and motivational influences of such feedback (for a comprehensive review, see [8]). As the two groups rated the value of trainers' feedback equally high, we attribute effects of the study to the *variance in frequency* of feedback. Notably, more feedback, more frequent feedback, or even concurrent feedback, may not always necessarily have a positive impact [19] or even degrade learning [53]. In our study, we were able to show that *intermittent* feedback – at both low and high frequency – supports trainees' performance in a complex motor skill.

This increase in performance among both study groups in terms of smoothness of the procedure (as reflected by our global rating scores), was more pronounced after a higher frequency of intermittent feedback. This is in contrast to other findings: Receiving no feedback in a surgical simulator setting led to more instrument smoothness than with receiving feedback – even though, as expected, more mistakes occurred [21]. Too intense feedback during the early stages of skill acquisition may even hinder learning [20]. This fact that reducing the frequency of feedback related to performance enhances motor skills learning has

been described previously [54] and has not yet been satisfactorily explained. In our study high frequency intermittent feedback more affects the secureness and smoothness of the procedure to be performed, than the accurate performance of the single sub-steps (task-specific clinical skill performance).

Limitations

Both groups already achieved high scores in their nasogastric skill performance during the very first independent performance, which could be attributed to the success of Peyton's four-step framework of deconstruction and the learners' comprehension before actually performing the skill. This only leaves small scope for improvement – but of note, there is still a significant effect for both groups. Notably, within Peyton's four-step framework, there is no scheduled specific feedback. In line with previous findings, our undergraduates reached mastery after their sixth repetition. For more experienced members of the medical community, i.e. graduates, reduced or delayed feedback as well as a more complex instruction might be more successful [19].

We measure training performance during practice and during early stages of skills learning; currently, a final conclusion for retention and transfer of this performance to future practice is hard to draw. Our work group has addressed this apparent gap in literature previously; our setting seems to provide more potential for retention than the traditional bedside teaching [2,5].

Conclusion

As a conclusion, the optimal benefit from feedback seems to be a question of timing and dosing. Regarding the pros and cons of frequency or timing of feedback, we were able to show that both *high-* and *low-frequency* intermittent positive feedback leave sufficient time for the learner for self-controlled practice as compared to continuous concurrent (= permanent) feedback [8]. Our intermittent feedback results in a positive impact on the on skill-specific self-efficacy and the learning curve of students' clinical procedural skill acquisition, with a moderate reduction of cognitive workload over the training. In contrast, continuous concurrent (= permanent) feedback during skill acquisition may degrade the learning of skills: "It may be better to wait" [55] for a trainee to finish performing a defined sequence. In this sense, we conclude that both *low-frequency* and *high-frequency* positive feedback leave sufficient resources for the learner's cognitive processing, with *high-frequency* positive feedback potentially being more demanding in this sense. However, apparently, this still leaves sufficient resources for this group to take advantage of their additional intermittent feedback in order to achieve mastery.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors substantially contributed to the conception and planning of study as well as to the drafting of the manuscript. All authors read and approved the final manuscript. CN, JM, and HMB were responsible for conception and design of the study. JM undertook the video documentation and blinding for rating; blinded video rating was performed by HMB, MK, and PW. BB, JM, and CN organized the training sessions and served as tutors for instruction and feedback. CN and JM organised the assessment. CN performed the statistical analysis. CN, HMB, JJ, PW and WH contributed to interpretation of data. CN, HMB, and JM drafted the manuscript, CN, HMB, JJ, JM, and WH wrote the manuscript.

Acknowledgments

We acknowledge financial support by Deutsche Forschungsgemeinschaft and Ruprecht-Karls-Universität Heidelberg within the funding programme Open Access Publishing.

Author details

¹General Pediatrics, Neonatology and Pediatric Cardiology, University Children's Hospital, Düsseldorf, Germany. ²Department of General Internal and Psychosomatic Medicine, University of Heidelberg Medical Hospital, Im Neuenheimer Feld 410, 69120 Heidelberg, Germany. ³Department of Nephrology, University of Heidelberg, Heidelberg, Germany. ⁴University Department of Medicine, Department of Internal Medicine IV, Endocrinology and Diabetology, Angiology, Nephrology and Clinical Chemistry, University of Tübingen, Tübingen, Germany.

Received: 30 June 2014 Accepted: 5 January 2015

Published online: 19 February 2015

References

1. Lynagh M, Burton R, Sanson-Fisher R. A systematic review of medical skills laboratory training: where to from here? *Med Educ*. 2007;41:879–87.
2. Lund F, Schultz JH, Maatouk I, Krautter M, Moltner A, Werner A, et al. Effectiveness of IV cannulation skills laboratory training and its transfer into clinical practice: a randomized, controlled trial. *PLoS One*. 2012;7:e32831.
3. Bradley P, Bligh J. One year's experience with a clinical skills resource centre. *Med Educ*. 1999;33:114–20.
4. Jünger J, Schäfer S, Roth C, Schellberg D, Friedman Ben-David M, Nikendei C. Effects of basic clinical skills training on objective structured clinical examination performance. *Med Educ*. 2005;39:1015–20.
5. Herrmann-Werner A, Nikendei C, Keifenheim K, Bosse HM, Lund F, Wagner R, et al. "Best practice" skills lab training vs. a "see one, do one" approach in undergraduate medical education: an RCT on students' long-term ability to perform procedural clinical skills. *PLoS One*. 2013;8:e76354.
6. Remmen R, Scherpbier A, van der Vleuten C, Denekens J, Derese A, Hermann I, et al. Effectiveness of basic clinical skills training programmes: a cross-sectional comparison of four medical schools. *Med Educ*. 2001;35:121–8.
7. Remmen R, Derese A, Scherpbier A, Denekens J, Hermann I, van der Vleuten C, et al. Can medical schools rely on clerkships to train students in basic clinical skills? *Med Educ*. 1999;33:600–5.
8. Wulf G, Shea C, Lewthwaite R. Motor skill learning and performance: a review of influential factors. *Med Educ*. 2010;44:75–84.
9. Issenberg S, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach*. 2005;27:10–28.
10. Lammers RL. Learning and retention rates after training in posterior epistaxis management. *Acad Emerg Med*. 2008;15:1181–9.
11. Ericsson KA. An expert-performance perspective of research on medical expertise: the study of clinical performance. *Med Educ*. 2007;41:1124–30.
12. Konrad C, Schüpfer G, Wietlisbach M, Gerber H. Learning manual skills in anesthesiology: is there a recommended number of cases for anesthetic procedures? *Anesth Analg*. 1998;86:635–9.
13. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. Effect of practice on standardised learning outcomes in simulation-based medical education. *Med Educ*. 2006;40:792–7.

14. Loukas C, Nikiteas N, Kanakis M, Moutsatsos A, Leandros E, Georgiou E. A virtual reality simulation curriculum for intravenous cannulation training. *Acad Emerg Med*. 2010;17:1142–5.
15. Uribe JJ, Ralph Jr WM, Glaser AY, Fried MP. Learning curves, acquisition, and retention of skills trained with the endoscopic sinus surgery simulator. *Am J Rhinol*. 2004;18:87–92.
16. Peyton JWR. *Teaching & Learning in Medical Practice*. Manticore Europe: Rickmansworth; 1998.
17. Auerbach M, Kessler D, Foltin JC. Repetitive pediatric simulation resuscitation training. *Pediatr Emerg Care*. 2011;27:29–31.
18. Kruglikova I, Grantcharov TP, Drewes AM, Funch-Jensen P. The impact of constructive feedback on training in gastrointestinal endoscopy using high-fidelity Virtual-Reality simulation: a randomised controlled trial. *Gut*. 2010;59:181–5.
19. Vickers JN, Livingston LF, Umeris-Bohnert S, Holden D. Decision training: the effects of complex instruction, variable practice and reduced delayed feedback on the acquisition and transfer of a motor skill. *J Sports Sci*. 1999;17:357–67.
20. Stefanidis D, Korndorffer Jr JR, Heniford BT, Scott DJ. Limited feedback and video tutorials optimize learning and resource utilization during laparoscopic simulator training. *Surgery*. 2007;142:202–6.
21. Boyle E, Al-Akash M, Gallagher AG, Traynor O, Hill AD, Neary PC. Optimising surgical training: use of feedback to reduce errors during a simulated surgical procedure. *Postgrad Med J*. 2011;87:524–8.
22. Metheny NA, Meert KL, Clouse RE. Complications related to feeding tube placement. *Curr Opin Gastroenterol*. 2007;23:178–82.
23. De Aguilar-Nascimento JE, Kudsk KA. Clinical costs of feeding tube placement. *J Parenter Enteral Nutr*. 2007;31:269–73.
24. Nikendei C, Zeuch A, Dieckmann P, Roth C, Schafer S, Volkl M, et al. Role-playing for more realistic technical skills training. *Med Teach*. 2005;27:122–6.
25. Nikendei C, Schilling T, Nawroth P, Hensel M, Ho AD, Schwenger V, et al. Integrated skills laboratory concept for undergraduate training in internal medicine. *Dtsch Med Wochenschr*. 2005;130:1133–8.
26. Schwarzer R, Jerusalem M. Generalized Self-Efficacy Scale. In: Weinman J, Wright S, Johnston M, editors. *Measures In Health Psychology: A User's Portfolio Causal and Control Beliefs* (Pp 35-37). Windsor, UK: NFER-NELSON; 1995. p. 35–7.
27. Kolb AY. *The Kolb Learning Style Inventory–Version 3.1 2005 Technical Specifications*. 2005.
28. Lynch TG, Woelfl NN, Steele DJ, Hanssen CS. Learning style influences student examination performance. *Am J Surg*. 1998;176:62–6.
29. Martin IG, Stark P, Jolly B. Benefiting from clinical experience: the influence of learning style and clinical experience on performance in an undergraduate objective structured clinical examination. *Med Educ*. 2000;34:530–4.
30. McManus IC, Richards P, Winder BC, Sproston KA. Clinical experience, performance in final examinations, and learning style in medical students: prospective study. *BMJ*. 1998;316:345–50.
31. Coker CA. Consistency of learning styles of undergraduate athletic training students in the traditional classroom versus the clinical setting. *J Athl Train*. 2000;35:441–4.
32. Coker CA. Learning style consistency across cognitive and motor settings. *Percept Mot Skills*. 1995;81(3 Pt 1):1023–6.
33. Nikendei C, Kraus B, Schrauth M, Weyrich P, Zipfel S, Herzog W, et al. Integration of role-playing into technical skills training: a randomized controlled trial. *Med Teach*. 2007;29:956–60.
34. Nikendei C. *Legen einer Magensonde*. In: Jünger J, Nikendei C, editors. *Kompetenzzentrum für Prüfungen in der Medizin OSCE Prüfungsvorbereitung Innere Medizin*. Stuttgart: Thieme; 2005. p. 40–141.
35. Krautter M, Weyrich P, Schultz JH, Buss SJ, Maatouk I, Junger J, et al. Effects of Peyton's four-step approach on objective performance measures in technical skills training: a controlled trial. *Teach Learn Med*. 2011;23:244–50.
36. Veloski J, Boex JR, Grasberger MJ, Evans A, Wolfson DB. Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Med Teach*. 2006;28:117–28.
37. Parmar S, Delaney CP. The role of proximate feedback in skills training. *Surgeon*. 2011;9 Suppl 1:526–7.
38. Ende J. Feedback in clinical medical education. *JAMA*. 1983;250:777–81.

39. Weyrich P, Celebi N, Schrauth M, Moltner A, Lammerding-Koppel M, Nikendei C. Peer-assisted versus faculty staff-led skills laboratory training: a randomised controlled trial. *Med Educ.* 2009;43:113–20.
40. Tolsgaard MG, Gustafsson A, Rasmussen MB, Hoiby P, Muller CG, Ringsted C. Student teachers can be as good as associate professors in teaching clinical skills. *Med Teach.* 2007;29:553–7.
41. Weyrich P, Schrauth M, Kraus B, Habermehl D, Netzhammer N, Zipfel S, et al. Undergraduate technical skills training guided by student tutors—analysis of tutors' attitudes, tutees' acceptance and learning progress in an innovative teaching model. *BMC Med Educ.* 2008;8:18.
42. Hart SG. NASA-Task Load Index (NASA-TLX); 20 Years Later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Volume 50. 2006. p. 904–8.
43. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med.* 1998;73:993–7.
44. Kneebone R, Bello F, Nestel D, Yadollahi F, Darzi A. Training and assessment of procedural skills in context using an Integrated Procedural Performance Instrument (IPPI). *Stud Health Technol Inform.* 2007;125:229–31.
45. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA.* 2011;306:978–88.
46. Bandura A. Perceived self-efficacy in cognitive development and functioning. *Educ Psychol.* 1993;28:117–48.
47. Davis A. MP: Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *JAMA.* 2006;296:1094–102.
48. Chen W, Liao S, Tsai C, Huang C, Lin C, Tsai C. Clinical skills in final-year medical students: the relationship between self-reported confidence and direct observation by faculty or residents. *Ann Acad Med Singapore.* 2008;37:3–8.
49. Young A, Miller JP, Azarow K. Establishing learning curves for surgical residents using Cumulative Summation (CUSUM) analysis. *Curr Surg.* 2005;62:330–4.
50. Van Rij AM, McDonald JR, Pettigrew RA, Putterill MJ, Reddy CK, Wright JJ. Cusum as an aid to early assessment of the surgical trainee. *Br J Surg.* 1995;82:1500–3.
51. Schlup MM, Williams SM, Barbezat GO. ERCP: a review of technical competency and workload in a small unit. *Gastrointest Endosc.* 1997;46:48–52.
52. Sutton RM, Niles D, Meaney PA, Aplenc R, French B, Abella BS, et al. Low-dose, high-frequency CPR training improves skill retention of in-hospital pediatric providers. *Pediatrics.* 2011;128:e145–51.
53. Schmidt RA, Wulf G. Continuous concurrent feedback degrades skill learning: implications for training and simulation. *Hum Factors.* 1997;39:509–25.
54. Winstein CJ, Schmidt RA. Reduced frequency of knowledge of results enhances motor skill learning. *J Exp Psychol Learn Mem Cogn.* 1990;16:677.
55. Walsh CM, Ling SC, Wang CS, Carnahan H. Concurrent versus terminal feedback: it may be better to wait. *Acad Med.* 2009;84(10 Suppl):S54–7.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

