

Multisource feedback in residency training: A quantitative study to investigate the feedback conversation

Abstract

Introduction: Multisource Feedback (MSF) is one form of assessment in medical training. It provides individual feedback based on multiple ratings, which can then be used to develop learning goals during a feedback conversation between supervisor and resident. We conducted this study to investigate how the resident and supervisor set the learning goals and what the learning goals refer to.

Methods: The study comprised 75 sets of MSF, each consisting of 12–15 external ratings per resident and one self-rating per resident. Data included 1015 external ratings and 75 self-ratings. As ten sets missed learning goals, only 65 sets of MSF could be analysed. Data comprised written MSF feedback, including scale-based ratings and narrative comments, structured minutes from the feedback conversations and the resulting learning goals, which were sorted into themes. We used multiple linear regressions to determine the associations between feedback data, conversation topics and learning goals.

Results: Topics were more likely to be discussed as strengths during the feedback conversation if scale-based ratings were high and there were many favourable comments on an item. Topics were more likely to be discussed as areas for improvement if the number of unfavourable comments was high. Topics with many (favourable and unfavourable) comments and topics discussed as an area for improvement were more likely to result in learning goals. We found a number of learning goals beyond the competences on the MSF-questionnaire, that can be understood as connected to Professional Identity Formation.

Conclusion: As the feedback and learning goals clearly exceeded the expected competences from the MSF questionnaire, we see the need for addressing these broader topics of residents' development. Hence, we encourage supervisors and residents to explicitly include Professional Identity Formation topics such as career plans, engagement in research or personal attitudes into regular assessments and feedback conversations. Thus, MSF might be a fitting tool to support professional identity development.

Keywords: professional identity formation, formative assessment, multisource feedback, residency training, workplace-based assessment

Eva K. Hennel^{1,2}

Felicitas-M. Lahner^{1,3}

Noemi Zweifel⁴

Sigrid Harendza⁵

Kathrin Neuhaus⁴

Sören Huwendiek¹

1 University of Bern, Institute for Medical Education, Department for Assessment and Evaluation (AAE), Bern, Switzerland

2 Schweizerisches Institut für ärztliche Weiter- und Fortbildung SIWF, Bern, Switzerland

3 University of Applied Sciences, Department of Health Professions, Bern, Switzerland

4 University Children's Hospital Zürich, Department of Surgery, Zürich, Switzerland

5 University Medical Centre Hamburg-Eppendorf, III. Department of Internal Medicine, Hamburg, Germany

1. Introduction

Multisource Feedback (MSF), or 360° assessment, is used to obtain specific feedback to support medical training, mainly the development of skills and competences [1], [2], [3], [4]. In this article we argue how MSF might also be used to support Professional Identity Formation as described by Jarvis-Selinger [5]. Unlike other workplace-based assessments, MSF includes several perspectives on a resident's performance in various situations over time. In some settings of MSF, a supervisor summarises and "transmits" the feedback, including all scale-based ratings and written narrative comments, to the resident. During this feedback conversation, they compare the feedback to the resident's self-assessment

and define the resident's learning goals, guiding each resident individually.

Most studies on MSF focused on consultants and specialists as recipients of the feedback and investigated factors that influence the effectiveness and outcomes of MSF [6], [7], [8], [9], [10], [11]. Despite recommendations for feedback conversations to guide emotional responses [11] and reflection [9], few studies have investigated factors influencing MSF in settings where these conversations actually occurred [6], [7], [9], [11]. Some studies based the need for feedback conversations on participant expectations rather than experiences [8], [10], [12]. No study analysed the content or process of these conversations, or observed the facilitators and residents interacting. We thus lack understanding of how facilitators filter

and summarize MSF ratings or use them to create learning goals. In non-medical settings, where feedback was given without a conversation, the learning goals were shaped by the number of narrative comments, the polarity of comments (favourable or unfavourable), and the focus of comments (task-focused versus trait-focused) [13]. A better knowledge on the content of the feedback conversation between supervisor and resident is needed to understand how learning goals are set and what they refer to. In this study, we investigate how MSF ratings influenced the feedback conversation and how both shaped the number and content of the resident's learning goals. We aim to better inform the training of supervisors and residents to improve the targeted use of MSF.

2. Methods

2.1. Ethics

The local committee of the Association of Swiss Ethics Committees deemed the study exempt from further approval. All participants gave written informed consent to participate and publish their anonymised data.

2.2. Setting

MSF was conducted at the Department of Surgery of the University Children's Hospital Zurich, Switzerland, where it is a mandatory part of residency training with a formative purpose. Implemented in 2015, based on best-practice literature [6], [7], [9], [14], [15], [16], [17], [18], [19], [20], all participants (residents, raters and supervisors) were taught the objectives, content, and the MSF questionnaire, including the rating scale and feedback rules. Supervisors learned to give specific feedback while preserving rater anonymity. The MSF questionnaire, was based on the Can-MEDS roles [21] and previously described in detail [22].

The MSF process in this study involved up to 15 raters from predefined groups of co-workers providing feedback twice a year. Raters completed the MSF questionnaire online with feedback later presented anonymously as part of a summary with other raters' feedback. During the feedback conversation the supervisor shared the means of the scale-based ratings with the resident and summarised anonymised narrative comments they felt were important. The supervisor and the resident decided which items on the MSF questionnaire to discuss, set the topics of the feedback conversation, and decided of the number and content of learning goals. They documented MSF ratings, strengths or areas for improvement, and the learning goals on a structured form.

2.3. Study design and data collection

To address the specific literature gap around the feedback conversation, we deliberately decided for a quantitative study design. To trace the way of data during the turn of

the conversation as exactly as possible, we disassembled the conversation into three smaller steps, each of which containing documented, quantifiable data.

This study comprised 75 sets of MSF for residents, each consisting of 12-15 external ratings per residents and one self-rating per resident. Data included 1015 external ratings and 75 self-ratings. As ten sets missed learning goals, only 65 sets of MSF could be analysed. More detailed information and an overview on the three components of the feedback conversation are shown in table 1.

2.4. Definition and calculation of narrative comments, learning goals, and influencing factors

Data coding was based on [13] and [23].

2.4.1. Polarity of narrative comments

Narrative comments that clearly praised or reinforced were coded "favourable". Corrections and critical comments, including those made in a critical tone ("sometimes too motivated", "not sure, if knowledge is enough") were coded "unfavourable". Comments that could not clearly be sorted were coded "cannot be allocated".

2.4.2. Quality of narrative comments

Narrative comments were also coded as task-focused or trait-focused. Comments that clearly described the way a certain task had been or should be conducted, were coded "task-focused". Comments that directly commented on the person or their attitudes were coded "trait-focused". Comments that could not clearly be sorted were coded "cannot be allocated".

2.4.3. Number of learning goals

We defined and counted learning goals based on their content. We had access to 75 sets of minutes from feedback conversations but excluded 13 of them from the analysis of learning goals because they contained no notes on learning goals (these residents were at the end of their training at the department). Notes from two other conversations were lost, so our data set included the text of 60 of the 75 sets of minutes.

2.4.4. Content of learning goals

We checked to see if each documented learning goal could be assigned to one or more items on the MSF questionnaire. Goals associated with more than one item (e.g., "acquisition of knowledge") or goals that did not seem to be connected to the MSF questionnaire were analysed qualitatively and counted separately. We then assigned all learning goals to CanMEDS roles (see table 2).

Table 1: Data collection and analyses according to the three parts of MSF investigated in this study

	Data collected on 75 occasions of MSF	Coding and calculations as part of the analysis
Part (i) MSF questionnaires		
Components of the MSF questionnaires	1015 external ratings: 12 to 15 scale-based external ratings per resident, consisting of scale-based ratings for each item and narrative comments for each item	Number of narrative comments, including polarity (favourable/unfavourable) and focus (task-focused/ trait-focused)
	75 self-assessments: 1 scale-based self-rating per MSF per resident, consisting of scale-based ratings for each item and a narrative comment for each item	Gap between external ratings and self-rating
Part (ii) topics discussed during feedback conversation		
	List of all items with information, if discussed as a strength or an area for improvement or not discussed	
Part (iii) learning goals generated during feedback conversation		
	List of learning goals	Learning goals: number, formulation, content assigned to CanMEDS roles

Table 2: Assumptions for multiple linear models

		Model 1: Discussion of strengths	Model 2: discussion of areas for improvement	Model 3: Influence on learning goals
VIF	Scale-based external rating	1.63	1.63	1.79
	Number of comments	8.61	8.61	9.35
	Number of unfavourable comments	2.72	2.72	9.58
	Number of favourable comments	8.07	8.07	3.99
	Gap between external ratings and self-rating	1.11	1.11	1.06
	Discussed as strength			1.19
	Discussed as an area for improvement			1.14
Breusch-Pagan Test	Chi2	118.43	1129.80	26.23
	p	<0.05	<0.05	<0.05
Shapiro-Wilk Test	W	0.91	0.46	0.81
	p	0.09	0.12	0.23

2.4.5. Influencing factors

We calculated influencing factors based on 1015 scale-based external ratings and 75 scale-based self-ratings. All 16 items of the MSF questionnaire were rated on a 5-point scale, from “below my expectations” (1) to “far above my expectations” (5), or alternatively “unable to comment” with space provided for narrative comments directly after each item.

2.5. Regression analyses

We used multiple linear regression analyses to determine the influence of scale-based external ratings, the number of comments, the number of favourable comments, the number of unfavourable comments and the gap between external ratings and self-ratings. To check that the assumptions for multiple linear regressions were met, we used the Breusch-Pagan test for analysing homoscedasti-

city, the VIF value for analysing multicollinearity, and the Shapiro-Wilk test for normal distribution of residuals. As all three models showed a significant Breusch-Pagan test indicating heteroscedasticity we used heteroscedasticity consistent standard errors. In all three models VIF values were below 10, indicating no multicollinearity. According to the Shapiro-Wilk test residuals were normally distributed. Results of the statistical analysis can be found in table 2.

We displayed the unstandardized beta (B), the standard error for the unstandardized beta (SE(B)), the standardized beta (β), the t-test statistic (t), and the probability value (p). Our regression analyses first investigated the influence of those variables on the discussion of strengths, and, second, on the discussion of areas of improvement, and third, on learning goals set during the feedback conversation. As an index of effect size, we report R^2 . The level of significance was set at $p < 0.05$. All analyses were conducted using R (version 3.2.0) [24].

Table 3: Allocation of items of the MSF questionnaire and learning goals to CanMEDS roles

Three goals which concerned the overall performance and 24 goals which concerned career planning were not allocated and not included into the calculations

CanMEDS role	Corresponding items of the MSF questionnaire (beginning with “The doctor...”) and/or topics of learning goals
Medical Expert	The doctor diagnoses patient problems correctly.
	The doctor formulates appropriate management plans.
	The doctor is aware of her/his own limitations and asks for help in that situation.
	The doctor has good manual and technical skills.
	Acquiring knowledge
Communicator	The doctor keeps records in a timely and accurate manner.
	The doctor communicates adequately with patients and family members.
	The doctor involves psychosocial aspects.
	Showing presence
Collaborator	The doctor communicates adequately with colleagues.
	The doctor is accessible and reliable.
	Working independently
	Self-confident appearance
	Language skills
Manager	The doctor orders investigations in awareness of costs.
	The doctor manages time effectively and prioritises.
	The doctor shows initiative and assumes responsibility.
	Dealing with stress
Health Advocate	The doctor keeps an eye on patient safety.
Scholar	The doctor likes to teach younger colleagues.
	The doctor is open to feedback and implements it.
	Engaging in research
Professional	

3. Results

3.1. Narrative comments

MSF data included 3024 narrative comments, averaging 2.63 comments per item per resident (between 0 and 12, SD=1.90) and 39.51 comments in sum per resident (between 15 and 66, SD=12.09). Of the narrative comments, 77% were favourable, 15% were unfavourable, 7% could not be allocated; 95% of narrative comments were task-focused, 0,9% trait-focused, and 3,5% could not be allocated.

3.2. Learning goals

3.2.1. Number of learning goals

In total, we analysed 132 separate learning goals, averaging 2.13 (between 0 and 4) learning goals per resident per feedback conversation.

3.2.2. Content of learning goals

Of these 132 goals, 68 could be matched directly to single items on the questionnaire, 64 matched multiple items or referred to topics not mentioned in the MSF question-

naire. These 64 goals were of different quality compared to the competence-based items of the questionnaire and addressed broader concepts that can be understood as referring to Professional Identity Formation. They concerned career planning (n=24), gaining knowledge (n=15), scientific work (n=9), appearing self-confident (n=7), working independently (n=3), and other topics (e.g. language skills, showing presence, dealing with stress; n=6). To integrate as many goals as possible into the calculations, we collated them, despite three overall goals and 24 goals on career planning, to the related CanMEDS roles, see table 3 and figure 1.

3.3. Influencing factors

Our findings are summarised in figure 2.

When we analysed influencing factors, we found significant regression equations ($F(5/1116)=48.3$, $p<0.05$) with an R^2 of 0.17 for topics discussed as strengths ($F(5/1116)=27.5$, $p<0.05$), an R^2 of 0.11 for topics discussed as areas for improvement ($F(7/352)=7.8$, $p<0.05$), and an R^2 of 0.11 for learning goals, see table 3, table 4 and table 5 for regression coefficients.

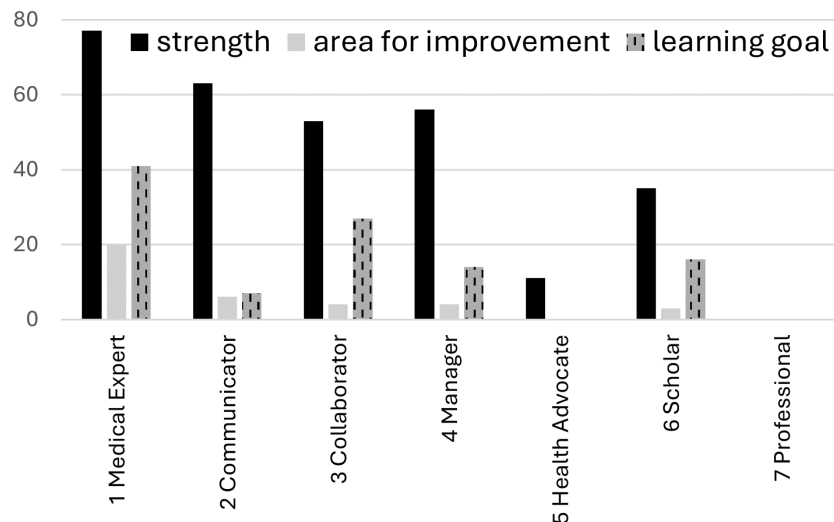


Figure 1: Content of the conversations and learning goals, documented on the structured form for the feedback conversation, dependent on CanMEDS roles

Note: this chart excludes the 3 goals that focused on overall performance and the 24 career planning goals.

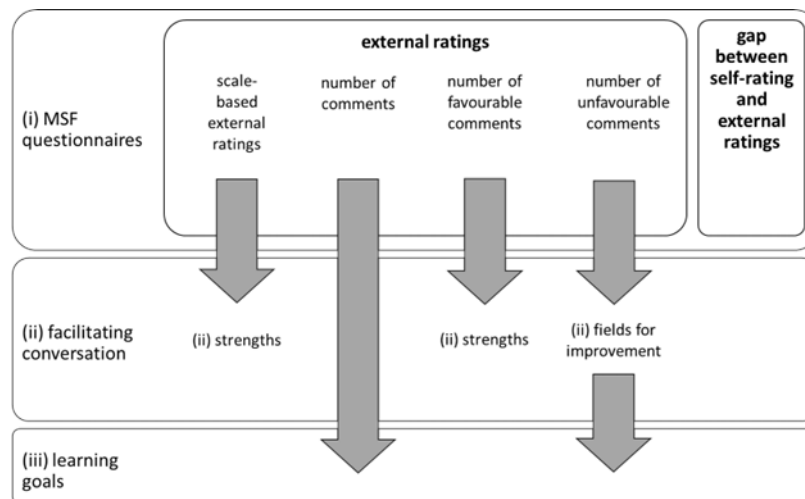


Figure 2: Overview of factors that influence the feedback conversation and the learning goals
Arrows represent significant influence.

Table 4: Influences on the discussion of strengths

Influences on the discussion of strenghts	B	SE (B)	β	t	p
Intercept	-1.22	0.15	0.00	0.05	0.96
Scale-based external rating	0.39	0.04	0.30	8.83	<0.05
Number of comments	0.00	0.02	0.00	0.00	1.00
Number of unfavourable comments	0.00	0.03	0.01	0.13	0.90
Number of favourable comments	0.05	0.02	0.18	2.39	<0.05
Gap between external ratings and self-rating	0.03	0.02	0.04	1.59	0.11

B: unstandardized regression coefficient, SE (B): Standard error of the coefficient,

β (Beta): standardized coefficient, t: t-statistic, p: p-value.

Explanation and interpretation examples: B: The unstandardized regression coefficient shows the direction and size of the effect. SE (B): The standard error of the coefficient indicates the precision of the estimate. β (Beta): The standardized coefficient allows comparison of the relative strength of each predictor. t: The t-statistic is used to test if the coefficient is significantly different from zero. p: The p-value indicates whether the effect is statistically significant ($p < 0.05$ is considered significant).

Examples: Scale-based external rating: Significant positive effect. Higher external ratings are associated with more discussion of strengths. Number of comments: No effect at all. The total number of comments does not influence the discussion of strengths.

Table 5: Influences on the discussion of areas for improvement

Influences on the discussion of areas for improvement	B	SE (B)	β	t	p
Intercept	0.16	0.07	0.00	2.25	<0.05
Scale-based external rating	-0.03	0.02	-0.06	-1.57	0.12
Number of comments	0.02	0.01	0.15	1.80	0.07
Number of unfavourable comments	0.06	0.01	0.23	4.93	<0.05
Number of favourable comments	-0.02	0.01	-0.13	-1.66	0.10
Gap between external ratings and self-rating	-0.01	0.01	-0.04	-1.49	0.14

B: unstandardized regression, SE (B): Standard error of the coefficient,
 β (Beta): standardized coefficient, t: t-statistic, p: p-value

3.3.1. Number of task-focused or trait-focused comments

Of the narrative comments, 95.5% were task-focused; 1% were trait-focused, and 3.5% could not be allocated. Because of this distribution we decided not to add calculations about the differences between the number of task-focused or trait-focused comments to our calculations of the overall number of comments.

3.4. Influences of MSF ratings (part i) on strengths discussed during feedback conversations (part ii)

3.4.1. Scale-based external ratings

Higher scale-based external ratings per item ($B=0.39$, $\beta=0.30$, $p<0.05$) significantly increased the likelihood the topic would be addressed as a strength.

3.4.2. Favourable comments

A higher number of favourable comments per item ($B=0.05$, $\beta=0.18$, $p<0.05$) significantly increased the likelihood the topic would be discussed as a strength. The numbers of overall comments ($B=0.00$, $\beta=0.00$, $p=1.00$), the number of unfavourable comments ($B=0.00$, $\beta=0.01$, $p=0.90$), and the gap between external ratings and self-rating ($B=0.03$, $\beta=0.04$, $p=0.11$) did not have significant influence on whether a topic was discussed as a strength (see table 4).

3.5. Influence of MSF ratings (part i) on topics discussed as areas for improvement during the feedback conversation (part ii)

3.5.1. Unfavourable comments

A higher number of unfavourable comments per item ($B=0.06$, $\beta=0.23$, $p<0.05$) significantly increased the likelihood the topic would be discussed as an area for improvement. Scale-based external ratings ($B=-0.03$, $\beta=-0.06$, $p=0.12$), overall number of comments ($B=0.02$,

$\beta=0.15$, $p=0.07$), the number of favourable comments ($B=-0.02$, $\beta=-0.13$, $p=0.10$), and the gap between external ratings and self-rating ($B=-0.01$, $\beta=-0.04$, $p=0.14$) did not have significant influence on whether a topic was discussed as an area for improvement (see table 5).

3.6. Influences of MSF ratings (part i) and topics discussed (part ii) on learning goals (part iii)

3.6.1. Number of comments

A higher number of comments per item significantly increased ($B=0.04$, $\beta=0.43$, $p<0.05$) the likelihood the topic would be translated into a learning goal.

3.6.2. Discussed as an area for improvement

Discussing a topic as an area for improvement ($B=0.029$, $\beta=0.19$, $p<0.05$) significantly increased the likelihood the topic would be translated into a learning goal. Scale-based external ratings ($B=-0.06$, $\beta=-0.04$, $p=0.52$), the number of favourable comments ($B=-0.03$, $\beta=-0.21$, $p=0.21$), the number of unfavourable comments ($B=-0.01$, $\beta=-0.03$, $p=0.76$), the gap between external ratings and self-rating ($B=0.01$, $\beta=0.01$, $p=0.88$), and discussing a topic as a strength did not significantly influence its translation to a learning goal ($B=0.01$, $\beta=0.01$, $p=0.79$) (see table 6).

4. Discussion

Our investigation revealed that topics discussed as strengths depended on scale-based ratings and the number of favourable comments while areas for improvement depended on the number of unfavourable comments. Gaps between self-ratings and external ratings did not increase the likelihood of discussion. Each learning goal was influenced by the number of narrative comments and whether the topic had been discussed as an area for improvement. Unexpectedly, we found a range of narrative comments and learning goals that exceeded the competences suggested on the MSF questionnaire and related to broader concepts. We see those goals connected to professional identity formation, discuss

Table 6: Influences on learning goals

Influences on learning goals	B	SE (B)	β	t	p
Intercept	0.27	0.34	0.00	0.00	1.00
Scale-based external rating	-0.06	0.10	-0.04	-0.64	0.52
Number comments	0.04	0.02	0.43	2.37	<0.05
Number of favourable comments	-0.03	0.02	-0.21	-1.26	0.21
Number of unfavourable comments	-0.01	0.03	-0.03	-0.30	0.76
Gap between external ratings and self-rating	0.01	0.05	0.01	0.15	0.88
Discussed as strength	0.01	0.05	0.01	0.27	0.79
Discussed as an area for improvement	0.29	0.08	0.19	3.59	<0.05

B: unstandardized regression, SE (B): Standard error of the coefficient, β (Beta): standardized coefficient, t: t-statistic, p: p-value

them under 4.2. and propose implications for practice based on this finding.

4.1. Narrative comments

The high proportion of favourable narrative comments aligns with other studies [23], [25]. We found a higher proportion of task-focused comments [13], [23], likely due to differences in rater training or questionnaire design. Both Dory et al. [26], who studied in-training assessment reports based on written feedback, and Lockyer et al. [23], who studied MSF, indicate that even small adjustments in the format of the questionnaire can increase the quality of feedback. We concluded that the specific rater training and item formulations, asking for task-focused comments, raised the proportion of task-focused comments.

4.2. Learning goals and connection to Professional Identity Formation

We observed between 1 and 4 learning goals per feedback conversation. Many of these learning goals could not be clearly tied to single items of the MSF questionnaire; some addressed combinations of items, and others addressed aspects of the residents' development not covered by the questionnaire, e.g., career plans, engagement in research or personal attitudes like appearing more self-confident. This finding can be interpreted through the lens of professional identity formation theory [5], which explains that during a resident's development, the focus of learning goals might shift from single, separate skills to broader, more personal goals. Thus, MSF conversations with residents might be expanded to encompass both *development* goals and *learning* goals. Our finding that some learning goals seemed unrelated to items on the MSF questionnaire hints at the need to adjust the questionnaire to meet learners' additional needs. These additional learning goals might also indicate gaps in the curriculum. An analysis of learning goals could be used to actively improve the residency curriculum [27], by including overarching goals.

When learning goals exceed the scope of the MSF questionnaire, it points out the complexity of measuring the

effectiveness of MSF since the MSF process addresses both competences and overarching goals. Orienting the feedback conversation to meet these additional goals might not directly increase the effectiveness of MSF as a formative assessment of competences, but expand its range. Embracing the added use might exceed the assessment of competences and might support professional identity development. The MSF could better fulfil this second aim by extending the training of supervisors and residents to encompass the added content and prepare them to better formulate these overarching goals. To see whether this also holds true in an international context, more cross-validation is needed, as MSF is idiosyncratic in every setting, surrounded by different learning cultures including other forms of assessment in different curricula of residency training.

4.3. Influencing factors

Items were more likely to be discussed as strengths with more favourable comments and as an area for improvement with more unfavourable comments. Higher ratings predicted discussion of strengths, but lower ratings did not predict discussion of improvements; only the number of unfavourable comments did. This is in contrast to studies by Sargeant et al. [7] and Overeem et al. [15], who reported low ratings as a main motivator for change. A recent qualitative study [28] described that feedback conversations, perceived as summative, affected the residents' behaviour leading them to conceal weaknesses. A review of the assessor's behaviour [29] found this group also felt this tension, which led them to avoid delivering unpleasant results. This might partially explain our findings that low ratings and the gap between external ratings and self-ratings were not discussed as often as high ratings. Focussing on the formative aspect of feedback might lower this barrier to honest conversation [29]. Also, our study found significant differences between external and self-ratings, as other studies did, but we did not confirm the influence of this difference on conversation or the learning goals [15], [18]. Again, it is possible that supervisors avoided directly pointing out the differences to reduce friction during the feedback conversation. A study of managers by Smither and Walker [13] found that a small number of unfavourable task-focused com-

ments had the strongest effect on performance, but a high number of unfavourable comments was discouraging. The number of written comments in their study averaged 6.8, while we observed even 15 to 66 unfavourable comments per resident. We hypothesise that in our study the supervisors prevented residents from being overwhelmed by too many comments.

Despite the high number of favourable narrative comments, feedback conversations focused more on areas for improvement, possibly because corrective feedback was felt to be more actionable than reinforcing feedback. A model on reflection supports this argument; Sargeant et al. [11] found that feedback inconsistent with self-perception leads to longer lasting reflection; the resident and supervisor might have unconsciously been engaging in this process. Besides, supervisor training in our study focussed on a balance between strengths and areas for improvement, but did not focus on reinforcing goals.

4.4. Implications for practice

4.4.1. Professional Identity Formation

We suggest expanding the MSF questionnaire and participant training to meet the residents' need for overarching development goals. By addressing both specific competences and broader developmental goals, MSF could support professional identity development, besides assessing competences.

4.4.2. Format of the MSF questionnaire and rater training

For the more clearly defined competences, we encourage the use of clear task-related items to achieve high quality feedback with specific task-focused comments.

4.4.3. Feedback conversation

We suggest a formative focus in feedback conversations to encourage honest discussions and reduce tension. When supervisors and residents decide which topics to discuss, they should combine scale-based ratings with narrative comments because both provide valuable information for learning goals.

4.5. Limitations and strengths of the study

We were limited by our inability to observe the conversation directly; we analysed only data going into the feedback conversation, limited notes, and the learning goals as its outcome. Because our study focused in detail on the quantitative data of MSF, we could not study other factors that also influenced the process like the perceptions of the residents or their emotional response to the feedback.

Our unique approach of breaking MSF into components helped us understand the interplay between scale-based

ratings, narrative comments, feedback conversations, and learning goals, elucidating the complex process. Since few studies quantitatively investigated influences on the feedback conversation, and the way feedback was translated into learning goals, our data provides valuable insights. As a quantitative study can only show connections between data, but not explain them, qualitative studies are needed to investigate the residents' and supervisors' perspective on our findings and recommendations.

5. Conclusion

Our study focused on the feedback conversation, a crucial element in the MSF process. We found that a part of the learning goals set during the conversation concerned the professional development of residents. Thus we suggest to use MSF not only as from of a formative assessment of competences but as a guidance process that helps residents develop their professional identity. We encourage supervisors and residents to explicitly include Professional Identity Formation topics such as career plans, engagement in research or personal attitudes into regular assessments and feedback conversations. Our insights offer practical recommendations for improving the use of MSF and shows a way how Professional Identity Formation might be supported during residency training.

Notes

Authorship

The authors Kathrin Neuhaus and Sören Huwendiek share the last authorship.

Authors' ORCIDs

- Eva K. Hennel: [0000-0002-7625-5785]
- Felicitas-M. Lahner: [0000-0001-6697-3698]
- Noemi Zweifel: [0000-0001-7313-3929]
- Sigrid Harendza: [0000-0002-7920-8431]
- Kathrin Neuhaus: [0000-0003-2438-1779]
- Sören Huwendiek: [0000-0001-6116-9633]

Competing interests

The authors declare that they have no competing interests.

References

- Sargeant J. Reflecting upon multisource feedback as 'assessment for learning'. *Perspect Med Educ*. 2015;4(2):55-56. DOI: 10.1007/s40037-015-0175-y
- Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach*. 2007;29(9):855-871. DOI: 10.1080/01421590701775453
- Boursicot K, Etheridge L, Setna Z, Sturrock A, Ker J, Smee S, Sambandam E. Performance in assessment: consensus statement and recommendations from the Ottawa conference. *Med Teach*. 2011;33(5):370-383. DOI: 10.3109/0142159X.2011.565831
- Boursicot K, Kemp S, Wilkinson T, Findyartini A, Canning C, Cilliers F, Fuller R. Performance assessment: Consensus statement and recommendations from the 2020 Ottawa Conference. *Med Teach*. 2021;43(1):58-67. DOI: 10.1080/0142159X.2020.1830052
- Jarvis-Selinger S, Pratt DD, Regehr G. Competency is not enough: integrating identity formation into the medical education discourse. *Acad Med*. 2012;87(9):1185-1190. DOI: 10.1097/ACM.0b013e3182604968
- Overeem K, Lombarts M, Arah OA, Klazinga NS, Grol RP, Wollersheim HC. Three methods of multi-source feedback compared: a plea for narrative comments and coworkers' perspectives. *Med Teach*. 2010;32(2):141-147. DOI: 10.3109/01421590903144128
- Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Challenges in multisource feedback: intended and unintended outcomes. *Med Educ*. 2007;41(6):583-591. DOI: 10.1111/j.1365-2923.2007.02769.x
- Burford B, Illing J, Kergon C, Morrow G, Livingston M. User perceptions of multi-source feedback tools for junior doctors. *Med Educ*. 2010;44(3):165-176. DOI: 10.1111/j.1365-2923.2009.03565.x
- Overeem K, Wollersheim H, Driessen E, Lombarts K, Van De Ven G, Grol R, Arah O. Doctors' perceptions of why 360° degree feedback does (not) work: a qualitative study. *Med Educ*. 2009;43(9):874-882. DOI: 10.1111/j.1365-2923.2009.03439.x
- Sargeant J, Mann KV, Ferrier SN, Langille DB, Muirhead PD, Hayes VM, Sinclair DE. Responses of rural family physicians and their colleague and coworker raters to a multi-source feedback process: a pilot study. *Acad Med*. 2003;78(10 Suppl):S42-S44. DOI: 10.1097/00001888-200310001-00014
- Sargeant JM, Mann KV, van der Vleuten CP, Metsemakers JF. Reflection: a link between receiving and using assessment feedback. *Adv Health Sci Educ Theory Pract*. 2009;14(3):399-410. DOI: 10.1007/s10459-008-9124-4
- Hall J, Atkinson A, Chan M-K, Tourian L, Thoma B, Pattani R. The Clinical Learning Environment in CanMEDS 2025. *Can Med Educ J*. 2023;14(1):41-45. DOI: 10.36834/cmej.75537
- Smither JW, Walker AG. Are the Characteristics of Narrative Comments Related to Improvement in Multirater Feedback Ratings Over Time? *J Appl Psychol*. 2004;89(3):575-581. DOI: 10.1037/0021-9010.89.3.575
- Ferguson J, Wakeling J, Bowie P. Factors influencing the effectiveness of multisource feedback in improving the professional practice of medical doctors: a systematic review. *BMC Med Educ*. 2014;14:76. DOI: 10.1186/1472-6920-14-76
- Overeem K, Wollersheim HC, Arah OA, Cruijsberg JK, Grol RP, Lombarts KM. Factors predicting doctors' reporting of performance change in response to multisource feedback. *BMC Med Educ*. 2012;12:52. DOI: 10.1186/1472-6920-12-52
- Al Khalifa K, Al Ansari A, Violato C, Donnon T. Multisource Feedback to Assess Surgical Practice: A Systematic Review. *J Surg Educ*. 2013;70(4):475-486. DOI: 10.1016/j.jsurg.2013.02.002
- Violato C, Lockyer JM, Fidler H. Changes in performance: a 5 year longitudinal study of participants in a multisource feedback programme. *Med Educ*. 2008;42(10):1007-1013. DOI: 10.1111/j.1365-2923.2008.03127.x
- Wood L, Hassell A, Whitehouse A, Bullock A, Wall D. A literature review of multi-source feedback systems within and without health services, leading to 10 tips for their successful design. *Med Teach*. 2006;28(7):e185-e191. DOI: 10.1080/01421590600834286
- Sargeant J, McNaughton E, Mercer S, Murphy D, Sullivan P, Bruce DA. Providing feedback: Exploring a model (emotion, content, outcomes) for facilitating multisource feedback. *Med Teach*. 2011;33(9):744-749. DOI: 10.3109/0142159X.2011.577287
- Mackillop L, Parker-Swift J, Crossley J. Getting the questions right: non-compound questions are more reliable than compound questions on matched multi-source feedback instruments. *Med Educ*. 2011;45(8):843-848. DOI: 10.1111/j.1365-2923.2011.03996.x
- Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Med Teach*. 2007;29(7):642-647. DOI: 10.1080/01421590701746983
- Hennel EK, Subotic U, Berendonk C, Stricker D, Harendza S, Huwendiek S. A german-language competency-based multisource feedback instrument for residents: development and validity evidence. *BMC Med Educ*. 2020;20(1):357. DOI: 10.1186/s12909-020-02259-2
- Lockyer JM, Sargeant J, Richards SH, Campbell JL, Rivera LA. Multisource Feedback and Narrative Comments: Polarity, Specificity, Actionability, and CanMEDS Roles. *J Contin Educ Health Prof*. 2018;38(1):32-40. DOI: 10.1097/CEH.0000000000000183
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria: R Core Team; 2021. Zugänglich unter/available from: <https://www.R-project.org/>
- Brown JM, Lowe K, Fillingham J, Murphy PN, Bamforth M, Shaw NJ. An investigation into the use of multi-source feedback (MSF) as a work-based assessment tool. *Med Teach*. 2014;36(11):997-1004. DOI: 10.3109/0142159X.2014.909920
- Dory V, Cummings BA, Mondou M, Young M. Nudging clinical supervisors to provide better in-training assessment reports. *Perspect Med Educ*. 2020;9(1):66-70. DOI: 10.1007/s40037-019-00554-3
- Thomas PA, Kern DE, Hughes MT, Chen BY. Curriculum development for medical education: a six-step approach. Baltimore (MA): JHU Press; 2016.
- Huffman BM, Hafferty FW, Bhagra A, Leasure EL, Santivasi WL, Sawatsky AP. Resident impression management within feedback conversations: A qualitative study. *Med Educ*. 2020;55(2):266-274. DOI: 10.1111/medu.14360
- Scarff CE, Bearman M, Chiavaroli N, Trumble S. Keeping mum in clinical supervision: private thoughts and public judgements. *Med Educ*. 2019;53(2):133-142. DOI: 10.1111/medu.13728

Corresponding author:

Eva K. Hennel
Universität Bern, Institut für Medizinische Lehre,
Department for Assessment and Evaluation (AAE),
Mittelstrasse 43, CH-3012 Bern, Switzerland
eva.hennel@faculty.unibe.ch

This article is freely available from
<https://doi.org/10.3205/zma001791>

Received: 2024-11-12
Revised: 2025-06-10
Accepted: 2025-07-28
Published: 2025-11-17

Please cite as

Hennel EK, Lahner FM, Zweifel N, Harendza S, Neuhaus K, Huwendiek S. Multisource feedback in residency training: A quantitative study to investigate the feedback conversation. *GMS J Med Educ.* 2025;42(5):Doc67.
DOI: 10.3205/zma001791, URN: urn:nbn:de:0183-zma0017919

Copyright

©2025 Hennel et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Multisource-Feedback in der ärztlichen Weiterbildung: Eine quantitative Studie zur Untersuchung des Feedbackgesprächs

Zusammenfassung

Einleitung: Multisource Feedback (MSF) ist ein Bewertungsformat in der medizinischen Aus- und Weiterbildung. Es liefert individuelles Feedback auf Grundlage mehrerer Bewertungen, das dann in einem Feedbackgespräch zwischen der Supervisionsperson und der Person in Weiterbildung zur Entwicklung von Lernzielen verwendet werden kann. Wir haben diese Studie durchgeführt, um zu untersuchen, wie Weiterzubildende und Supervisionspersonen die Lernziele festlegen und worauf sich diese Lernziele beziehen.

Methoden: Die Studie umfasste 75 MSF-Datensätze, die pro Person in Weiterbildung aus 12 bis 15 externen Bewertungen und einer Selbstbewertung bestanden. Die Daten umfassten insgesamt 1015 externe Bewertungen und 75 Selbstbewertungen. Da zehn Datensätze keine Lernziele enthielten, konnten nur 65 MSF-Datensätze analysiert werden. Die Datensätze enthielten schriftliches MSF-Feedback, einschliesslich skalenbasierter Bewertungen und narrativer Kommentare sowie strukturierte Protokolle der Feedbackgespräche mit den daraus resultierenden Lernzielen, die nach Themen sortiert wurden. Wir verwendeten multiple lineare Regressionen, um die Zusammenhänge zwischen Feedbackdaten, Gesprächsthemen und Lernzielen zu ermitteln.

Ergebnisse: Themen wurden in den Feedbackgesprächen eher als Stärken diskutiert, wenn die skalenbasierten Bewertungen hoch waren und es viele positive Kommentare zu einem Item gab. Themen wurden eher als Bereiche zur Verbesserung diskutiert, wenn die Anzahl der negativen Kommentare hoch war. Themen mit vielen (positiven und negativen) Kommentaren und Themen, die als verbesserungswürdig diskutiert wurden, führten eher zu Lernzielen. Wir fanden eine Reihe von Lernzielen, die über die Kompetenzen des MSF-Fragebogens hinausgehen und im Zusammenhang mit der Entwicklung der professionellen Identität gesehen werden können.

Fazit: Da das Feedback und die Lernziele über die im MSF-Fragebogen aufgelisteten Kompetenzen hinausgingen, scheint es uns notwendig, diese umfassenderen Themen, die die Entwicklung der Weiterzubildenden betreffen, anzusprechen. Daher empfehlen wir Supervisionspersonen und Weiterzubildenden, Aspekte der Entwicklung der professionellen Identität wie Karrierepläne, Engagement in der Forschung oder persönliche Haltungen ausdrücklich in regelmässige Bewertungen und Feedbackgespräche einzubeziehen. Auf diese Weise könnte MSF ein geeignetes Instrument sein, um die Entwicklung der professionellen Identität zu unterstützen.

Schlüsselwörter: Entwicklung der professionellen Identität, formatives Prüfen, Multisource Feedback, ärztliche Weiterbildung, arbeitsplatzbasiertes Assessment

Eva K. Henkel^{1,2}

Felicitas-M. Lahner^{1,3}

Noemi Zweifel⁴

Sigrid Harendza⁵

Kathrin Neuhaus⁴

Sören Huwendiek¹

1 Universität Bern, Institut für Medizinische Lehre, Abteilung für Assessment und Evaluation (AAE), Bern, Schweiz

2 Schweizerisches Institut für ärztliche Weiter- und Fortbildung SIWF, Bern, Schweiz

3 Berner Fachhochschule, Departement Gesundheit, Bern, Schweiz

4 Universitätskinderspital Zürich, Chirurgische Klinik, Zürich, Schweiz

5 Universitätsklinikum Hamburg-Eppendorf, III. Medizinische Klinik, Hamburg, Deutschland

1. Einleitung

Multisource Feedback (MSF) oder 360° Feedback wird verwendet, um die medizinische Aus- und Weiterbildung zu unterstützen, insbesondere die Entwicklung von Fähigkeiten und Kompetenzen [1], [2], [3], [4]. In diesem Artikel stellen wir dar, wie MSF auch die Entwicklung der professionellen Identität unterstützen kann, wie sie von Jarvis-Selinger beschrieben ist [5]. Im Gegensatz zu anderen arbeitsplatzbasierten Bewertungen umfasst MSF mehrere Perspektiven auf die Leistung einer Person in Weiterbildung in verschiedenen Situationen und über einen längeren Zeitraum. Häufig fasst beim MSF eine Supervisionsperson das Feedback zusammen und „übermittelt“ es der Person in Weiterbildung, einschliesslich aller skalenbasierter Bewertungen und schriftlicher Kommentare. Während dieses Feedbackgesprächs wird das Feedback mit der Selbsteinschätzung der Person in Weiterbildung verglichen und Lernziele definiert, um die Person in Weiterbildung individuell zu begleiten.

Die meisten Studien zu MSF konzentrierten sich auf Feedback, das an Fach- und Oberärzt*innen gegeben wird und untersuchten Faktoren, die die Wirksamkeit und die Ergebnisse von MSF beeinflussen [6], [7], [8], [9], [10], [11]. Obgleich es Empfehlungen gibt, wie in Feedbackgesprächen die emotionalen Reaktionen gesteuert [11] und die Reflexion gefördert werden können [9], haben nur wenige Studien Faktoren in den realen Umgebungen untersucht, in denen diese Gespräche tatsächlich stattfanden [6], [7], [9], [11]. Einige Studien begründeten die Notwendigkeit von Feedbackgesprächen eher mit den Erwartungen von Teilnehmenden als mit ihren Erfahrungen [8], [10], [12]. Keine Studie analysierte den Inhalt oder den Ablauf dieser Gespräche oder beobachtete die Interaktion zwischen Supervidierenden und Weiterzubildenden. Daher ist unbekannt, wie Supervisionspersonen die MSF-Bewertungen filtern und zusammenfassen oder diese zur Erstellung von Lernzielen verwenden. In nicht-medizinischen Bereichen, in denen Feedback schriftlich ohne Gespräch gegeben wurde, wurden die Lernziele durch die Anzahl der narrativen Kommentare, die Polarität der Kommentare (positiv oder negativ) und den Fokus der Kommentare (aufgabenorientiert versus eigenschaftsorientiert) geprägt [13].

Um zu verstehen, wie Lernziele in der Weiterbildung festgelegt werden und worauf sie sich beziehen, sind bessere Kenntnisse über den Inhalt der Feedbackgespräche zwischen der Supervisionsperson und der Person in Weiterbildung erforderlich. In dieser Studie untersuchen wir, wie MSF-Bewertungen die Feedbackgespräche beeinflussen haben und wie beide, die Bewertungen und die Gespräche, die Anzahl und den Inhalt der Lernziele der Weiterzubildenden geprägt haben. Unser Ziel ist es, die Schulung von Supervisionspersonen und Weiterzubildenden anhand dieser Erkenntnisse zu verbessern, um den gezielten Einsatz von MSF zu ermöglichen.

2. Methoden

2.1. Ethik

Die lokale Ethikkommission der Schweizerischen Vereinigung der Forschungsethikkommissionen beurteilte die Studie als nicht bewilligungspflichtig. Alle Teilnehmenden gaben ihre schriftliche Einwilligung zur Teilnahme und zur Veröffentlichung ihrer anonymisierten Daten.

2.2. Setting

Die Studie wurde an der Abteilung für Chirurgie des Universitätskinderspitals Zürich, Schweiz, durchgeführt. MSF wird dort als obligatorischer Bestandteil der fachärztlichen Ausbildung mit formativem Zweck eingesetzt und wurde 2015 auf der Grundlage von Best-Practice-Literatur [6], [7], [9], [14], [15], [16], [17], [18], [19], [20] eingeführt. Alle Teilnehmenden (Weiterzubildende, Bewertende und Supervisionspersonen) wurden über die Ziele, Inhalte und den MSF-Fragebogen einschliesslich der Bewertungsskala und der Feedbackregeln informiert. Die Supervisionspersonen lernten, spezifisches Feedback zu geben und dabei die Anonymität der Bewertenden zu wahren. Der MSF-Fragebogen basierte auf den Can-MEDS-Rollen [21] und wurde bereits ausführlich beschrieben [22].

Der MSF-Prozess in dieser Studie umfasste bis zu 15 externe Bewertende aus vordefinierten Gruppen von Mitarbeitenden, die zweimal jährlich Feedbacks gaben. Die Bewertenden füllten den MSF-Fragebogen online aus, ihr Feedback wurde später anonymisiert und zusammengefasst verwendet. Während des Feedbackgesprächs teilte die Supervisionsperson den Mittelwert der skalenbasierten Bewertungen mit und fasste narrative Kommentare zusammen, die ihr wichtig erschienen. Die Supervisionspersonen und die Weiterzubildenden entschieden jeweils gemeinsam, welche Items des MSF-Fragebogens besprochen werden sollten, legten die Themen für das Feedbackgespräch fest und bestimmten die Anzahl und den Inhalt der Lernziele. Sie dokumentierten die MSF-Bewertungen, Stärken oder Bereiche der Verbesserung, sowie die Lernziele in einem strukturierten Formular.

2.3. Studiendesign und Datenerhebung

Entsprechend der einleitend beschriebenen spezifischen Literaturlücke zum Feedbackgespräch haben wir uns bewusst für ein quantitatives Studiendesign entschieden. Um den Weg der Daten während des Gesprächs so genau wie möglich nachzuvollziehen, haben wir das Gespräch in drei kleinere Teile aufgetrennt, die jeweils dokumentierte, quantifizierbare Daten enthalten.

Die Studie umfasste 75 MSF-Datensätze, die pro Person in Weiterbildung aus 12-15 externen Bewertungen und einer Selbstbewertung bestanden. Die Daten umfassten insgesamt 1015 externe Bewertungen und 75 Selbstbewertungen. Da in zehn Datensätzen die Lernziele fehlten, konnten nur 65 MSF-Datensätze analysiert werden. De-

Tabelle 1: Datenerhebung und -analysen entsprechend der drei in dieser Studie untersuchten Teile vom MSF

	Daten, die bei 75 MSF-Einsätzen erhoben wurden	Kodierung und Berechnung im Rahmen der Analyse
Teil (i) MSF-Fragebögen		
Bestandteile der MSF-Fragebögen	1015 externe Bewertungen: 12 bis 15 skalenbasierte externe Bewertungen pro Person in Weiterbildung, bestehend aus skalenbasierten Bewertungen für jedes Item und narrativen Kommentaren zu jedem Item	Anzahl der narrativen Kommentare, einschliesslich Polarität (positiv/negativ) und Fokus (aufgabenorientiert/haltungsorientiert)
	75 Selbstbewertungen: 1 skalenbasierte Selbstbewertung pro MSF pro Person in Weiterbildung, bestehend aus skalenbasierten Bewertungen für jedes Item und einem narrativen Kommentar zu jedem Item	Diskrepanz zwischen externen Bewertungen und Selbstbewertung
Teil (ii) Themen, die während des Feedbackgesprächs besprochen wurden		
	Liste aller Items mit Angaben dazu, ob sie als Stärke oder Bereich zur Verbesserung diskutiert wurden oder nicht diskutiert wurden	
Teil (iii) Lernziele, die während des Feedbackgesprächs generiert wurden		
	Liste der Lernziele	Lernziele: Anzahl, Formulierung, Zuordnung zu CanMEDS-Rollen

tailliertere Informationen und eine Übersicht über die drei Teile des Feedbackgesprächs sind in Tabelle 1 dargestellt.

2.4. Definition und Berechnung von narrativen Kommentaren, Lernzielen und Einflussfaktoren

Die Datenkodierung orientierte sich an [13] und [23].

2.4.1. Polarität der narrativen Kommentare

Narrative Kommentare, die eindeutig lobend oder bestätigend waren, wurden als „positiv“ kodiert. Korrekturen und kritische Kommentare, einschliesslich solcher in kritischem Ton („manchmal zu motiviert“, „nicht sicher, ob das Wissen ausreicht“), wurden als „negativ“ kodiert. Kommentare, die nicht eindeutig zugeordnet werden konnten, wurden als „nicht zuzuordnen“ kodiert.

2.4.2. Qualität der narrativen Kommentare

Narrative Kommentare wurden ausserdem als *aufgabenorientiert* oder *eigenschaftsorientiert* kodiert. Kommentare, die eindeutig beschrieben, wie eine bestimmte Aufgabe durchgeführt wurde oder durchgeführt werden sollte, wurden als „aufgabenorientiert“ kodiert. Kommentare, die sich direkt auf die Person oder ihre Haltungen bezogen, wurden als „eigenschaftsorientiert“ kodiert. Kommentare, die nicht eindeutig zugeordnet werden konnten, wurden als „nicht zuzuordnen“ kodiert.

2.4.3. Anzahl der Lernziele

Wir haben die Lernziele anhand ihres Inhalts definiert und gezählt. Wir hatten Zugang zu 75 Protokollen von Feedbackgesprächen, haben jedoch 13 davon aus der Analyse der Lernziele ausgeschlossen, da sie keine Notizen zu Lernzielen enthielten (diese Weiterzubildenden befanden sich am Ende ihrer Ausbildung auf der Abteilung). Die Notizen aus zwei weiteren Gesprächen waren verloren, sodass unser Datensatz den Text von 60 der 75 Protokolle umfasste.

2.4.4. Inhalt der Lernziele

Wir überprüften, ob jedes dokumentierte Lernziel sich einem oder mehreren Items des MSF-Fragebogens zuordnen lässt. Ziele, die mit mehr als einem Item verbunden waren (z. B. „Erwerb von Wissen“) oder Ziele, die keinen Bezug zum MSF-Fragebogen zu haben schienen, wurden qualitativ analysiert und separat gezählt. Anschliessend haben wir alle Lernziele den CanMEDS-Rollen zugeordnet (siehe Tabelle 2).

2.4.5. Einflussfaktoren

Wir haben Einflussfaktoren auf der Grundlage von 1015 skalenbasierten externen Bewertungen und 75 skalenbasierten Selbstbewertungen berechnet. Alle 16 Items des MSF-Fragebogens wurden auf einer 5-Punkte-Skala bewertet, von „unter meinen Erwartungen“ (1) bis „weit über meinen Erwartungen“ (5) oder alternativ mit „keine Angabe“ mit Platz für narrative Kommentare direkt nach jedem Punkt.

Tabelle 2: Annahmen für multiple lineare Regressionen

		Modell 1: Diskussion der Stärken	Modell 2: Diskussion von Bereichen der Verbesserungen	Modell 3: Einfluss auf die Lernziele
VIF	Skalenbasierte externe Bewertungen	1.63	1.63	1.79
	Anzahl der Kommentare	8.61	8.61	9.35
	Anzahl negativer Kommentare	2.72	2.72	9.58
	Anzahl positiver Kommentare	8.07	8.07	3.99
	Diskrepanz zwischen externen Bewertungen und Selbstbewertungen	1.11	1.11	1.06
	Als Stärke diskutiert			1.19
	Als Bereich zur Verbesserung diskutiert			1.14
Breusch-Pagan Test	Chi2	118.43	1129.80	26.23
	p	<0.05	<0.05	<0.05
Shapiro-Wilk Test	W	0.91	0.46	0.81
	p	0.09	0.12	0.23

2.5. Regressionsanalysen

Wir verwendeten multiple lineare Regressionen, um den Einfluss der skalenbasierten externen Bewertungen, der Anzahl der Kommentare, der Anzahl der positiven Kommentare, der Anzahl der negativen Kommentare und der Diskrepanz zwischen externen Bewertungen und Selbstbewertungen zu ermitteln. Um zu überprüfen, ob die Annahmen für multiple lineare Regressionen erfüllt waren, verwendeten wir den Breusch-Pagan-Test zur Analyse der Homoskedastizität, den VIF-Wert zur Analyse der Multikollinearität und den Shapiro-Wilk-Test zur Prüfung der Normalverteilung der Residuen. Da alle drei Modelle einen signifikanten Breusch-Pagan-Test auf Heteroskedastizität zeigten, haben wir heteroskedastizitätskonsistente Standardfehler verwendet. In allen drei Modellen lagen die VIF-Werte unter 10, was auf keine Multikollinearität hindeutet. Nach dem Shapiro-Wilk-Test waren die Residuen normalverteilt. Die Ergebnisse der statistischen Analyse sind in Tabelle 2 zu finden.

Wir haben das unstandardisierte Beta (B), den Standardfehler für das unstandardisierte Beta (SE(B)), das standardisierte Beta (β), die t-Teststatistik (t) und den Wahrscheinlichkeitswert (p) dargestellt. Die Regressionsanalysen untersuchten zunächst den Einfluss dieser Variablen auf die Diskussion der Stärken, dann auf die Diskussion der Bereiche zur Verbesserung und schliesslich auf die während des Feedbackgesprächs festgelegten Lernziele. Als Index für die Effektgrösse wurde R^2 verwendet. Das Signifikanzniveau wurde auf $p < 0,05$ festgelegt. Alle Analysen wurden mit R (Version 3.2.0) [24] durchgeführt.

3. Ergebnisse

3.1. Narrative Kommentare

Die MSF-Datensätze umfassten 3024 narrative Kommentare, durchschnittlich 2,63 Kommentare pro Item pro

Person in Weiterbildung (zwischen 0 und 12, $SD=1,90$) und insgesamt 39,51 Kommentare pro Person in Weiterbildung (zwischen 15 und 66, $SD=12,09$). Von den narrativen Kommentaren waren 77% positiv, 15% negativ und 7% konnten nicht zugeordnet werden; 95% der narrativen Kommentare waren aufgabenorientiert, 0,9% eigenchaftsorientiert und 3,5% konnten nicht zugeordnet werden.

3.2. Lernziele

3.2.1. Anzahl der Lernziele

Insgesamt haben wir 132 separate Lernziele analysiert, was einem Durchschnitt von 2,13 (zwischen 0 und 4) Lernzielen pro Person in Weiterbildung und Feedbackgespräch entspricht.

3.2.2. Inhalt der Lernziele

Von diesen 132 Zielen konnten 68 direkt einzelnen Items des MSF-Fragebogens zugeordnet werden, 64 entsprachen mehreren Items oder bezogen sich auf Themen, die im MSF-Fragebogen nicht erwähnt wurden. Diese 64 Ziele wiesen im Vergleich zu den kompetenzbasierten Items des Fragebogens eine andere Qualität auf und bezogen sich auf weiter gefasste Konzepte, die mit Bezug auf die Entwicklung der professionellen Identität verstanden werden können. Sie betrafen die Karriereplanung ($n=24$), den Erwerb von Wissen ($n=15$), wissenschaftliches Arbeiten ($n=9$), selbstbewusstes Auftreten ($n=7$), selbstständiges Arbeiten ($n=3$) und andere Themen (z. B. Sprachkenntnisse, Präsenz zeigen, Umgang mit Stress; $n=6$). Um möglichst viele Ziele in die Berechnungen einbeziehen zu können, haben wir sie, mit Ausnahme von drei übergeordneten Zielen und 24 Zielen zur Karriereplanung, den entsprechenden CanMEDS-Rollen zugeordnet, siehe Tabelle 3 und Abbildung 1.

Tabelle 3: Zuordnung der Items des MSF-Fragebogens und der Lernziele zu den CanMEDS-Rollen. Die CanMEDS-Rollen wurden bewusst nicht ins Deutsche übersetzt

Drei Ziele, die sich auf die Gesamtleistung bezogen, und 24 Ziele, die sich auf die Karriereplanung bezogen, wurden nicht zugeordnet und nicht in die Berechnungen einbezogen.

CanMEDS-Rolle	Entsprechende Punkte des MSF-Fragebogens (beginnend mit „Die ärztliche Fachperson ...“) und/oder Themen der Lernziele
Medical Expert	... stellt korrekte Diagnosen.
	... entwickelt angemessene Behandlungspläne.
	... ist sich der eigenen Grenzen bewusst und bittet in der entsprechenden Situation um Hilfe.
	... verfügt über gute manuelle/technische Fähigkeiten.
	Erwerb von Wissen
Communicator	... führt die Krankengeschichte und Berichte zeitgerecht und präzise.
	... kommuniziert adäquat mit Patienten und Angehörigen.
	... bezieht psychosoziale Aspekte mit ein.
	Präsenz zeigen
Collaborator	... kommuniziert adäquat mit Kollegen.
	... ist erreichbar und zuverlässig.
	Selbständig Arbeiten
	Selbstbewusstes Auftreten
	Sprachkenntnisse
Manager	Die ärztliche Fachperson ordnet Untersuchungen unter Berücksichtigung der Kosten an.
	... hat ein gutes Zeitmanagement und setzt Prioritäten.
	... ist initiativ und übernimmt Verantwortung.
	Umgang mit Stress
Health Advocate	... behält die Patientensicherheit im Blick.
Scholar	... gibt gern Wissen an junge Kollegen weiter.
	... ist offen für Feedback und setzt es um.
	Beteiligung an Forschung
Professional	

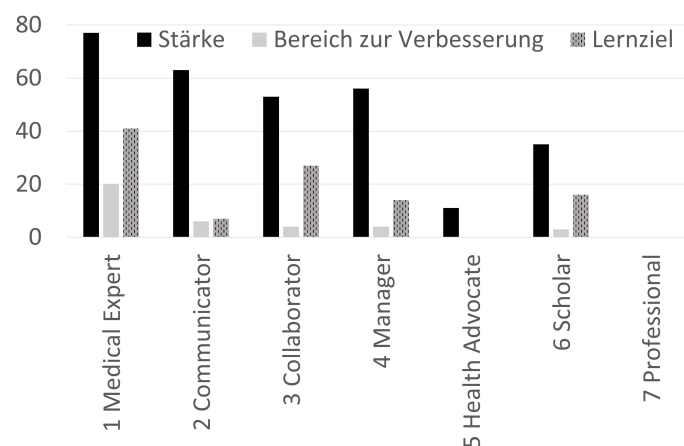


Abbildung 1: Inhalt der Gespräche und Lernziele, dokumentiert auf dem strukturierten Formular für das Feedbackgespräch, abhängig von den CanMEDS-Rollen. Die CanMEDS-Rollen wurden bewusst nicht ins Deutsche übersetzt.

Hinweis: Diese Grafik enthält nicht die 3 Ziele, die sich auf die Gesamtleistung konzentrierten und nicht die 24 Ziele zur Karriereplanung.

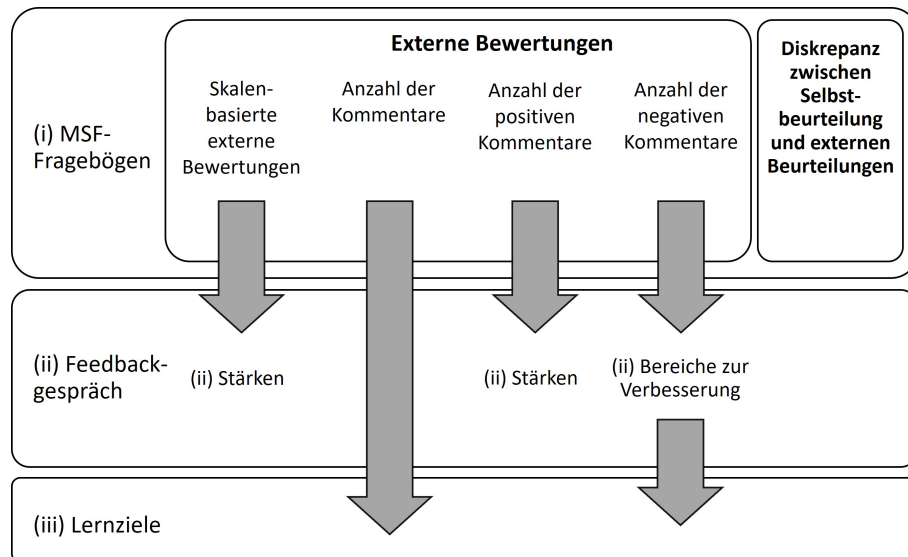


Abbildung 2: Überblick über Faktoren, die das Feedbackgespräch und die Lernziele beeinflussen
Pfeile stehen für einen signifikanten Einfluss.

3.3. Einflussfaktoren

Unsere Ergebnisse sind in Abbildung 2 zusammengefasst. Bei der Analyse der Einflussfaktoren fanden wir signifikante Regressionsgleichungen ($F(5/1116)=48,3$, $p<0,05$) mit einem R^2 von 0,17 für Themen, die als Stärken diskutiert wurden ($F(5/1116)=27,5$, $p<0,05$), einem R^2 von 0,11 für Themen, die als Bereiche der Verbesserung diskutiert wurden ($F(7/352)=7,8$, $p<0,05$), und einem R^2 von 0,11 für Lernziele, siehe Tabelle 3, Tabelle 4 und Tabelle 5 für die Regressionskoeffizienten.

3.3.1. Anzahl der aufgabenorientierten oder eigenschaftsorientierten Kommentare

Von den narrativen Kommentaren waren 95,5% aufgabenorientiert, 1% eigenschaftsorientiert und 3,5% konnten nicht zugeordnet werden. Aufgrund dieser Verteilung haben wir beschlossen, keine Berechnungen über die Unterschiede zwischen der Anzahl der aufgabenorientierten und eigenschaftsorientierten Kommentare zu unseren Berechnungen der Gesamtzahl der Kommentare hinzuzufügen.

3.4. Einflüsse der MSF-Bewertungen (Teil i) auf die während der Feedbackgespräche diskutierten Stärken (Teil ii)

3.4.1. Skalenbasierte externe Bewertungen

Höhere skalierte externe Bewertungen pro Item ($B=0,39$, $\beta=0,30$, $p<0,05$) erhöhten die Wahrscheinlichkeit, dass das Thema als Stärke angesprochen wurde, signifikant.

3.4.2. Positive Kommentare

Eine höhere Anzahl positiver Kommentare pro Item ($B=0,05$, $\beta=0,18$, $p<0,05$) erhöhte die Wahrscheinlichkeit, dass das Thema als Stärke diskutiert wurde, signifikant. Die Anzahl der Gesamtkommentare ($B=0,00$, $\beta=0,00$, $p=1,00$), die Anzahl der negativen Kommentare ($B=0,00$, $\beta=0,01$, $p=0,90$) und die Diskrepanz zwischen externen Bewertungen und Selbstbewertungen ($B=0,03$, $\beta=0,04$, $p=0,11$) hatten keinen signifikanten Einfluss darauf, ob ein Thema als Stärke diskutiert wurde (siehe Tabelle 4).

3.5. Einfluss der MSF-Bewertungen (Teil i) auf die Themen, die während des Feedbackgesprächs als Bereiche zur Verbesserung diskutiert wurden (Teil ii)

3.5.1. Negative Kommentare

Eine höhere Anzahl negativer Kommentare pro Item ($B=0,06$, $\beta=0,23$, $p<0,05$) erhöhte die Wahrscheinlichkeit, dass das Thema als verbesserungswürdig diskutiert wurde, signifikant. Skalenbasierte externe Bewertungen ($B=-0,03$, $\beta=-0,06$, $p=0,12$), die Gesamtzahl der Kommentare ($B=0,02$, $\beta=0,15$, $p=0,07$), die Anzahl der positiven Kommentare ($B=-0,02$, $\beta=-0,13$, $p=0,10$) und die Diskrepanz zwischen externen Bewertungen und Selbstbewertung ($B=-0,01$, $\beta=-0,04$, $p=0,14$) hatten keinen signifikanten Einfluss darauf, ob ein Thema als Bereich der Verbesserung diskutiert wurde (siehe Tabelle 5).

Tabelle 4: Einflüsse auf die Diskussion der Stärken

Einflüsse auf die Diskussion der Stärken	B	SE (B)	β	t	p
Intercept	-1,22	0,15	0,00	0,05	0,96
Skalenbasierte externe Bewertungen	0,39	0,04	0,30	8,83	<0,05
Anzahl der Kommentare	0,00	0,02	0,00	0,00	1,00
Anzahl negativer Kommentare	0,00	0,03	0,01	0,13	0,90
Anzahl positiver Kommentare	0,05	0,02	0,18	2,39	<0,05
Unterschied zwischen externen Bewertungen und Selbstbewertung	0,03	0,02	0,04	1,59	0,11

B: unstandardisierter Regressionskoeffizient, SE (B): Standardfehler des Koeffizienten, β (Beta): standardisierter Koeffizient, t: t-Statistik, p: p-Wert.

Erläuterung und Interpretationsbeispiele: B: Der unstandardisierte Regressionskoeffizient zeigt die Richtung und Grösse des Effekts. SE (B): Der Standardfehler des Koeffizienten gibt die Genauigkeit der Schätzung an. β (Beta): Der standardisierte Koeffizient ermöglicht den Vergleich der relativen Stärke der einzelnen Prädiktoren. t: Die t-Statistik wird verwendet, um zu testen, ob der Koeffizient signifikant von Null abweicht. p: Der p-Wert gibt an, ob der Effekt statistisch signifikant ist (<0,05 gilt als signifikant). Beispiele: Skalenbasierte externe Bewertungen: Signifikanter positiver Effekt. Höhere externe Bewertungen gehen mit einer verstärkten Diskussion der Stärken einher. Anzahl der Kommentare: Gar kein Effekt. Die Gesamtzahl der Kommentare hat keinen Einfluss auf die Diskussion von Stärken.

Tabelle 5: Einflüsse auf die Diskussion von Bereichen zur Verbesserung

Einflüsse auf die Diskussion von Bereichen zur Verbesserung	B	SE (B)	β	t	p
Intercept	0,16	0,07	0,00	2,25	<0,05
Skalenbasierte externe Bewertung	-0,03	0,02	-0,06	-1,57	0,12
Anzahl der Kommentare	0,02	0,01	0,15	1,80	0,07
Anzahl negativer Kommentare	0,06	0,01	0,23	4,93	< 0,05
Anzahl positiver Kommentare	-0,02	0,01	-0,13	-1,66	0,10
Diskrepanz zwischen externen Bewertungen und Selbstbewertung	-0,01	0	-0,04	-1,49	0

B: unstandardisierte Regression, SE (B): Standardfehler des Koeffizienten, β (Beta): standardisierter Koeffizient, t: t-Statistik, p: p-Wert

3.6. Einflüsse der MSF-Bewertungen (Teil i) und der diskutierten Themen (Teil ii) auf die Lernziele (Teil iii)

3.6.1. Anzahl der Kommentare

Eine höhere Anzahl von Kommentaren pro Item erhöhte signifikant ($B=0,04$, $\beta=0,43$, $p<0,05$) die Wahrscheinlichkeit, dass das Thema in ein Lernziel umgesetzt wurde.

3.6.2. Als Bereich zur Verbesserung diskutiert

Die Diskussion eines Themas als Bereich zur Verbesserung ($B=0,029$, $\beta=0,19$, $p<0,05$) erhöhte die Wahrscheinlichkeit, dass das Thema in ein Lernziel umgesetzt wurde, signifikant.

Skalenbasierte externe Bewertungen ($B=-0,06$, $\beta=-0,04$, $p=0,52$), die Anzahl der positiven Kommentare ($B=-0,03$, $\beta=-0,21$, $p=0,21$), die Anzahl der negativen Kommentare ($B=-0,01$, $\beta=-0,03$, $p=0,76$), die Diskrepanz zwischen externen Bewertungen und Selbstbewertung ($B=0,01$, $\beta=0,01$, $p=0,88$) und die Erwähnung eines Themas als Stärke hatten keinen signifikanten Einfluss auf dessen

Übertragung in ein Lernziel ($B=0,01$, $\beta=0,01$, $p=0,79$) (siehe Tabelle 6).

4. Diskussion

Unsere Untersuchung zeigte, dass die als Stärken diskutierten Themen von den skalenbasierten Bewertungen und der Anzahl positiver Kommentare abhingen, während Bereiche zur Verbesserung von der Anzahl negativer Kommentare abhingen. Diskrepanzen zwischen der Selbstbewertung und den externen Bewertungen erhöhten nicht die Wahrscheinlichkeit einer Diskussion. Jedes Lernziel wurde durch die Anzahl der narrativen Kommentare beeinflusst und dadurch, ob das Thema als Bereich zur Verbesserung diskutiert worden war. Überraschend fanden wir eine Reihe von narrativen Kommentaren und Lernzielen, die über die im MSF-Fragebogen vorgeschlagenen Kompetenzen hinausgingen und sich auf weiter gefasste Konzepte bezogen. Wir sehen diese Ziele im Zusammenhang mit der Entwicklung der professionellen Identität, diskutieren sie unter 4.2. und schlagen auf der Grundlage dieser Ergebnisse Implikationen für die Praxis vor.

Tabelle 6: Einflüsse auf Lernziele

Einflüsse auf Lernziele	B	SE (B)	β	t	p
Intercept	0,27	0,34	0,00	0,00	1,00
Skalenbasierte externe Bewertung	-0,06	0,10	-0,04	-0,64	0,52
Anzahl Kommentare	0,04	0,02	0,43	2,37	<0,05
Anzahl positiver Kommentare	-0,03	0,02	-0,21	-1,26	0,21
Anzahl negativer Kommentare	-0,01	0	-0,03	-0,30	0,76
Unterschied zwischen externe Bewertungen und Selbstbewertung	0,01	0,05	0,01	0,15	0,88
Als Stärke diskutiert	0,01	0,05	0,01	0,27	0,79
Als Bereich zur Verbesserung diskutiert	0,29	0,08	0	3,59	<0,05

B: unstandardisierte Regression, SE (B): Standardfehler des Koeffizienten, β (Beta): standardisierter Koeffizient, t: t-Statistik, p: p-Wert

4.1. Narrative Kommentare

Der hohe Anteil an positiven narrativen Kommentaren deckt sich mit anderen Studien [23], [25]. Wir stellten einen höheren Anteil an aufgabenbezogenen Kommentaren fest [13], [23], was wahrscheinlich auf Unterschiede in der Schulung der Bewertenden oder das Fragebogendesign zurückzuführen ist. Sowohl Dory et al. [26], die „in-training assessment reports“ mit schriftlichem Feedback untersuchten, als auch Lockyer et al. [23], die MSF untersuchten, weisen darauf hin, dass bereits kleine Anpassungen im Format des Fragebogens die Qualität des Feedbacks verbessern können. Wir kamen zu dem Schluss, dass die spezifische Schulung der Bewertenden und die Formulierung der Fragen, die zu aufgabenbezogenen Kommentaren aufforderten, den Anteil der aufgabenbezogenen Kommentare erhöhten.

4.2. Lernziele und Zusammenhang mit der Entwicklung der professionellen Identität

Wir haben zwischen 1 und 4 Lernziele pro Feedbackgespräch beobachtet. Viele dieser Lernziele konnten nicht eindeutig mit einzelnen Items des MSF-Fragebogens in Verbindung gebracht werden; einige bezogen sich auf Kombinationen von Items, andere auf Aspekte der Entwicklung der Weiterzubildenden, die im Fragebogen nicht abgedeckt waren, z. B. Karrierepläne, Engagement in der Forschung oder persönliche Haltungen wie ein selbstbewussteres Auftreten. Dieser Fund kann durch die Theorie der Professional Identity Formation [5] betrachtet werden, welche erklärt, dass sich der Schwerpunkt der Lernziele während der Entwicklung einer Person in Weiterbildung von einzelnen, separaten Fähigkeiten zu breiteren, persönlicheren Zielen verlagern kann. Daher schlagen wir vor, die MSF-Gespräche mit Weiterzubildenden zu erweitern, um sowohl Entwicklungsziele als auch Lernziele zu berücksichtigen.

Unsere Erkenntnis, dass einige Lernziele offenbar keinen Bezug zu den Fragen des MSF-Fragebogens hatten, deutet darauf hin, dass der Fragebogen angepasst werden muss, um den zusätzlichen Bedürfnissen der Lernen-

den gerecht zu werden. Diese zusätzlichen Lernziele könnten auch auf Lücken im Curriculum hinweisen. Eine Analyse der Lernziele könnte daher genutzt werden, um das Weiterbildungscurriculum zu verbessern [27], indem übergeordnete Ziele aufgenommen werden.

Wenn die Lernziele über den Umfang des MSF-Fragebogens hinausgehen, wird deutlich, wie komplex die Messung der Wirksamkeit von MSF ist, da der MSF-Prozess sowohl Kompetenzen als auch übergeordnete Ziele berücksichtigt. Die Ausrichtung des Feedbackgesprächs auf diese zusätzlichen Ziele erhöht möglicherweise nicht direkt die Wirksamkeit von MSF als formative Kompetenzbewertung, erweitert jedoch deren Anwendungsbereich. Diese zusätzlichen Funktion könnte über die Bewertung von Kompetenzen hinausführen und die professionelle Identitätsentwicklung unterstützen. MSF könnte letzteres Ziel besser erfüllen, indem die Schulung von Supervisionspersonen und Weiterzubildenden um die zusätzlichen Inhalte erweitert wird und sie darauf vorbereitet werden, die übergeordneten Ziele besser zu formulieren. Um zu sehen, ob dies auch in einem internationalen Kontext gilt, sind weitere Untersuchungen erforderlich, da das MSF kontextabhängig ist, geprägt von unterschiedlichen Lernkulturen, einschliesslich anderer Formen der Bewertung in unterschiedlichen Weiterbildungscurricula.

4.3. Einflussfaktoren

Items wurden eher als Stärken diskutiert, wenn sie mehr positive Kommentare hatten und als Bereiche zur Verbesserung, wenn sie negativen Kommentare hatten. Höhere Bewertungen liessen eine Diskussion über Stärken erwarten, niedrigere Bewertungen hingegen liessen keine Diskussion über Bereiche zur Verbesserung erwarten; nur die Anzahl der negativen Kommentare war ein Indikator für eine Diskussion über Verbesserungen. Dies steht im Gegensatz zu Studien von Sargeant et al. [7] und Overeem et al. [15], die niedrige Bewertungen als Hauptmotiv für Veränderungen nannten. Eine aktuelle qualitative Studie [28] beschrieb, dass Feedback-Gespräche, die als summativ empfunden wurden, das Verhalten der Weiterzubildenden beeinflussten und sie dazu veranlassten, Schwächen zu verbergen. Ein Literatur-Review zum Verhalten der Beurteilenden [29] fasste zusammen, dass

auch diese Gruppe diese Spannung empfand, was sie dazu veranlasste, unangenehme Information zu verschweigen. Dies könnte teilweise unsere Ergebnisse erklären, dass niedrige Bewertungen und die Diskrepanz zwischen externen Bewertungen und Selbstbewertungen nicht so häufig diskutiert wurden wie hohe Bewertungen. Eine Fokussierung auf den formativen Aspekt des Feedbacks könnte diese Barriere für ehrliche Gespräche verkleinern [29].

Ausserdem fanden wir, wie andere Studien, signifikante Unterschiede zwischen externen Bewertungen und Selbstbewertungen, konnten jedoch den beschriebenen Einfluss dieser Unterschiede auf die Gespräche oder die Lernziele nicht bestätigen [15], [18]. Auch hier ist es möglich, dass Supervisionspersonen es vermieden, direkt auf die Unterschiede hinzuweisen, um Spannungen während des Feedbackgesprächs zu vermeiden.

Eine Studie von Smither und Walker [13] mit Führungskräften ergab, dass eine kleine Anzahl negativer aufgabenbezogener Kommentare den stärksten Einfluss auf die Leistung hatte, während eine hohe Anzahl negativer Kommentare demotivierend wirkte. Die Anzahl der schriftlichen Kommentare lag in ihrer Studie bei durchschnittlich 6,8, während wir sogar 15 bis 66 negative Kommentare pro Person in Weiterbildung beobachteten. Wir vermuten, dass die Supervisionspersonen in unserer Studie verhindert haben, die Weiterzubildenden durch zu viele Kommentare zu überfordern.

Trotz der hohen Anzahl positiver narrativer Kommentare konzentrierten sich die Feedback-Gespräche eher auf Bereiche zur Verbesserung, möglicherweise weil korrigierendes Feedback als umsetzbarer empfunden wurde als bestätigendes Feedback. Ein Reflexionsmodell stützt diese Argumentation: Sargeant et al. [11] fanden heraus, dass Feedback, das nicht mit der Selbstwahrnehmung übereinstimmt, zu einer länger anhaltenden Reflexion führt; möglicherweise haben Weiterzubildende und der Supervisionspersonen unbewusst diesen Prozess berücksichtigt. Ausserdem wurde in unserer Studie bei der Schulung der Supervisionspersonen betont, dass Stärken und Bereiche zur Verbesserung im Gleichgewicht diskutiert werden sollten, jedoch nicht geschult, Ziele zu formulieren, die das bestehende Verhalten bestärken.

4.4. Implikationen für die Praxis

4.4.1. Entwicklung der professionellen Identität

Wir schlagen vor, den MSF-Fragebogen und die Schulung zu erweitern, um dem Bedürfnis der Weiterzubildenden nach übergeordneten Entwicklungszielen gerecht zu werden. Durch die Berücksichtigung sowohl spezifischer Kompetenzen als auch allgemeinerer Entwicklungsziele könnte MSF neben der Bewertung von Kompetenzen auch die Entwicklung der professionellen Identität unterstützen.

4.4.2. Format des MSF-Fragebogens und Schulung der Bewertenden

Für die klar definierten Kompetenzen empfehlen wir die Verwendung aufgabenbezogener Items, um ein qualitativ hochwertiges Feedback mit spezifischen, aufgabenorientierten Kommentaren zu erzielen.

4.4.3. Feedbackgespräch

Wir empfehlen einen formativen Fokus in Feedbackgesprächen, um ehrliche Diskussionen zu fördern und Spannungen abzubauen. Wenn Supervisionspersonen und Weiterzubildende entscheiden, welche Themen besprochen werden sollen, sollten sie skalenbasierte Bewertungen mit narrativen Kommentaren kombinieren, da beide wertvolle Informationen für die Lernziele liefern.

4.5. Limitationen und Stärken der Studie

Wir waren dadurch eingeschränkt, dass wir die Gespräche nicht direkt beobachten konnten; wir analysierten nur die Daten, die in das Feedback-Gespräch einfließen, begrenzte Notizen und die Lernziele als Ergebnis. Da sich unsere Studie detailliert auf die quantitativen Daten des MSF konzentrierte, konnten wir andere Faktoren, die den Prozess ebenfalls beeinflussten, wie die Wahrnehmung der Weiterzubildenden oder ihre emotionale Reaktion auf das Feedback, nicht untersuchen.

Unser einzigartiger Ansatz, den MSF-Prozess in einzelne Teile zu betrachten, half uns, die Wechselwirkungen zwischen skalenbasierten Bewertungen, narrativen Kommentaren, Feedback-Gesprächen und Lernzielen zu verstehen und so den komplexen Prozess zu beleuchten. Da nur wenige Studien die Einflüsse auf das Feedbackgespräch und die Art und Weise, wie Feedback in Lernziele umgesetzt wurde, quantitativ untersucht haben, liefern unsere Daten wertvolle Erkenntnisse. Da eine quantitative Studie nur Zusammenhänge zwischen Daten aufzeigen, aber nicht erklären kann, sind qualitative Studien erforderlich, um die Perspektive der Weiterzubildenden und Supervisionspersonen zu unseren Ergebnissen und Empfehlungen zu untersuchen.

5. Fazit

Unsere Studie konzentrierte sich auf das Feedbackgespräch, ein entscheidendes Element im MSF-Prozess. Wir haben festgestellt, dass ein Teil der während des Gesprächs festgelegten Lernziele die professionelle Entwicklung der Weiterzubildenden betraf. Daher schlagen wir vor, MSF nicht nur als formatives Bewertungsformat zu nutzen, sondern als Begleitmassnahme, welche Weiterzubildende bei der Entwicklung der professionellen Identität unterstützt. Wir ermutigen Supervisionspersonen und Weiterzubildende, Themen der Entwicklung der professionellen Identität, wie Karrierepläne, Engagement in der Forschung oder persönliche Haltungen, ausdrücklich

in regelmäßige Bewertungen und Feedbackgespräche einzubeziehen. Unsere Erkenntnisse bieten praktische Empfehlungen zur Verbesserung der Nutzung von MSF und zeigen einen Weg auf, wie die Entwicklung der professionellen Identität während der ärztlichen Weiterbildung unterstützt werden kann.

Anmerkungen

Autorenschaft

Die Autor*in Kathrin Neuhaus und Sören Huwendiek teilen sich die Letztautorenschaft.

ORCIDs der Autor*innen

- Eva K. Hennel: [0000-0002-7625-5785]
- Felicitas-M. Lahner: [0000-0001-6697-3698]
- Noemi Zweifel: [0000-0001-7313-3929]
- Sigrid Harendza: [0000-0002-7920-8431]
- Kathrin Neuhaus: [0000-0003-2438-1779]
- Sören Huwendiek: [0000-0001-6116-9633]

Interessenkonflikt

Die Autor*innen erklären, dass sie keinen Interessenkonflikt im Zusammenhang mit diesem Artikel haben.

Literatur

1. Sargeant J. Reflecting upon multisource feedback as 'assessment for learning'. *Perspect Med Educ*. 2015;4(2):55-56. DOI: 10.1007/s40037-015-0175-y
2. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Med Teach*. 2007;29(9):855-871. DOI: 10.1080/01421590701775453
3. Boursicot K, Etheridge L, Setna Z, Sturrock A, Ker J, Smee S, Sambandam E. Performance in assessment: consensus statement and recommendations from the Ottawa conference. *Med Teach*. 2011;33(5):370-383. DOI: 10.3109/0142159X.2011.565831
4. Boursicot K, Kemp S, Wilkinson T, Findyartini A, Canning C, Cilliers F, Fuller R. Performance assessment: Consensus statement and recommendations from the 2020 Ottawa Conference. *Med Teach*. 2021;43(1):58-67. DOI: 10.1080/0142159X.2020.1830052
5. Jarvis-Selinger S, Pratt DD, Regehr G. Competency is not enough: integrating identity formation into the medical education discourse. *Acad Med*. 2012;87(9):1185-1190. DOI: 10.1097/ACM.0b013e3182604968
6. Overeem K, Lombarts M, Arah OA, Klazinga NS, Grol RP, Wollersheim HC. Three methods of multi-source feedback compared: a plea for narrative comments and coworkers' perspectives. *Med Teach*. 2010;32(2):141-147. DOI: 10.3109/01421590903144128
7. Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Challenges in multisource feedback: intended and unintended outcomes. *Med Educ*. 2007;41(6):583-591. DOI: 10.1111/j.1365-2923.2007.02769.x
8. Burford B, Illing J, Kergon C, Morrow G, Livingston M. User perceptions of multi-source feedback tools for junior doctors. *Med Educ*. 2010;44(3):165-176. DOI: 10.1111/j.1365-2923.2009.03565.x
9. Overeem K, Wollersheim H, Driessen E, Lombarts K, Van De Ven G, Grol R, Arah O. Doctors' perceptions of why 360° degree feedback does (not) work: a qualitative study. *Med Educ*. 2009;43(9):874-882. DOI: 10.1111/j.1365-2923.2009.03439.x
10. Sargeant J, Mann KV, Ferrier SN, Langille DB, Muirhead PD, Hayes VM, Sinclair DE. Responses of rural family physicians and their colleague and coworker raters to a multi-source feedback process: a pilot study. *Acad Med*. 2003;78(10 Suppl):S42-S44. DOI: 10.1097/00001888-200310001-00014
11. Sargeant JM, Mann KV, van der Vleuten CP, Metsemakers JF. Reflection: a link between receiving and using assessment feedback. *Adv Health Sci Educ Theory Pract*. 2009;14(3):399-410. DOI: 10.1007/s10459-008-9124-4
12. Hall J, Atkinson A, Chan M-K, Tourian L, Thoma B, Pattani R. The Clinical Learning Environment in CanMEDS 2025. *Can Med Educ J*. 2023;14(1):41-45. DOI: 10.36834/cmej.75537
13. Smither JW, Walker AG. Are the Characteristics of Narrative Comments Related to Improvement in Multirater Feedback Ratings Over Time? *J Appl Psychol*. 2004;89(3):575-581. DOI: 10.1037/0021-9010.89.3.575
14. Ferguson J, Wakeling J, Bowie P. Factors influencing the effectiveness of multisource feedback in improving the professional practice of medical doctors: a systematic review. *BMC Med Educ*. 2014;14:76. DOI: 10.1186/1472-6920-14-76
15. Overeem K, Wollersheim HC, Arah OA, Cruisberg JK, Grol RP, Lombarts KM. Factors predicting doctors' reporting of performance change in response to multisource feedback. *BMC Med Educ*. 2012;12:52. DOI: 10.1186/1472-6920-12-52
16. Al Khalifa K, Al Ansari A, Violato C, Donnon T. Multisource Feedback to Assess Surgical Practice: A Systematic Review. *J Surg Educ*. 2013;70(4):475-486. DOI: 10.1016/j.jsurg.2013.02.002
17. Violato C, Lockyer JM, Fidler H. Changes in performance: a 5 year longitudinal study of participants in a multisource feedback programme. *Med Educ*. 2008;42(10):1007-1013. DOI: 10.1111/j.1365-2923.2008.03127.x
18. Wood L, Hassell A, Whitehouse A, Bullock A, Wall D. A literature review of multi-source feedback systems within and without health services, leading to 10 tips for their successful design. *Med Teach*. 2006;28(7):e185-e191. DOI: 10.1080/01421590600834286
19. Sargeant J, McNaughton E, Mercer S, Murphy D, Sullivan P, Bruce DA. Providing feedback: Exploring a model (emotion, content, outcomes) for facilitating multisource feedback. *Med Teach*. 2011;33(9):744-749. DOI: 10.3109/0142159X.2011.577287
20. Mackillop L, Parker-Swift J, Crossley J. Getting the questions right: non-compound questions are more reliable than compound questions on matched multi-source feedback instruments. *Med Educ*. 2011;45(8):843-848. DOI: 10.1111/j.1365-2923.2011.03996.x
21. Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based framework of physician competencies. *Med Teach*. 2007;29(7):642-647. DOI: 10.1080/01421590701746983
22. Hennel EK, Subotic U, Berendonk C, Stricker D, Harendza S, Huwendiek S. A german-language competency-based multisource feedback instrument for residents: development and validity evidence. *BMC Med Educ*. 2020;20(1):357. DOI: 10.1186/s12909-020-02259-2

23. Lockyer JM, Sargeant J, Richards SH, Campbell JL, Rivera LA. Multisource Feedback and Narrative Comments: Polarity, Specificity, Actionability, and CanMEDS Roles. *J Contin Educ Health Prof.* 2018;38(1):32-40. DOI: 10.1097/CEH.0000000000000183
24. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria: R Core Team; 2021. Zugänglich unter/available from: <https://www.R-project.org/>
25. Brown JM, Lowe K, Fillingham J, Murphy PN, Bamforth M, Shaw NJ. An investigation into the use of multi-source feedback (MSF) as a work-based assessment tool. *Med Teach.* 2014;36(11):997-1004. DOI: 10.3109/0142159X.2014.909920
26. Dory V, Cummings BA, Mondou M, Young M. Nudging clinical supervisors to provide better in-training assessment reports. *Perspect Med Educ.* 2020;9(1):66-70. DOI: 10.1007/s40037-019-00554-3
27. Thomas PA, Kern DE, Hughes MT, Chen BY. Curriculum development for medical education: a six-step approach. Baltimore (MA): JHU Press; 2016.
28. Huffman BM, Hafferty FW, Bhagra A, Leasure EL, Santivasi WL, Sawatsky AP. Resident impression management within feedback conversations: A qualitative study. *Med Educ.* 2020;55(2):266-274. DOI: 10.1111/medu.14360
29. Scarff CE, Bearman M, Chiavaroli N, Trumble S. Keeping mum in clinical supervision: private thoughts and public judgements. *Med Educ.* 2019;53(2):133-142. DOI: 10.1111/medu.13728

Korrespondenzadresse:

Eva K. Hennel
Universität Bern, Institut für Medizinische Lehre, Abteilung für Assessment und Evaluation (AAE), Mittelstr. 43, CH-3012 Bern, Schweiz
eva.hennel@faculty.unibe.ch

Bitte zitieren als

Hennel EK, Lahner FM, Zweifel N, Harendza S, Neuhaus K, Huwendiek S. Multisource feedback in residency training: A quantitative study to investigate the feedback conversation. *GMS J Med Educ.* 2025;42(5):Doc67.
DOI: 10.3205/zma001791, URN: urn:nbn:de:0183-zma0017919

Artikel online frei zugänglich unter

<https://doi.org/10.3205/zma001791>

Eingereicht: 12.11.2024

Überarbeitet: 10.06.2025

Angenommen: 28.07.2025

Veröffentlicht: 17.11.2025

Copyright

©2025 Hennel et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.