



Published in final edited form as:

Am J Surg. 2022 July ; 224(1 Pt B): 391–395. doi:10.1016/j.amjsurg.2021.12.035.

Developing a longitudinal database of surgical skills performance for practicing surgeons: A formal feasibility and acceptance inquiry

Megan K. Applewhite^a, LaDonna E. Kearse^b, Hossein Mohamadipanah^b, Anna Witt^b, Cassidi Goll^b, Brett Wise^b, James R. Korndorffer Jr.^b, Carla M. Pugh^{b,*}

^aDepartment of Surgery, Albany Medical College, 43 New Scotland Avenue, Albany, NY, 12208, USA

^bDepartment of Surgery, Stanford University, 300 Pasteur Drive, Stanford, CA, 94305, USA

Abstract

Background: We explored the feasibility and surgeons' perceptions of the utility of a longitudinal skills performance database.

Methods: A 10-station surgical skills assessment center was established at a national scientific meeting. Skills assessment volunteers (n = 189) completed a survey including opinions on practicing surgeons' skills evaluation, ethics, and interest in a longitudinal database. A subset (n = 23) participated in a survey-related interview.

Results: Nearly all participants reported interest in a longitudinal database and most believed there is an ethical obligation for such assessments to protect the public. Several interviewees specified a critical role for both formal and informal evaluation is to first create a safe and supportive environment.

Conclusions: Participants support the construction of longitudinal skills databases that allow information sharing and establishment of professional standards. In a constructive environment, structured peer feedback was deemed acceptable to enhance and diversify surgeon skills. Large scale skills testing is feasible and scientific meetings may be the ideal location.

Keywords

Performance database; Surgical skills assessment; Surgeon skills tracking

* Corresponding author. Department of Surgery, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA, USA. cpugh@stanford.edu (C.M. Pugh).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.amjsurg.2021.12.035>.

Declaration of competing interest

The authors have no conflicts of interest to disclose.

1. Introduction

Evaluation of technical skills is widely accepted as a critical part of surgical training, both inside and outside the operating room. Residents and fellows are closely supervised in the operating room, with ongoing critique of their technical competency by attending surgeons throughout their training. Outside of the operating room, surgical simulation centers are embedded in institutions to facilitate execution of skills curricula and satisfactory completion of Fundamentals of Laparoscopic Surgery and Fundamentals of Endoscopic Surgery examinations required prior to graduating residency to ensure standards of technical competency.¹ However, this framework of training and competency assessment does not exist for practicing surgeons.

With the advent of advanced technologies, practicing surgeons are interested in exploring various modalities to enhance continued training such as artificial intelligence, big data, and video-based coaching.²⁻⁴ Artificial intelligence, especially when equipped with large data sets and accompanying sensor technology, can aid surgeons in identifying technical preferences and predicting decision-making strategies that portend better or worse outcomes.⁵ Furthermore, video-based coaching allows the opportunity for deliberate practice, individualized feedback, and peer coaching.⁶ Although efforts have been made to continue surgical training while in independent practice, the close intraoperative supervision and formal technical assessments present during residency or fellowship dissolve.

In spite of the fact that the safety of the general public relies on quality and consistency of surgeon skill and decision-making, attending physicians do not have the ability to evaluate their technical capabilities over time. Given this, there is a role for such longitudinal self- and peer-assessment of technical skills and tracking decision-making. In many other safety-conscious industries strict metrics are imposed for evaluation of their employees to ensure competence⁷ and surgeon variability in technique have been shown to affect patient outcomes.⁸⁻¹⁰ However, it is not known how surgeons would respond to structured, low-stakes skills assessment throughout their careers, or how this would be meaningfully and logistically carried out in the profession.

Despite strong interest in artificial intelligence, big data and video-based coaching for practicing surgeons, there is a paucity of research reporting feasibility and surgeon interest in performance databases. Establishment of a longitudinal database to better understand both surgeon-specific and profession-wide performance changes over time would allow for creation of performance assessments that establish benchmarks of competence and identify opportunities for intervention when appropriate. This experiential project explored the feasibility (cost, personnel, etc.) and surgeons' perceptions of the utility of a longitudinal skills performance database. In this qualitative analysis, we investigate if surgeons will be supportive of profession-directed skills performance databases and the factors driving that support.

2. Material and methods

2.1. Data collection

A 10-station surgical skills assessment center was established in the exhibit hall at the 2019 American College of Surgeons Clinical Congress in San Francisco, CA. Attendees of the meeting were voluntarily recruited to participate in this study. The study was approved by the Institutional Review Board at Stanford University.

2.2. Skills assessment center

The skills assessment center was divided into 10 simulated operating rooms and a cognitive task center (Fig. 1) Each participant was instructed to locate and repair any injuries encountered on a porcine segment of bowel. Audio, video, motion tracking, and electroencephalogram (EEG) data were collected during the task. The cognitive task center captured baseline EEG to be used for comparison during EEG captured throughout the bowel repair.

2.3. Survey

A survey was administered prior to the hands-on surgical skills assessment that queried participants regarding age, gender, specialty, and years as a practicing surgeon. In addition, participants indicated their interest in contributing data to and tracking performance using a longitudinal database, using this database for self-assessment of performance, and whether they believed there was an ethical obligation to have surgeons' skills and judgement tested to protect the public.

2.4. Interview

After participating in a 30-min hands-on surgical skills assessment, a subset of attendees participated in a 30-min interview with in-depth questions relating to the survey items. A pool of survey questions were developed by the interviewer as well as a surgeon colleague with significant experience and background in ethics¹¹ to gain a qualitative understanding of surgeons' motivations to participate, understand how they felt while performing the technical test and to elicit some more personal viewpoints on the idea of skills assessment. Survey questions were then shared with a research advisory committee, consisting of a surgeon ethicist, ACS Board of Governors members, ACS Executive staff, and American Board of Surgery Executive staff.¹² After finalizing which of the survey questions would be included, the interviewer and one of the research advisory committee members (CP) generated a list of interview questions that would complement the survey. The interview script was piloted with 8 surgeons at a single institution. There were no major changes made after this event.

Participants underwent a semi-structured interview by a single interviewer (MA) using the previously developed interview script which included questions regarding their interest in participating in the study, how they thought the hands-on skills assessment went, and their thoughts about standardized testing for surgeons or peer evaluation of efficiency feedback or technical tips (Fig. 2). The interviewer is a practicing surgeon ethicist.^{13(p)} She had no

established knowledge of the participants. Participants' qualitative responses on views of their hands-on skills assessment was transcribed live by the interviewer.

2.5. Statistical analysis

Data were analyzed using descriptive statistics and chi-square tests. Thematic analysis after inductive coding by a research resident (LK) was used to analyze qualitative responses and transcription was evaluated until saturation was reached. No interrater reliability was performed given one coder. Resident physician and retired physician survey and interview responses were excluded from analysis. Two-sided p-values were calculated, and $p < 0.05$ was used to determine statistical significance. Analyses were completed using IBM SPSS Statistics for Mac (Version 27.0, IBM Corp., Armonk, NY, USA).

3. Results

3.1. Survey

The results included survey participants ($n = 189$) and interview participants ($n = 23$) ranging in age from 29 to 77 with no significant difference in gender, specialty, or years as a practicing surgeon (Table 1). Specialties of participants included General Surgery, Minimally Invasive Surgery, Trauma Surgery, Colorectal Surgery, Surgical Oncology, and Other. Of those surveyed, 67% reported having a clinical workload of at least 75%, with 66% of those surgeons spending at least half of their clinical time operating. 96% of participants ($n = 181$) indicated an interest in contributing their data to a longitudinal performance database and 96% demonstrated interest in tracking their individual performance against this database. 82% ($n = 149$) believed there was an ethical obligation to have skills assessment to protect the public (see Table 2).

3.2. Interview

Several themes emerged during qualitative analysis of interviewee responses (Table 3). More than half of interviewees (56.5%, $n = 13$) stated they participated due to curiosity and a desire to contribute to research. The Surgical Metrics Project announcement by American College of Surgeons leadership also motivated surgeons ($n = 8$) to participate in our study. Additionally, some surgeons were motivated to participate based on assessing a difference in their perception of simulated bowel repair performance compared to their intraoperative performance in real clinical environments. Surgeons' perceptions of their performance were evaluated with some surgeons reporting confidence in the integrity of their repair before the subsequent leak test. However, several surgeons reported decreased confidence in the quality of their bowel repair due to concerns such as porcine tissue quality and surgical instruments and suture supplied. Furthermore, assistant skill level was also reported to impact confidence in performing the procedure. In analyzing responses regarding standardized testing for surgeons or peer evaluation, 69.5% ($n = 13$) of participants largely expressed interest in standardized testing or peer evaluation. However, 21.7% ($n = 5$) of participants stated that there was no role for formal evaluation without first creating a safe and supportive environment.

3.3. Cost of skills assessment

At the skills assessment center, a surgical assistant was required at each station ($n = 10$) and a person to collect video and motion data was required for every 2 stations ($n = 5$). In addition, administrative personnel ($n = 1$) were needed to manage the booth and guide participants to different stations. The cost of personnel and disposables such as porcine bowel, storage, surgical instruments, sutures, gloves, and cleaning supplies were around \$30,000. This cost did not cover exhibit booth space, as this was provided by the American College of Surgeons. On average, booth space at society meetings ranges from free (reserved education/research space) to an average cost of \$2000 for small 10×10 feet booths to $> \$20,000$ for large exhibit booths greater than 40×40 feet.

4. Discussion

In the group studied, there was overwhelming support for longitudinal performance databases, with greater than 95% of all participants expressing willingness to contribute data, track individual performance, and utilize testing to assess skills and clinical judgement over time. To our knowledge, this is the first study to evaluate surgical metrics of surgeons and understand surgeon positions on contribution to longitudinal databases and skill assessment. We found that surgeons are interested in learning more about their own skills at any given time, and as they evolve over the course of their careers. Knowing this, development of a framework for implementation is the next step. Unlike the aviation industry which has modeled a method of strict and rigid training and assessment guidelines, this longitudinal database would need to be highly adaptive in its construction.^{14,15} Ideally, low-stakes, formative skills assessments of surgeons and creation of a longitudinal database would be designed, implemented, and managed within the confines of the profession of surgery, as opposed to being run by a business, institution, governmental entity, or other outside facilitator. This would allow an opportunity for self-regulation and ensure that the environment is constructive, focused, and efficient, with the end goal that patients are experiencing safe and consistent care when having surgery by board certified surgeons.

Given the wide variety of geographic locations of surgeons as well as varied capabilities to travel, the logistical components of cognitive and technical testing on a large scale proposes some challenges. One potential solution is to gather surgical performance data at society meetings (local, regional, and national) in order to capture as many surgeons as possible, and this study demonstrated feasibility of such a model. While this format may not be easily accessible to all surgeons and may pose limitations due to cost, it would certainly be able to provide reproducible and structured throughput that does not require additional extensive travel or expenses for those who are able to attend such conferences.

Peer evaluation for efficiency feedback or technical tips is a less structured and more informal strategy to “cross-pollinate”, varied experiences of individual surgeons and what many participants thought was a way to enhance surgeon skill level at the institution level. As surgeons advance from training, many develop idiosyncratic tricks or innovative ideas in the operating room that may make a given operation more efficient or technically more straightforward. The best structure for peer-mediated efficiency feedback would be institution-dependent, but what is clear from participants in this study is for such a

mechanism assessment/tool sharing to be successful, it must be done in a safe environment that is entirely separate of annual evaluations, promotions, or compensation. This requires a culture intentionally built to be constructive and collaborative without fear of punishment or judgement, particularly when the surgeon is of junior rank, thereby reinforcing the need for surgical coaching.¹⁶

Interestingly, only about 8 out of every 10 participants thought surgeons have an ethical obligation to submit to skills assessment in an effort to protect the public. Surgery is among the few professions wherein an individual's technical capabilities and the capacity to make timely critical decisions has a safety impact on the general public. To draw a parallel to the aviation profession wherein the operator's skills affect the public safety, pilots are required to undergo annual flight review tests with a flight instructor every 24 months throughout their career to ensure safe conduct of flight operations.¹⁷ Certainly discussions regarding how to carry out testing and how to define a threshold for "safety" will need to be well thought out and studied. Using objective metrics, such as motion tracking, may be such a way to begin to define surgical "safety" thresholds.¹⁸ Doing so in a collaborative and transparent way will optimize the likelihood of buy in from the most surgeons possible.

We recognize that our study faces several limitations. Selection bias may be present, as those participants who opted to participate may share similar characteristics and perceptions of the role of technical skills assessment and the implementation of a skills database. We acknowledge that ascertaining surgeon perceptions of ethical obligations using closed-ended questions may introduce framing bias. We also recognize that bias may have been introduced given live transcription by the interviewer and the potential for themes to be missed. However, biases were mitigated during the data analysis process by having a non-practicing surgeon conduct the coding. The use of a single coder also limited our ability to recognize any potential conflicts in the coding process or assess inter-coder reliability. Similarly, the surgeons who volunteered to participate may not be representative of the practicing surgeon demographic, leading to a sampling bias. This, in turn, may also present a challenge to implementation, as capturing a wide variety of surgeon data may be limited to only those able to attend local, regional, or national scientific meetings. Additionally, as survey data was collected, this introduces some recall bias in regard to reported clinical and administrative duties. Furthermore, variability in the quality of tissue and comfort repairing animal tissue models in a simulated environment could also pose a challenge if used strictly for testing purposes.

Future studies will explore the implementation of longitudinal skills tracking and its potential relationship to patient outcomes and patient-reported quality of life. Such a database could also be helpful in identifying "best practices," big data, and offering additional opportunities for collaborative education and training at the practicing surgeon level.

5. Conclusions

Large scale skills testing is feasible and national or regional scientific meetings may be a potential location for such testing. This study identifies specific interests in safe and

supportive environments as well as interest in performance databases that allow information sharing and peer/self-assessment.

Funding

This research was partially funded by the American College of Surgeons [The Surgical Metrics Project, XVQU].

References

1. ACS/APDS. Surgery resident skills curriculum. American College of surgeons. <http://www.facs.org/education/program/resident-skills>. Accessed May 10, 2021.
2. Greenberg CC, Dombrowski J, Dimick JB. Video-based surgical coaching: an emerging approach to performance improvement. *JAMA Surg.* 2016;151(3): 282–283. 10.1001/jamasurg.2015.4442. [PubMed: 26650338]
3. Mathias B, Lipori G, Moldawer LL, Efron PA. Integrating “big data” into surgical practice. *Surgery.* 2016;159(2):371–374. 10.1016/j.surg.2015.08.043. [PubMed: 26603852]
4. Loftus TJ, Upchurch GR, Bihorac A. Building an artificial intelligence–competent surgical workforce. *JAMA Surg.* 2021;156(6):511. 10.1001/jamasurg.2021.0045. [PubMed: 33688933]
5. Wall J, Krummel T. The digital surgeon: how big data, automation, and artificial intelligence will change surgical practice. *J Pediatr Surg.* 2020;55:47–50. 10.1016/j.jpedsurg.2019.09.008.
6. Singh P, Aggarwal R, Tahir M, Pucher PH, Darzi A. A randomized controlled study to evaluate the role of video-based coaching in training laparoscopic skills. *Ann Surg.* 2015;261(5):862–869. 10.1097/SLA.0000000000000857. [PubMed: 25185469]
7. Dellinger EP, Pellegrini CA, Gallagher TH. The aging physician and the medical profession: a review. *JAMA Surg.* 2017;152(10):967. 10.1001/jamasurg.2017.2342. [PubMed: 28724142]
8. Birkmeyer JD, Finks JF, O’Reilly A, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med.* 2013;369(15):1434–1442. 10.1056/NEJMs1300625. [PubMed: 24106936]
9. Stulberg JJ, Huang R, Kreutzer L, et al. Association between surgeon technical skills and patient outcomes. *JAMA Surg.* 2020;155(10):960. 10.1001/jamasurg.2020.3007. [PubMed: 32838425]
10. Fecso AB, Bhatti JA, Stotland PK, Queresby FA, Grantcharov TP. Technical performance as a predictor of clinical outcomes in laparoscopic gastric cancer surgery. *Ann Surg.* 2019;270(1):115–120. 10.1097/SLA.0000000000002741. [PubMed: 29578907]
11. Angelos P The ethics of introducing new surgical technology into clinical practice: the importance of the patient–surgeon relationship. *JAMA Surg.* 2016;151(5):405. 10.1001/jamasurg.2016.0011. [PubMed: 26982505]
12. Looking forward - september 2019. The bulletin. Published September 1 <https://bulletin.facs.org/2019/09/looking-forward-september-2019/>; 2019. Accessed November 9, 2021.
13. Reis-Dennis S, Applewhite MK. Ethical considerations in vaccine allocation. *Immunol Invest.* 2021;50(7):857–867. 10.1080/08820139.2021.1924771. [PubMed: 34036865]
14. Gogalniceanu P, Calder F, Callaghan C, Sevdalis N, Mamode N. Surgeons are not pilots: is the aviation safety paradigm relevant to modern surgical practice? *J Surg Educ.* 2021;78(5):1393–1399. 10.1016/j.jsurg.2021.01.016. [PubMed: 33579654]
15. Lipshy KA, Britt LD. How do we improve patient safety? A look at the issues and an interview with. Dr. Britt. *Bull Am Coll Surg* 2017;102(2):22–29. [PubMed: 28925176]
16. Pradarelli J, et al. Surgical coaching to achieve the ABMS vision for the future of continuing board certification. *Am. J. Surg* 2020. 10.1016/j.amjsurg.2020.06.014. In press.
17. Ac 61–98D. Currency requirements and guidance for the flight review and instrument proficiency check – document information. https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentid/1033391. Accessed May 10, 2021.
18. Perrone K, et al. Translating motion tracking data into resident feedback: An opportunity for streamlined video coaching. *Am. J. Surg* 2020. 10.1016/j.amjsurg.2020.01.032. In press.

Survey Questions		
I am interested in contributing by data to a longitudinal database of physician performance.	Yes	No
I would be interested in tracking my own performance relative to a larger database year after year.	Yes	No
If testing was reliable and confidential, would you utilize testing to assess your own skills and judgment over time?	Yes	No
Do you believe that there is an ethical obligation to have surgeons' skills and judgment tested in order to protect the public?	Yes	No
Interview Questions		
What made you stop by our booth to participate in the study?		
How do you think the [technical and cognitive skills test] went?		
What do you think about standardized testing for surgeons or peer evaluation for efficiency feedback or technical tips?		

Fig. 1.
Survey and interview questions.

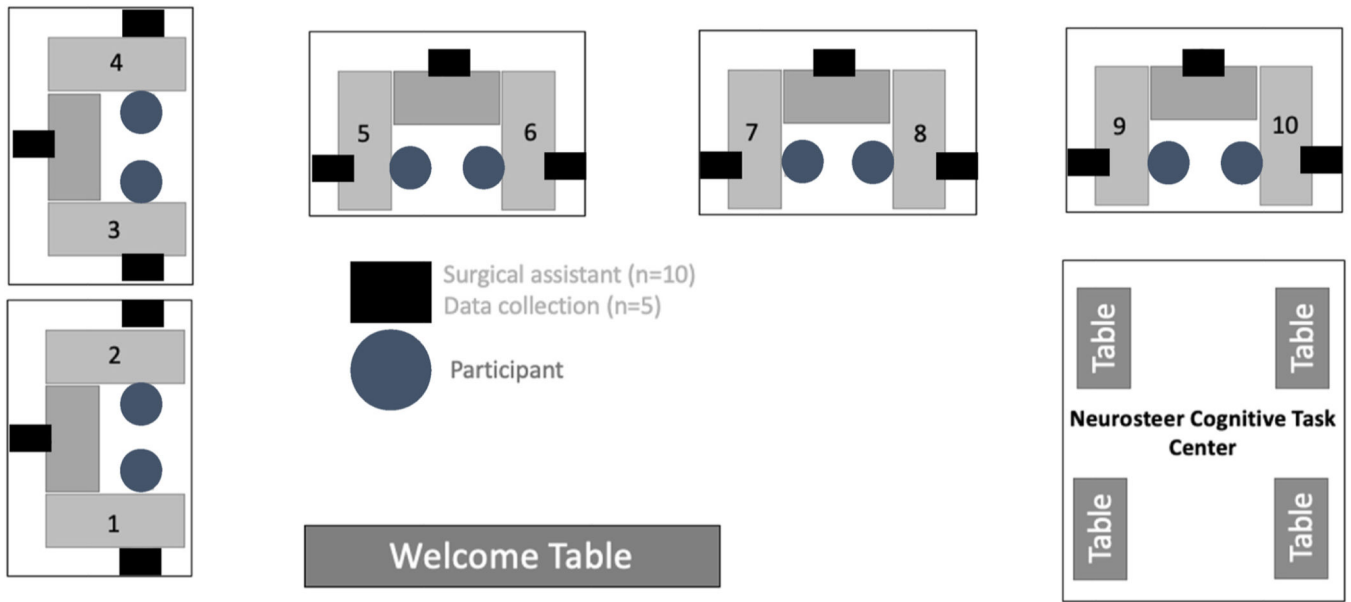


Fig. 2.
Surgical Skills Assessment Center set-up.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Demographic data of survey and interview participants.

Demographics	Survey Participants (n = 189)	Interview Participants (n = 23)
Age	29 to 77	35 to 70
Gender		
<i>Male</i>	142 (75.9%)	12 (54.5%)
<i>Female</i>	45 (24.1%)	10 (45.5%)
Years as a Practicing Surgeon		
<i>0 to 10</i>	65 (34.4%)	9 (40.9%)
<i>11 to 20</i>	60 (31.7%)	5 (22.7%)
<i>21</i>	64 (38.9%)	8 (36.4%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Survey responses.

Survey Responses	n (%)
Clinical workload of at least 75% (n = 183)	123 (67.2%)
Clinical time spent operating < 50% (n = 180)	120 (66.7%)
Interest in contributing data to longitudinal performance database (n = 188)	181 (96.3%)
Interest in tracking individual performance against database (n = 186)	178 (95.7%)
If reliable and confidential, interest utilizing testing to self-assess skills and judgement over time (n = 185)	181 (97.8%)
Ethical obligation to have skills assessment to protect the public (n = 181)	149 (82.3%)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Sample interview responses regarding feedback^a.

Interview Question	Themes Identified	Quotes
What made you stop by our booth to participate in the study?	Surgical Metrics Project presentation/ Word of mouth	The talk at the AWS convinced me to come by the booth. I like surgical simulation and education. This is a very interesting way to look at things.
	Curiosity	I came by out of curiosity of what this booth was all about.
	Contribution to research	I stopped by because I like surgical simulation and I want to contribute to research.
	Skills assessment	I am getting old and I think it is important to make sure that I am still okay. As the age, we can lose skills and does not know. I came by because it is important for individual performance to be evaluated and I wanted to see how it worked.
How do you think the [technical and cognitive skills test] went?	Went well/Confident in performance	It went well. I feel good. I am pretty sure that I won't have a leak.
	Issues with bowel quality/type	I do not think that it went that well. Technically, the bowel was too thin, it was deflated, it was difficult to see, it was too dark.
	Difficulties with assistant skill level	It was difficult working with a novice assistant.
What do you think about standardized testing for surgeons or peer evaluation for efficiency feedback or technical tips?	Desire for evaluation/feedback	I would like to have feedback. Even Tiger Woods has a coach. We do need to give it to each other. I am comfortable receiving it, I would like to know what I can do better. I do watch my peers operate, and I steal their tricks liberally. It is not a bad idea to offer feedback, but it can be tricky depending on the way that it is phrased. Certainly, residents can cross pollinate technical tips between attendings. It is important to give each other feedback. I do not think people will be receptive to it, however, I do think it is important to be evaluated and see feedback.
	No role for evaluation/feedback	There is no role for efficiency feedback. It is difficult to evaluate the skills of surgeons because what may work for me does not work for someone else. We can always learn from one another. What would probably be better is quality metrics and if these are not being met something is going on could be dangerous.
	Need for safe and supportive environment	A junior person should receive feedback in a safe environment. It should not be linked with other evaluations. It should be discussed beforehand that the feedback is being given only for constructive purposes.

ACS, American College of Surgeons.

^aWe received over 100 different responses/comments and suggestions. The above table represents some of the common themes that emerged.