Global Surgical Education - Journal of the Association for Surgical Education
https://doi.org/10.1007/s44186-022-00029-w

ORIGINAL ARTICLE

Multi-disciplinary assessment of the entrustable professional activities of surgery residents

S. Jung[1,4] · C. C. Stahl[1] · A. A. Rosser[1] · A. S. Kraut[2] · B. H. Schnapp[2] · M. Westergaard[2] · A. G. Hamedani[2] · R. M. Minter[1] · J. A. Greenberg[3]

Received: 22 November 2021 / Revised: 4 April 2022 / Accepted: 14 June 2022

Abstract

Purpose Medicine is practiced in a collaborative and interdisciplinary manner. However, medical training and assessment remain largely isolated in traditional departmental silos. Two Entrustable Professional Activities (EPAs) developed by the American Board of Surgery are multidisciplinary in nature and offer a unique opportunity to study interdisciplinary assessment.

Methods EPA microassessments were collected from Surgery and Emergency Medicine (EM) faculty between July 2018 and May 2020. Differences in feedback provided by faculty were assessed using natural language processing (NLP) techniques, (1) automated algorithms; and (2) topic modeling. Summative content analysis was used to identify themes in text feedback. We developed automated coding algorithms for these themes using regular expressions. Topic modeling was performed using latent Dirichlet allocation.

Results 549 assessments were collected for two EPAs: 198 for GS Consultation and 351 for Trauma. 27 EM and 27 Surgery faculty provided assessments for 71 residents. EM faculty were significantly more likely than Surgery faculty to submit feedback coded as *Communication*, *Demeanor*, and *Timeliness,* (all chi-square test p-values < 0.01). No significant differences were found for *Clinical Performance*, *Skill Level*, or *Areas for Improvement*. Similarly, topic modeling indicated that assessments submitted by EM faculty focused on communication, timeliness, and interpersonal skills, while those submitted by Surgery faculty focused on the residents' abilities to effectively gather information and correctly diagnose the underlying pathology.

Conclusions Feedback from EM and Surgery faculty differed significantly based on NLP analyses. EPA assessments should stem from multiple sources to avoid assessment gaps and represent a more holistic picture of performance.

Keywords Entrustable professional activities · Natural language processing · Assessment · Feedback · Interdisciplinary · Residents

Introduction

The practice of medicine is increasingly collaborative and interdisciplinary in nature. Furthermore, growing evidence suggests that well-delivered interdisciplinary care improves patient satisfaction and health outcomes [5, 7]. Collaboration among different disciplines and health professions is so important that the Interprofessional Education Collaborative (IPEC) has been established to provide guidance on this practice and has established a set of core competencies for interprofessional practice and training [15]. However, despite an increased emphasis on interdisciplinary and team-based patient care, medical training and assessment often remain largely isolated in traditional departmental silos. This is problematic given that

✉ S. Jung
jungs@surgery.wisc.edu

1 Department of Surgery, University of Wisconsin, Madison, WI, USA

2 BerbeeWalsh Department of Emergency Medicine, University of Wisconsin, Madison, WI, USA

3 Department of Surgery, Medical College of Georgia, Augusta, GA, USA

4 UW Hospital, K6/126 Clinical Science Center, 600 Highland Avenue, Madison, WI 53792, USA

Published online: 04 July 2022

no single individual observes all the professional behaviors of trainees, and thus getting feedback from a wider selection of sources paints a more complete picture of the improvement of skills and professional development. Further, trainees have indicated that they are open to feedback from other disciplines when working to develop skills in the domain of surgery [4].

As surgical training continues to evolve, there are more opportunities to engage in interdisciplinary training practices. In particular, the Entrustable Professional Activities (EPAs) for general surgery, developed by the American Board of Surgery, provide a context in which to reflect on how we assess residents and give them feedback on their professional skills and interactions as part of a patient care team. In short, the EPAs define the professional behaviors of a general surgeon and what the development of these behaviors looks like in practice [8]. To develop professional behaviors, feedback is required, but providing meaningful feedback in the clinical learning environment can be difficult. It has previously been shown that feedback can be improved by offering more frequent formative assessments with the goal of improving learners' performance [17]. Additionally, these professional behaviors of trainees are rarely performed in the absence of some form of supervision from medical professionals, including nurses and physicians from other specialties, thus providing an opportunity to receive feedback on performance from multiple team members. In particular, two of the currently developed EPAs, providing general surgical consultation (GS Consultation) and providing initial care for a traumatically injured patient (Trauma), are multidisciplinary in nature and offer a unique opportunity to study interdisciplinary assessment.

No single colleague or set of colleagues (such as physicians, nurses, or other groups of healthcare professionals) observes a given surgical resident for the entirety of their training. To paint the most comprehensive picture of a resident's skills development, it is key that feedback should come from a variety of interdisciplinary sources. However, very little research has explored interdisciplinary assessment in surgery and surgical education. As such, the goal of this work was to investigate the contributions of assessment and feedback from both surgeons and emergency medicine physicians when providing microassessments of the GS Consultation and Trauma EPAs for general surgery residents. We hypothesized that faculty from these different specialties would focus on different skills and facets of surgical residents' performance on these EPAs.

Materials and methods

Data collection

This study was reviewed and determined to be exempt by the Institutional Review Board. EPA microassessments were collected between July 2018 and May 2020 for general surgery residents at a Midwestern academic surgery program using an in-house smartphone application and stored in a secure Department of Surgery database. Paper microassessment forms were also available to the Emergency Medicine faculty. Data included information about attending faculty department as well as their text-based feedback for residents on cases categorized as General Surgery (GS) Consultation or Trauma. Prior to completing EPA microassessments, faculty were introduced to the app and the EPA assessment levels at either a faculty meeting or grand rounds and engaged in a discussion and group reflection on making decisions about performance levels and behaviors associated with those levels [1]. A total of 549 assessments were collected regarding these two EPAs: 198 for GS Consultation and 351 for Trauma. These two EPAs were chosen because they require multidisciplinary interaction, often involving faculty members from the Emergency Medicine department. For these EPAs, 27 Emergency Medicine (EM) and 27 Surgery faculty attendings provided assessments for 71 general surgery residents.

Data analyses

Differences in feedback provided by faculty were assessed using two natural language processing (NLP) techniques: (1) automated algorithms to capture qualitative themes related to performance and (2) topic modeling. Summative content analysis was used to identify themes in our corpus of text feedback. We developed automated coding algorithms for these codes using regular expressions. Topic modeling was performed using latent Dirichlet allocation. Broadly speaking, NLP identifies patterns in text based on the frequency and order of words and has been used in domains such as political forecasting, biological science, and medical education to measure the presence of themes and sentiments [10, 12, 18].

Automated code development and validation

A summative content analysis was used to identify competency-related themes in our corpus of text feedback. In summative content analysis, a corpus is evaluated (a) qualitatively, to identify potentially meaningful themes or patterns and (b) quantitatively, to determine whether those themes or patterns occur in frequencies that can be statistically analyzed [14]. Codes were developed inductively based upon themes in the corpus of data. Our analysis identified six common themes, or *codes*: (1) *Clinical Performance*, (2) *Communication*, (3) *Demeanor*, (4) *Skill Level*, (5) *Areas for Improvement*, and (6) *Timeliness* (see Table 1 for descriptions and examples).

We developed automated coding algorithms for each of these codes using regular expressions, which are code snippets that identify patterns in text. While automating the code *Timeliness*, for example, we developed an algorithm that identifies patterns in text related to timeliness and efficiency of tasks carried out.

To illustrate, the regular expression `[(?<!to) prompt\b|promptly]` searches the text for instances of the word "prompt" that (a) do not follow the word "to" and (b) include "promptly", but *not* other words which use "prompt" as a base, such as "prompted" and "prompting". This distinction allows the algorithm to capture text where "prompt" is used as an adjective or adverb (which is relevant to the concept of timeliness) and ignore text where "prompt" is used as a verb (which is not related to timeliness). As a result, the automated coding algorithm for *Timeliness* identifies the excerpts "[Resident] was **prompt** and courteous in their evaluation" and "The resident saw the patient **promptly**" as *Timeliness*, but excludes unrelated text such as "[Resident] developed an appropriate plan for workup and dispo independently w/o **prompt**ing to follow up *(sic)*" as well as "I had to **prompt** [resident] on the next steps."

Two trained human raters (a research specialist and an educational psychologist specializing in surgical education) established the reliability of all six automated coding algorithms. For each code, the human raters and the coding algorithm independently rated a random sample of at least 50 instances of feedback. Cohen's kappa was calculated to measure agreement both between the two human raters and between each human rater and the coding algorithm. To determine whether the kappa values obtained for these samples could be reasonably generalized to the whole dataset, Shaffer's rho (ρ) was calculated for each kappa using the rhoR package for the R statistical software platform. Rho can be interpreted similarly to a *p*-value and thus allowed us to measure the significance of the agreement between raters [11]. Because kappa was greater than or equal to 0.81 and rho was less than 0.05 for every code and for all combinations of raters (Table 2), we used the automated coding algorithms to code all feedback in the dataset.

Topic modeling using LDA

After removal of stop words, contractions, words less than three letters long, and special characters, topic modeling was performed using latent Dirichlet allocation (LDA) [6] from the topic models package in R [13]. LDA is a common machine learning algorithm used to identify latent topics in free text data (documents). The number of topics (*k*) is set, and then the unsupervised, generative algorithm analyzes the probability distributions of word frequencies and the structure of words within documents to identify the most likely underlying topics. Gamma scores are measures

Table 1 Overview of codes

Code	Description	Example
Clinical performance	Skills related to workup, appropriateness of tests conducted, documentation, history-taking, medical/diagnostic knowledge	"[Resident] did a nice job recognizing a patients' deteriorating medical condition (seizure) and appropriately upgrading the trauma."
Communication	Skills related to communication, reasoning, dissemination of plans, delegation within a team, communication with patient and/or patient's family	"Coordinated care with neurosurgical team and ED team. Overall very organized and able to think ahead without prompting."
Demeanor	How personality or demeanor as relates to working with a team	"Managed trauma calmly and efficiently."
Skill level	Appropriateness of domain-agnostic skills to resident training level; ability to carry out tasks independently; ability to supervise and/or mentor other learners	"Working at fellow level, as expected." "[Resident] supervises junior residents well." "Independent practice with trauma evaluation."
Areas for improvement	Skills that need improvement, identification of mistakes made, next steps a more advanced trainee should carry out	"[Resident] appropriately understood how sick (or not) the patient was, and came up with a reasonable interventional plan. The next step would be to appreciate that the fluid collections were too small to stick a drain in, but minor detail."
Timeliness	Comments regarding speed or efficiency of tasks carried out	"Recognized need for limited providers in the room due to COVID-19 and assigned roles clearly and quickly."

Table 2 Coding validation

Code	Human 1 vs. human 2		Human 1 vs. computer		Human 2 vs. computer	
	Kappa	Rho	Kappa	Rho	Kappa	Rho
Clinical performance	0.90	< 0.01	0.88	< 0.01	0.85	0.01
Communication	0.94	0.02	0.90	0.01	0.87	0.03
Demeanor	0.98	< 0.01	0.81	0.01	0.96	0.02
Skill level	0.98	< 0.01	0.92	< 0.01	0.93	< 0.01
Areas for improvement	0.98	0.01	0.87	0.02	0.98	0.02
Timeliness	0.99	< 0.01	0.98	< 0.01	0.98	< 0.01

Table 3 Differences in surgery vs. emergency medicine faculty feedback code proportions

	Faculty department		*p*-value
	Emergency medicine	Surgery	
Clinical performance	85.8%	82.4%	0.371
Communication	70.7%	39.4%	6.979e-12*
Demeanor	24.8%	7.1%	2.006e-06*
Skill level	20.6%	24.7%	0.3322
Areas for improvement	16.1%	19.4%	0.4058
Timeliness	33.2%	18.2%	0.0004717*

*Statistically significant at $p < 0.05$

of topic-document association (i.e., a gamma of 1 means a document consists 100% of that topic). After topics are identified, gamma scores were calculated for text feedback submitted by Surgery and EM faculty. The closer the gamma scores to "1" the more representative the topics were considered to be of the feedback given by the two departments. Once representative topics were found, the word lists representing the topics were compared for salient differences between feedback given by the EM vs Surgery groups.

Results

Automated coding

After automatically coding our corpus of text, we conducted chi-square tests to calculate differences in how EM and Surgery faculty attendings gave feedback for General Surgery Consultation and Trauma EPAs. Table 3 summarizes the proportions of assessments submitted by faculty that were coded as each of our six codes. A total of 549 assessments were analyzed. The percentages reported are calculated out of the total assessments submitted by each faculty division (379 from EM and 170 from Surgery).

We measured differences in code frequencies in text feedback from EM and Surgery faculty attendings. EM faculty were significantly more likely to submit feedback coded as *Communication*, *Demeanor*, and *Timeliness*. No significant differences were found for *Clinical Performance*, *Skill Level*, or *Areas for Improvement* (Table 3). These results indicate that EM faculty are significantly more likely than Surgery faculty to remark upon skills which have to do with teamwork, demeanor, and efficiency.

We did not further subdivide our data to analyze differences in EM versus Surgery faculty feedback for GS Consultation and Trauma separately or for residents' program year, as there were insufficient data to yield valid chi-square results.

Topic modeling

The results of LDA topic modeling are displayed in Table 4. Although k was initially set to 2 to identify topics differing between EM and Surgery faculty over both EPAs, the topics identified by the algorithm using this method had low gamma scores (~ 0.5), illustrating that the topics were not well-representative of the feedback data. The topics identified corresponded to the latent topics of Trauma and Consult EPAs, rather than faculty differences. However, after changing k to 4, the latent topics identified corresponded very well with feedback submitted by EM vs Surgery faculty for the two different EPAs. The LDA Topics A–D indicate the gamma scores for the latent topics after changing k to 4 (Table 4, gammas 1 or ~ 0).

Consult EPAs submitted by EM faculty focused primarily on communication skills and timeliness (words such as "communication", "prompt", "family", "recommendations", "promptly", "quickly"), while consulting EPAs submitted by Surgery faculty focused primarily on the residents' ability to effectively gather information and then correctly diagnose the underlying pathology leading to the consult (words such as "appropriately", "history", "physical", "information", "recognized", "accurate", "diagnosis", "complete", "correct") (Table 4).

Similarly, Trauma EPAs submitted by EM faculty focused heavily on communication and interpersonal skills while performing the initial evaluation (words such

Table 4 Results of latent dirichlet allocation (LDA) topic modeling

	Consult EPA		Trauma EPA	
	EM faculty	Surgery faculty	EM faculty	Surgery faculty
LDA Topic A	1	0.000272	0.00000612	0.0000195
LDA Topic B	0.0000139	0.000272	0.00000612	1
LDA Topic C	0.0000139	0.000272	1	0.0000195
LDA Topic D	0.0000139	1	0.00000612	0.0000195
	Consult EPA		Trauma EPA	
Topic word lists	EM faculty	Surgery faculty	EM faculty	Surgery faculty
1 (Stronger)	Patient	Patient	Trauma	Patient
2	Team	Job	Team	Trauma
3	Plan	Plan	Patient	Team
4	Communication	Consult	Plan	Evaluation
5	Job	Appropriately	Job	Job
6	Prompt	History	Communication	Plan
7	Care	Physical	Evaluation	Imaging
8	Evaluation	Care	Nice	Management
9	Consult	Information	Care	Level
10	Consultation	Recognized	Excellent	Injury
11	Family	Accurate	Assessment	Recognized
12	Recommendations	Diagnosis	Calm	Exam
13	Decision	Excellent	Patients	Communication
14	Promptly	Complete	Imaging	Excellent
15 (Weaker)	Quickly	Correct	Approach	Pediatric

LDA Topics A–D indicate the gamma scores for the latent topics

as "communication", "nice", "calm", "approach"), while Trauma EPAs submitted by Surgery faculty focused more on the outcomes of the workup and injury identification (words such as "imaging", "management', "injury", "recognized", "exam") (Table 4).

Discussion

The goal of this work was to contribute to the literature on interdisciplinary assessment in surgery education. The EPAs of GS Consultation and Trauma offered a multidisciplinary clinical workplace context in which to explore assessment from faculty in both EM and Surgery. Our hypothesis that faculty from these specialties would focus on different skills and abilities of surgical residents' performance was supported. However, where we did see significant differences, it was for EM faculty offering more feedback in the areas of *Communication*, *Demeanor*, and *Timeliness*, with no significant difference in feedback on *Clinical Performance*, *Skill Level*, or *Areas for Improvement.*

Potential explanations for these findings include both workflow and relationships across specialties. For example, at our institution, it is not uncommon for only EM faculty to be present for the primary and secondary survey in the trauma bay if the trauma is not categorized as Level 1. Therefore, the EM faculty would be the only faculty available to fill the critical role of giving residents feedback during those situations. Further, having less familiarity with the residents may also have influenced the focus on interpersonal skills, with EM faculty putting particular emphasis on these skills. This is an important finding, given the necessity of team-based care and positive interpersonal relationships across disciplines and medical professions. Due to the nature of clinical situations that require horizontal communications across teams, it is essential to provide feedback on these interpersonal and communication skills that can facilitate teamwork and patient care and safety.

This study extends work in the field of surgical education on both EPA assessment and feedback in multidisciplinary medical contexts. Our prior work showed that natural language processing techniques could be used to map feedback provided on EPAs to distinct entrustment levels [2]. We have also found no differences in faculty feedback by gender [3]. This work extends our previous findings by showing that we can also use these techniques to understand differences in feedback offered on the same EPAs by assessors from multiple disciplines. Moreover, research investigating the performance and communication behaviors of interprofessional teams suggests that trust, personality characteristics, and communication style play a large role in interdisciplinary team success and patient care [16]. Our findings align with this work by showing that EM faculty placed an emphasis on being able to depend upon a prompt response, residents' demeanor, and interpersonal communication as part of the team. Finally, recent work has also reported differences in how EM and Surgery faculty assess the communication and interpersonal skills of residents in trauma resuscitation scenarios, albeit in the context of simulation. This work found that faculty were more likely to assess trainees from their own programs more critically on their communication skills [9]. The chi-square analysis focused on feedback frequency and found that, when left open-ended as to aspects of performance on which to provide feedback, EM faculty were more likely to assess surgery residents on their communication skills. However, this does not mean that the Surgery faculty did not provide feedback on these skills, but EM faculty did this more often. In addition, the LDA topic modeling showed that while both Surgery and EM faculty are able to provide feedback on communication skills, they are seeing and highlighting different things in their feedback, showing that it is important to encourage and accept feedback from multiple people who are engaging with residents in their work. For example, the GS consultation assessments were done based on real-time interactions by the EM faculty. For the Surgery faculty, conversations were usually done after discussing the consultation with the resident, which likely contributed to differences in what was focused on in the feedback. Thus, while training surgeons in providing feedback may improve the quality and frequency, it will not change the fact that there are certain aspects of care processes engaged in by residents that they do not observe.

Limitations and future directions

There are limitations to our study that present opportunities for future research. We did not study the assessments of EM residents by the Surgery and EM faculty. Doing so in future work could help to shed light on why these differences exist. For example, based on the findings of our current work, one may hypothesize that EM faculty would focus more on clinical competencies and less on interpersonal relationships and communication due to the familiarity that exists with the residents in their program. We did not look at feedback quality or length, only general content and frequency, and we did not separate residents' data by year or by trauma level. It would be useful to look into the results of this analysis by program year and level of trauma using additional data, ideally from multiple institutions. Further, we did not look at the ways in which the Surgery versus EM faculty talked about the communication skills of surgery residents in their feedback, which would be an interesting addition to the literature going forward. Additionally, there may be other NLP techniques that might allow for increased accuracy in the modeling of feedback data. One such method that has been previously utilized for short text responses is the Dirichlet Multinomial Mixture model [19]. It will be important to explore these alternative techniques in future work. Finally, we would like to collect EPA microassessment data from other specialties and disciplines, such as hospitalist medicine and nursing, to investigate what these perspectives add to residents' feedback on their performance.

Conclusions

Feedback on surgical residents' performance from EM and Surgery faculty differed significantly based on multiple NLP analyses. As such, we can conclude that some gaps are present in assessment when residents are only assessed by surgeons that are addressed when EM faculty provide assessments and feedback. EPA assessments should stem from multiple sources to avoid assessment gaps and represent a more holistic picture of performance to ensure that trainees are prepared to practice in the collaborative modern medical environment.

Data availability The data that support the findings of this study are not openly available due to the sensitive nature of surgery residents' assessment information and are available from the corresponding author upon reasonable request with appropriate security measures in place from our secure database.

Declarations

Conflict of interest S. Jung is a member of the GSE Editorial Board. There are no other declarations from any of the other authors and there is no funding to report for this work.

References

1. Stahl CC, Collins E, Jung SA, Rosser AA, Kraut AS, Schnapp BH, Westergaard M, Hamedani AG, Minter RM, Greenberg JA. Implementation of entrustable professional activities into a general surgery residency. J Surg Educ. 2020;77(4):739–48.
2. Stahl CC, Jung SA, Rosser AA, Kraut AS, Schnapp BH, Westergaard M, Hamedani AG, Minter RM, Greenberg JA. Natural language processing and entrustable professional activity text feedback in surgery: a machine learning model of resident autonomy. Am J Surg. 2021;221(2):369–75.
3. Padilla EP, Stahl CC, Jung SA, Rosser AA, Schwartz PB, Aiken T, Acher AW, Abbott DE, Greenberg JA, Minter RM. Gender differences in entrustable professional activity evaluations of general surgery residents. Ann Surg. 2022;275(2):222–9.

4. Al-Jundi W, Wild J, Ritchie J, Daniels S, Robertson E, Beard J. Assessing the nontechnical skills of surgical trainees: views of the theater team. J Surg Educ. 2016;73(2):222–9.
5. Ashcraft S, Bordelon C, Fells S, George V, Thombley K, Shirey MR. Interprofessional clinical rounding: effects on processes and outcomes of care. J Healthc Qual. 2017;39(2):85–94. https://doi.org/10.1097/JHQ.0000000000000039.
6. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res. 2003;3:993–1022.
7. Boykin A, Wright D, Stevens L, Gardner L. Interprofessional care collaboration for patients with heart failure. Am J Health Syst Pharm. 2018;75(1):e45–9. https://doi.org/10.2146/ajhp160318.
8. Brasel KJ, Klingensmith ME, Englander R, Grambau M, Buyske J, Sarosi G, Minter R. Entrustable professional activities in general surgery: development and implementation. J Surg Educ. 2019;76(5):1174–86.
9. Caldwell KE, Lulla A, Murray CT, Handa RR, Romo EJ, Wagner JW, Wise PE, Leonard JM, Awad MM. Multi-Disciplinary Trauma Evaluation and Management Simulation (MD-TEAMS) training for emergency medicine and general surgery residents. Am J Surg. 2021;221(2):285–90.
10. Chary M, Parikh S, Manini AF, Boyer EW, Radeos M. A review of natural language processing in medical education. West J Emerg Med. 2019;20(1):78–86. https://doi.org/10.5811/westjem.2018.11.39725.
11. Eagan B, Rogers B, Pozen R, Marquart CL, Shaffer DW (2017) rhoR: rho for inter rater reliability. R package version 1.1.0.0, 2017
12. Haselmayer M, Jenny M. Sentiment analysis of political communication: combining a dictionary approach with crowdcoding. Qual Quant. 2017;51(6):2623–46. https://doi.org/10.1007/s11135-016-0412-4.
13. Hornik K, Grün B. Topicmodels: an R package for fitting topic models. J Stat Softw. 2011;40(13):1–30.
14. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res. 2005;15:1277–88.
15. Interprofessional Education Collaborative (IEC) (2016) Core competencies for interprofessional collaborative practice: 2016 update. Washington, DC. https://hsc.unm.edu/ipe/resources/ipec-2016-core-competencies.pdf
16. Kennedy E, Lingard L, Watling CJ, Alejandro RH, Leigh JP, Cristancho SM. Understanding helping behaviors in an interprofessional surgical team: How do members engage? Am J Surg. 2020;219(2):372–8.
17. McQueen SA, Petrisor B, Bhandari M, Fahim C, McKinnon V, Sonnadara RR. Examining the barriers to meaningful assessment and feedback in medical training. Am J Surg. 2016;211(2):464–75.
18. Thessen AE, Cui H, Mozzherin D. Applications of natural language processing in biodiversity science. Adv Bioinform. 2012. https://doi.org/10.1155/2012/391574.
19. Yin J, Wang J (2014) A dirichlet multinomial mixture model-based approach for short text clustering. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining, pp 233–242