Supplemental Digital Appendix 1

Summary of 51 Articles Included in an Integrative Review of the Literature on the Content of Feedback to Learners in Medical Education, 1980-2015

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Ferguson, 2010[21]	Audiotapes	Speech-pathology students	None	• Faculty did most of the talking, focused on behaviors. • Positive appraisals were explicit and negative appraisals implicit.
Hasley, 2009[22]	Audiotapes	Internal medicine residents and students	American Board of Internal Medicine evaluation form	• 86% of feedback sessions included general, positive statements, with a mean of 7 statements per session. • 41% of the time the learner was given improvement action plan. • Medical students received more positive statements than residents about their performance. • Faculty often did not engage learners in an interactive manner, and did not ask learners to discuss the learners' self-assessment
Spanager, 2015[23]	Audiotapes	Surgery 1st-3rd year residents	NOTSSdk (Non-Technical Skills for Surgeons in Denmark)	• Conversations lasted a median of 8 mins (2-15). • In few conversations (1 out of 8) were learning goals set (usually done by surgeon and not resident). • Conversations often ended by surgeon checking if resident understood feedback or reinforcing positive performance. • 47% of comments based on surgeons' "frames" (i.e., how they view the world) vs. 20% from residents' frames.

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Wen, 2015[24]	Audiotapes	5th-year medical students	Two-hour group discussion with tutor and peers	• 6 types of feedback from tutors, with exploring new knowledge about psychosocial issues most common (25.2%). • 8 types of feedback from peers, mostly focused on discussing psychosocial issues and action plans. • Tutor feedback focused on varied feedback types, whereas peer feedback was more limited. • Peers engaged in few confrontations or debates with each other.

Bok, 2016[25]	Clinical examination (CEX)	Veterinary students	Narrative feedback from mini-CEX form documented in digital portfolio	• 3 interrelated factors influenced teachers' use of the mini-CEX
				  ○ personal teacher
				  ○ context
				  ○ teacher-student
				• Teachers reluctant to document negative feedback in the mini-CEX.

Fernando, 2008[26]	CEX	Year 5 (final year) medical students	Mini-CEX form	• 5% of students failed to have any CEX encounters; 16% had only 1 encounter. Only 41% completed the required 3 evaluative encounters. • 21.2% had identical scores [i.e., no range]; only 1.3% had a range of 3 [out of 6]. On a 7-point scale, almost all rankings were 5,6,7 (and so 1-4 were merged). • 22.7% - no positive aspects were noted; 28.2% no suggestions for improvement; 49.7% no action plan. • Residents more likely than faculty to identify positive aspects, offer suggestions for improvement, and record action plans.

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Gauthier, 2015[27]	CEX	Endocrinology fellows	CEX form	• In 70% of 255 CEX evaluations, only a single element of deliberate practice noted (i.e., Task, Gap, or Action) • 56% -specific Task score; 3.9% specific Gap score; 13.7% specific Action score
Harvey, 2013[28]	CEX	2nd-year medical students	"Modified mini-CEX"	• Clinical supervisors underwent training on feedback strategies and use of the mini-CEX assessment tool • 60% of 1,000 records had no written feedback comments. • Structural analysis: significant variation; poor flow of info. • Content: 20% of statements did not even relate to student performance; for the remaining 80% of comments, 84% affirmed student competence, only 16% had goals for improvement.
Holmboe, 2004[29]	CEX	Internal medicine interns	Mini-CEX form	• Faculty received training in use of the Mini-CEX form • 0 to 9 recommendations per feedback session (mean 1.9). 20% had no recommendations. Only 1-2 % of recommendations for medical knowledge or professionalism • 61% of session's faculty asked for intern reactions; 34% involved self-assessments; 11% involved an action plan.
Kroboth, 1996[30]	CEX	Interns	CEX form	• 984 teaching points, 13.5/session. 48% of these were noted on Evaluator (EV)-Postfeedback Form (PFF); interns recalled hearing 46% of points on EV-PFF. • Interns recalled 75 points not on CEX form of EV-PFF. • 9.4% of CEX comments positive. 12.5% EV-PFF positive comments; Interns recalled only 30 positive comments [8.7%]. • Only ~75% of forms were completed. Interns only heard 25% of feedback on physical exam.

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Pelgrim, 2012[31]	CEX	1st- and 3rd-year postgraduate trainees	Mini-CEX form	• More comments for feedback [87-92%], less comments for trainee self-reflection [53%], and few action plan comments [3-34%]. • 57% of all comments were specific, <10% not specific. • Large variability between faculty-trainee pairs in specific comments. • 32% of evaluations showed specific reflection and specific feedback; the same percentage of evaluations showed *no* specific reflection and *no* specific feedback.
Playford, 2013[32]	CEX	Medical students in longitudinal integrated clerkship	Mini-CEX form	• More senior faculty gave lower ratings. • Monthly analysis showed progressive improvement.
Bandiera, 2008[33]	Feedback cards	PGY-1 residents in all specialties	Daily Encounter Cards (DECs)	• Only 1.3% of DECs said "needs attention." • 33/43 [73%] of faculty did not choose needs attention. • No feedback on Communicator, Collaborator, Professional roles.
Donata, 2015[34]	Feedback cards	Internal medicine residents	Minicard: 4 sections [history, physical exam, presentation, counseling]; 3 domains [knowledge, communication, professionalism]; 4 scoring levels	• 56% PGY1s were rated Good, 8% Marginal; 67% PGY3s Excellent, 2% Marginal. • Action plans: 50% action-oriented, 11% observational FB, 9% minimal feedback. • 30% of cards had no Action Plan. • 74% of encounters indicated verbal feedback given.

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Johnston, 2008[35]	Feedback cards	Internal medicine clerkship students	Structured Observation of Clinical Skills [SOCS] pocket card. History and physical exam focus. Observed behaviors on one side; 2 behaviors done well and 2 needing improvement on other side	• 10% cards were illegible. • 56% had documentation every rating field; 44% were incomplete. • 46% included the 4 requested comments; 54% were incomplete. • 92% of SOCS had general comments, 62% had specific behavioral comments. • 97% had praise. • 78% had advice for improvement, 44% of which had specific behaviors. • Students thought feedback sometimes was too general or brief, but feedback was timely and appreciated.
Johnston, 2008[36]	Feedback cards	Internal medicine clerkship students	Structured Observation of Clinical Skills [SOCS] pocket card. History and physical exam focus. Observed behaviors on one side; 2 behaviors done well and 2 needing improvement on other side	• No differences by gender of feedback provider • Female students received less advice with action plans for improvements [75 vs 90%]. • Fewer recommendations for improvement with gender concordant pairs. • 23% of cards not completed.

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Schum, 2003[37]	Feedback cards	Internal medicine and pediatrics residents, 3rd- and 4th-year medical students	Feedback "note" had 2 preprinted sections--"well done" and "needs improvement"	• 770 notes with 1,607 comments, but 2 faculty provided 73% of the comments. • There were more resident comments [75%] than student comments. • "Well done" noted 69%. • Based on their specificity coding, high degree of specificity for both Well done and Needs Improvement comments. • Specificity frequency increased going from PGY1 to PGY3 years. • The most commonly used content areas for the ''needs improvement'' comments were documentation (n=161, 33%) and didactic information (n =102, 21%). The content areas also were specific with 96% and 92%, respectively. Comments receiving the fewest feedback comments were communication and 'patient relations'
Sokol-Hessner, 2010[38]	Feedback cards	Clerkship students, all disciplines.	Cards had competency checklists, and a space for comments labeled "Action plan"	• 19% did not have a Comment; 3% were unintelligible. • Comments were brief [mean 10 words]; 1.2 action plans per card. • Feedback was positive 96%.
Bullock, 2009[39]	Multisource feedback (MSF)	Senior house officers and family practice physicians	Team Assessment of Behavior form (TAB)	• Only 6% of forms had "concerns." • Consultants expressed more concerns than peers, administrators or managers (i.e., "hawkish" behavior 3-4x more likely by consultants).

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Canavan, 2010[40]	MSF	Residents and fellows (6 disciplines)	Assessment of Professional Behaviors (APB) as part of National Board of Medical Examiners. Free text possible	• 74.5% non-behavioral/global comments. • 90.1% positive, 10.3% negative. • 41.1% general behavior of learner; 24.8% specific behavior. • 7.4% offered specific strategy for improvement. • 7.8% remarked inability to rate learner. • More comments were given to PGY1s than more senior year trainees.
Hayward, 2014[41]	MSF	Residents (Internal Medicine, OB/GYN, Neurology, Orthopedics)	ICAR (Interprofessional Collaborator Assessment Rubric [17 items with 9-point scale])	• Missing data decreased from 13.1% to 8.8% using daily assessments. • High internal consistency [Cronbach alpha 0.981]. • No significant differences between 3 rater groups (physicians, nurses, allied health). • Female raters scored residents lower than male raters.
Lockyer, 2002[42]	MSF	Practicing surgeons and Family Medicine physicians	Based on CanMEDS	• Surgeons more likely to over-rate themselves. • More than 70% contemplated change with feedback, but only 68% FM physicians and 27% surgeons initiated change.
Ogunyemi, 2009[43]	MSF	OB/GYN residents	Internally "validated" multisource feedback survey (4-point scale on 3 measures [interpersonal communication with patients; interactions with peers and staff; professionalism])	• Ratings on 3 measures ranged from 3.19 to 3.5. • As residents progressed, there were more negative evaluations. • Male residents had more negative evaluations by nurses (who were more likely female) than did female residents; for faculty, variable gender differences depending on which measure. • Residents on OB service had more negative evaluations than GYN service.

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Qu, 2012[44]	MSF	Residents	15-26 items depending on respondent, 5-point scale, no text	• Cronbach alpha >0.9; factor analysis accounted for 70-74% of total variance. • Most items rated >4.0
Sargeant, 2003[45]	MSF	Rural Family Medicine physicians	Physician Achievement Review (PAR)	• High mean PAR scores [>4.0 for 85 of 88 items]. • Physicians thought the review process was helpful, and thought the patient feedback was most appropriate. • Peers and coworkers who knew the physicians well, tended to rate them higher. • Most (89%) of physicians reported that feedback was useful and 61% planned to make a change based on the feedback. • Communication was the most common area for feedback.
Whitehouse, 2007[46]	MSF	Senior House Officers (SHOs)	Team Assessment of Behavior form (TAB)	• Only 94 of 171 learners [60%] received feedback; mainly by nurses [42.4%]. • 82-95% of the open comments in the 4 domains were positive; highest # [71] of negative comments was in verbal communications skills. • Though assessors thought process was positive, 53% worried negative assessment would damage working relationship, and 92% said they'd complete the TAB honestly if they liked the learner. • Only 64% of SHOs received "no concern." • Comments were included 623/1378 assessments, mostly positive comments. • The SHOs found the process practical and fair, but only 65% found it helpful. • Most faculty (77%) learned nothing about their SHOs.

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Wood, 2006[47]	MSF	OB/GYN trainees	Team Observation tool (4 domains, 4-point scale)	• Ratings: 0.7% Needs serious attention, 5.2% Progress needed, 53% Fine, 38% Outstanding [15x more likely to be told they're good vs they're not]. • 11% negative comments, 13% mixed, 40% positive.
Blatt, 2008[48]	Videotape	2nd-year medical students	Setting: communications skill exam, feedback by 4th-year students as SPs	• 59% were neutral comments; 25% were positive; 16% corrective; no negative comments. • Factual information mainly, no high-level cognitive information in feedback.
Fyre, 1996[49]	Videotape	Internal medicine interns	Setting: feedback from faculty during CEX CEX observation guide sheet	• 3 organization structures: checklist-driven pattern; topical pattern (reflected nature of physician's task interacting with patient)-most common; Learner-centered pattern (2, 10, 6 of 24 feedback sessions, respectively]. Remaining 6 had topical and 1 of the other 2 patterns. • 19 of 25 videos had two-way communication. 5 had one-way communication, driven by faculty. • 6 of 24 videos noted equal psychological size, one clearly unequal, 16 in between. • 3 of 24 videos noted feedback only, but 20 went beyond feedback and did teaching.

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Ghaderi, 2015[50]	Videotape	A single minimally invasive surgery fellow	Setting: fellow reviewed his videos daily. GOALS [Global Operative Assessment of Laparoscopic Skills] OSATS [Objectively Structured Assessment of Technical Skills] "HM (Heller myotomy)" assessment tool	• Significant differences between attending and fellow ratings except for GOALS. • Attending ratings higher. • Ratings got better over the year. • Text feedback had 672 segments [64% fellow, 36% attending]- attending more focused on efficiency and safety, fellow more focused on technical issues.
Govaerts, 2013[51]	Videotape	General Practice residents	Setting: faculty viewed 2 videos of GP residents, wrote down feedback, and verbalized what their feedback would be	• 4-5 feedback statements in writing per resident. • Verbal feedback mimicked what was written, with 2-4 additional feedback comments. • Mostly general feedback, minimal specific feedback. • Most of feedback aimed at level of the task, and less attention to transfer of knowledge to other tasks, or fostering self-regulated learning. • More negative than positive valence to statements. • Only 28-31% of feedback was specific. Verbal feedback has more instances of specific feedback compared to written. • Both experienced and non-experienced evaluators gave some negative-toned feedback (88%, 86%, respectively).

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Hollingsworth, 1994[52]	Videotape	Preclinical medical students	Setting: feedback during an OSCE (objective structured clinical examination)	• 87.8% had at least 1 summative positive statement, only 11.1% had a summative negative statement. • 26.7% specific statements were positive, 85.6% specific statements were negative. • 87% of 125 students liked getting feedback.
Hulsman, 2015[53]	Videotape	4th-year medical students	Setting: communication skills during regular history-taking program. Peer and self-feedback.	• Self-assessment: more negative [3.28 comments] than positive [2.36]. • Peer: more positive [1.4 comments] than negative [0.68 comments]. • Most comments focused on topic of structuring the conversation, less so on suggestions. • Annotations with a negative valence were more specific.
Rizan, 2014[54]	Videotape	Year 5 medical students	Setting: bedside teaching encounters (BTE)	• Correction strategies that were at the extreme poles of explicitness [high or low] tended to be brief interactions. • Implicit feedback strategies are akin to "all might be revealed" to student eventually, keeping student in state of unknowing suspense. • Embedded correction strategies seemed to be more effective (e.g., extended Q/A sequence; faculty treating answer as possible but needing revision).
Ball, 2009[55]	Written feedback	Nursing students	Annotations of written scripts	• Students and staff found annotations useful as feedback. • Negative tone though undermined confidence. • 24% of students thought the hand written annotations were difficult to read

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Byrd, 2015[56]	Written feedback	Medical students	Ratings and comments for peers and self at end of semester	• Students rated themselves lower than peers. • Over the year, self-ratings improved in communication and self-awareness, while peer ratings improved in participating, preparedness and self-awareness. • Narrative comments more on strengths than weaknesses, mainly focused on professionalism 59% vs 19% knowledge; negative comments more evenly spread around areas like professionalism and knowledge.
Cook, 2014[57]	Written feedback	Surgery residents: PGY1-2s compared to PGY3-5s	Procedure Feedback Form	• Technical feedback to senior residents more specific and nuanced; included more feedback re: team leadership and teaching. • Residents improved over time. • Few comments on case outcome.
Dannefer, 2013[58]	Written feedback	First year medical students	Problem-based learning (PBL) assessment form for each PBL block; and portfolios (2-page essay)	• Targeted Areas for Improvement [TAFIs] focused on interpersonal skills related to participation or not. • Peers more likely than tutors to give feedback on TAFIs; tutor feedback less detailed and only 28% of 288 tutor assessments had TAFI feedback. • More mid-PBL block than end of PBL block TAFIs; TAFIs also decreased over year. • 95% of students self-identified an area not identified by their peers or tutors.

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Dekker, 2013[59]	Written feedback	Preclinical medical students completing problem-based learning (PBL) Professional Development module over 1 year	Rating tool with 10 point scale, and qualitative narrative	• 3 dimensions of written feedback comments: format (question vs statement), focus (related to the levels of students' reflections), and tone (positive vs negative). • 11 of 43 feedback comments classified as stimulating reflection, mainly focused on format of the feedback and tone.
Evans, 2005[60]	Written feedback	Senior physiotherapy students doing internship	Web-based diaries	• Students perceived need for clear and explicit feedback delivery process. • "Conflict of openness"- students reluctant to disclose their knowledge deficits; "danger" in admitting or denying errors or deficiencies. • Relationship to instructor powerful factor whether internship positive or negative.
Fitzgerald, 2010[61]	Written feedback	2nd-year nursing students	Continuous Assessment of Practice (CAP) documents	• 7 of 17 (41%) had formative feedback inconsistent with scores at midway and final interview (e.g., deficiencies in comments were related to passing scores and vice versa). • Overall feedback documentation was brief, non-specific, and did not include references for improvements. • Action plans if completed were done on ad hoc basis, and did not relate to issues identified.

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Haffling, 2011[62]	Written feedback	Final-year medical students	Leicester Assessment Package- modified. 5 competency domains, 7 point scale	• Highest ratings were in Relationships With Patients and Medical Interview; lowest ratings in Working Diagnoses and Investigations and Treatment. • Supervisor mean scores 5.3 (3-7) and students self-assess 4.5 (2.4-6.8). • Male students rated themselves significantly higher than female students in 3 competency domains (Working Diagnoses, Problem Solving, Investigations and Treatment). • Of all supervisor and student narrative comments about agreed upon goals, 88% were specific, 6% general, and 6% included no goals. • Female students with female supervisors were provided with significantly more specific goals (95% vs 85% with male supervisor). • Increased stringency noted with longer supervisor experience using the tool.
Hughes, 2008[63]	Written feedback	Medical students (during first phase of 3 two-year phases)	*eMed-Teamwork* computer-based system to capture peer and self-feedback about teamwork on group projects	• After 2.5 years, system had 5,237 feedback comments, mainly from peers (4,798). Facilitators had 130 feedback comments. • Average word length of feedback comments: self-assessment 98 words, peer 95 words, Facilitator 52 words. • Only 9% of peer feedback identified the author.

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Jackson, 2015[64]	Written feedback	Internal medicine residents	Monthly evaluation form: 6 domains, 1-9 scale	• 21% of 6,603 evaluations had no written feedback. • Of 2,056 unique utterances, most (29%) were non-specific, 20% were about the resident personally, 16% about patient care, and 14% about interpersonal communication. • 88% of written comments were positive. Negative comments focused on 3 ACGME competencies (medical knowledge, practice-based learning and improvement [PBLI], systems-based practice [SBP]). • Based on criteria developed in 10 small groups, 65% of written feedback moderate quality, 22% high quality and 13% low. • Attendings with higher quality feedback rated residents lower and had higher spread of ratings on all 6 ACGME competencies. • No relationship of In-training exams and quality or polarity of feedback.
Lindon-Morris, 2014[65]	Written feedback	3rd-year medical students	Reflections on feedback from videotaped group discussion	• All students expressed apprehension about video peer review and feedback. • Many comments about feeling publicly self-aware, almost to the point of being detrimental. • Very self-critical about their own performance. • Peer feedback viewed as positive experience, but negative feedback not thought to add anything to their own assessment.

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Melton, 2015[66]	Written feedback	1st-year medical students completing small group clinical case presentations	"Evaluation rubric" with descriptive statements, rating scales, and a comment section	• Most peer feedback (76%) was positive, while faculty provided more constructive narrative feedback (74%). • Numeric peer ratings higher than faculty ratings for all 6 domains. • No differences in themes or ratings between male and female students. Though females more likely to leave comments, and males more likely to leave constructively critical comments.
Nesbitt, 2014[67]	Written feedback	Year 4 (of 6) medical students.	Formative feedback on Supervised Learning Event	• 63.1% of feedback comments were Weak. Reasons included: non-specific, unclear, illegible, left blank. • Large % of forms had "keep practising" as a comment.
Pelgrim, 2013[68]	Written feedback	GP (General practice) trainees, 1st and 3rd years of training	Formative assessment forms- trainee enters reflection on performance, trainer enters narrative feedback, then both agree on joint action plan	• 66% and 34% of forms contained specific feedback and specific reflections, respectively. • 0.53 specific comments related to an action plan. • Trainer-trainee pairs with the best Guttman pattern (specific feedback and specific reflections) had 1.02 comments per effect (i.e., large effect).
Renting, 2016[69] *(published online ahead of print 2015)*	Written feedback	Internal Medicine residents in first 3 years of postgraduate training	Five situation-specific forms developed space for strengths and suggestions	• Written feedback was provided on all CanMEDS roles; most frequently within the situations of Patient Encounters and Oral Presentations. • Strengths (78%) provided more frequently than suggestions for improvement (52%). • Feedback was scored as specific (n=1024), moderately specific (n=77), or non-specific (n=543).

First author, year of publication[ref]	Context of feedback content analysis	Learners	Feedback tool or setting	Content analysis findings
Sherbino, 2006[70]	Written feedback	Emergency Medicine residents, all PGY levels	Form used global assessment with 150 mm visual analog scale; raters asked to write down 3 things to improve on	• Mean global rating score 104.3 mm (slightly below the above average anchor). • Seven general categories emerged. • Frequency of feedback on themes differed between faculty and peers.
Sinclair, 2007[71]	Written feedback	Year 3 medical students	Common Assessment Scale (CAS) grade	• Less than half (46.4%) collected their feedback sheets. • Female students were more likely than males to seek feedback. • Those students with higher CAS marks more likely to seek feedback.