Objective classification of residents based on their psychomotor laparoscopic skills

**Magdalena K. Chmarra · Stefan Klein ·
Joost C. F. de Winter · Frank-Willem Jansen ·
Jenny Dankelman**

Abstract

Background From the clinical point of view, it is important to recognize residents' level of expertise with regard to basic psychomotor skills. For that reason, surgeons and surgical organizations (e.g., Acreditation Council for Graduate Medical Education, ACGME) are calling for assessment tools that credential residents as technically competent. Currently, no method is universally accepted or recommended for classifying residents as "experienced," "intermediates," or "novices" according to their technical abilities. This study introduces a classification method for recognizing residents' level of experience in laparoscopic surgery based on psychomotor laparoscopic skills alone.

Methods For this study, 10 experienced residents (>100 laparoscopic procedures performed), 10 intermediates (10–100 procedures performed), and 11 novices (no experience) performed four tasks in a box trainer. The movements of the laparoscopic instruments were recorded with the TrEndo tracking system and analyzed using six motion analysis parameters (MAPs). The MAPs of all participants were submitted to principal component analysis (PCA), a data reduction technique. The scores of the first principal components were used to perform linear discriminant analysis (LDA), a classification method. Performance of the LDA was examined using a leave-one-out cross-validation.

Results Of 31 participants, 23 were classified correctly with the proposed method, with 7 categorized as experienced, 7 as intermediates, and 9 as novices.

Conclusions The proposed method provides a means to classify residents objectively as experienced, intermediate, or novice surgeons according to their basic laparoscopic skills. Due to the simplicity and generalizability of the introduced classification method, it is easy to implement in existing trainers.

M. K. Chmarra (✉) · J. C. F. de Winter · F.-W. Jansen ·
J. Dankelman
Department of BioMechanical Engineering,
Faculty of Mechanical, Maritime and Materials Engineering
(3mE), Delft University of Technology, Mekelweg 2,
2628 CD Delft, The Netherlands
e-mail: m.k.chmarra@tudelft.nl

S. Klein
Image Sciences Institute, University Medical Center Utrecht,
Utrecht, The Netherlands

S. Klein
Departments of Radiology and Medical Informatics,
Biomedical Imaging Group Rotterdam, Erasmus MC,
Rotterdam, The Netherlands

F.-W. Jansen
Department of Gynecology, Leiden University Medical Center,
Leiden, The Netherlands

A surgeon needs to acquire a certain level of manual dexterity to perform surgery safely. For this reason, it is important to assess and classify residents' skills objectively during training. Currently, several training simulators used to train basic laparoscopic skills are commercially available [1–3]. Additionally, various scoring methods for assessing laparoscopic skills have been developed [4–6]. However, a method to determine objectively whether a resident can be called experienced, intermediate, or novice according to his or her laparoscopic skill does not exist.

From the clinical point of view, it is important to recognize residents' level of expertise with regard to basic

psychomotor skills. Surgeons and surgical organizations (e.g., Acredition Council for Graduate Medical Education, ACGME) are calling for assessment tools that credential residents as technically competent [7–9]. Because medical education extends over the lifetime of the surgeon, those methods would support regular certification and monitoring of the resident's progress, recertification after a loss of privileges, and maintenance of certification [10].

According to the literature, multiple performance measures exist to assess basic psychomotor skills in laparoscopy [11–13]. It has been demonstrated that one performance measure alone does not adequately measure proficiency [14]. Therefore, assessment of psychomotor skills usually is performed with at least two different outcome measures [11–13]. However, when measures are used for examination (in effect to determine whether a resident is an experienced, intermediate, or novice surgeon), a passing score (threshold) needs to be defined [15]. It is difficult to determine when a score is sufficient.

First, assessment is performed using multiple assessment measures. This introduces the question of how to combine the assessment measures. Second, there is no "gold standard" of surgical competency against which the validity of a competency assessment can be judged [15]. Third, it is not clear whether a resident should be assessed on the basis of individual tasks or a composite assessment of all tasks [16].

Determining the group (e.g., experienced, intermediate, or novice) to which the resident belongs is called classification [17, 18]. In general, classification methods put individual objects into groups according to quantitative information about the characteristics ("features") of the objects. Linear discriminant analysis (LDA) is a commonly used statistical technique for data classification [17]. In this study, we investigate whether a classification method based on LDA can be used in laparoscopy to recognize residents' level of experience objectively (i.e., as experienced, intermediate, or novice) based on his or her psychomotor skills. The classification is based on a set of motion analysis parameters (MAPs), which are assessment scores derived from the instruments' motions during a series of exercises.

The classification method consists of two stages: training and classification. In the training phase, the classification method learns the distribution of MAPs for the experienced, intermediate, and novice categories. In the classification phase, the motion analysis data of a new resident are compared with the previously learned distributions of each group (experienced, intermediate, novice). The classification method then estimates the group to which the resident most likely belongs. The proposed method combines MAPs from different exercises without using pragmatic decisions and therefore avoids the need to establish a passing score manually for each MAP.

Materials and methods

Participants

For this study, 10 experienced gynecologists (>100 laparoscopic surgical procedures performed), 10 gynecologic residents (10–100 laparoscopic surgical procedures performed), and 11 medical students (no prior experience in laparoscopic surgery) were personally invited to participate in this study. Because no standard method for determining the level of expertise has been found in the literature, we decided to use the number of performed laparoscopic surgical procedures as a cutoff point when forming the "experienced," "intermediate," and "novice" groups. All the participants completed a short questionnaire detailing demographic information and prior experience in surgical laparoscopy.

Tasks

The four tasks selected for this study have been validated and are used regularly to train eye–hand coordination in the skills lab located in the Department of Gynecology at the Leiden University Medical Centre in the Netherlands [19, 20]:

- *Pipe cleaner*. This task required passing a pipe cleaner through four rings from the left to the right sides of the rings (Fig. 1A). The pipe cleaner had to be passed successively through all consecutive rings, starting from the left-most one. At the end, the pipe cleaner had to be removed from the rings.
- *Rubber band*. This task required stretching a rubber band around 16 nails tacked on a wooden board (Fig. 1B). There was no predefined order of nails around which the rubber band had to be stretched.
- *Beads*. This task required placing of 13 beads at predefined positions using the instrument with the left hand (Fig. 1C). The beads had to be placed in a specified order (see Fig. 1C). The task was completed when the beads formed a letter "B."
- *Circle*. This task required cutting a circle from a rubber glove stretched over 16 nails tacked on a wooden board (Fig. 1D). Each participant kept scissors in his or her dominant hand.

All tests were performed in a box trainer with a built-in TrEndo tracking system (Fig. 2) [21, 22]. We chose to use a box trainer instead of a virtual reality trainer because it has a more realistic force feedback. Our previous study [23] showed that the presence or absence of force feedback does influence psychomotor laparoscopic skills when the executed tasks require application of pulling and pushing forces (e.g., as in case of the rubber band and circle tasks).

To provide the same conditions for all the participants, the position of the equipment in the box trainer, the
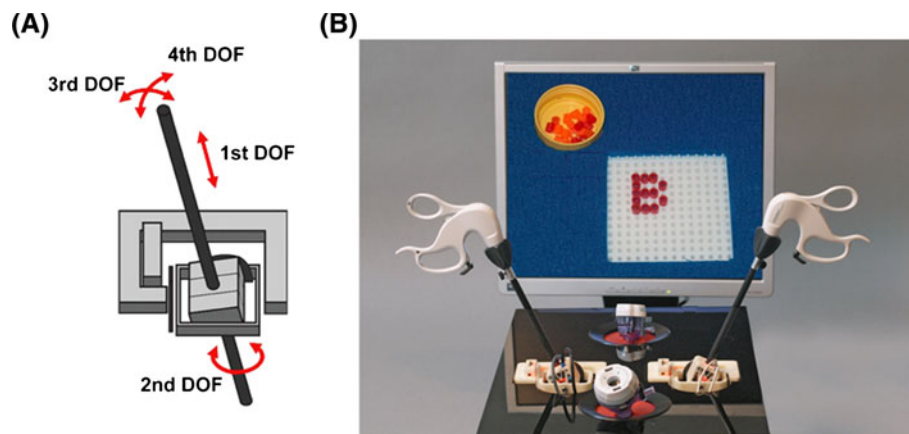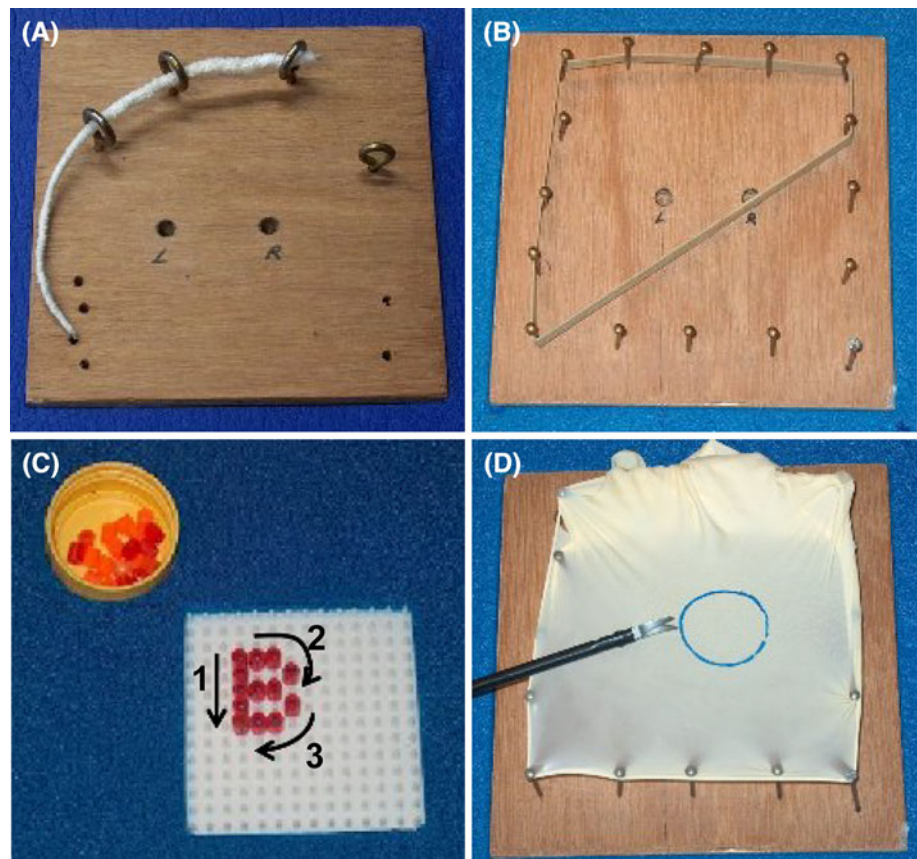
Fig. 2 The TrEndo tracking system for guiding and measuring real laparoscopic instruments in training setups. **A** A schematic drawing of the TrEndo tracking system. The TrEndo allows measurement and manipulation of the laparoscopic instrument in four degrees of freedom (DOFs): translation (first DOF) and rotation (second DOF) of the instrument around its axis, and left–right (third DOF) and forward–backward (fourth DOF) rotations of the instrument around the incision point. **B** A box trainer with a built-in TrEndo tracking system

instruments used, and the incision points for the camera and instruments in the box trainer were identical for everyone. The start and end positions of the laparoscopic instruments, predefined for each task, were the same for each participant. During the beads test, the participant held the camera in his or her right hand. During the pipe cleaner, rubber band, and circle tests, an assistant held the camera. The assistant was always the same. The image of a 0° laparoscope was presented on a monitor.

Motion analysis

The movements of the laparoscopic instruments were recorded with the TrEndo tracking system in four degrees of freedom (DOFs): translation (first DOF), rotation of the instrument around a longitudinal axis (second DOF), and left-right (third DOF) and forward-backward (fourth DOF) rotations of the instrument around the incision point (Fig. 2) [21, 24].

The recorded data were analyzed using six MAPs. The first four MAPs often are used to assess basic psychomotor skills in training setups [4, 11–13, 24]. These MAPs are related to time and the path traveled. The last two MAPs are new and have been introduced to measure how compact the instrument's tip movements are while performing tasks. The following six MAPs were chosen because time, path traveled, and compactness of the movement are important from the clinical point of view:

- *Time (T).* Total time taken to perform the task (in s).
- *Path length (PL).* Length of the curve described by the tip of the instrument while performing the task (in m).
- *Depth perception (DP), motion in depth.* Total distance traveled by the instrument along its axis (in m).
- *Motion smoothness (MS).* A motion analysis parameter based on the third time-derivative of position, which represents a change in acceleration (in m/s^3) [4, 24]. Before computation of motion smoothness, the raw data were filtered using a low-pass Butterworth filter.
- *Angular area (AA).* A motion analysis parameter related to the distances between the farthest positions of the instrument while performing a task. The angular area is defined as $AA = (Max[\alpha] - Min[\alpha]) \times (Max[\beta] - Min[\beta])$ (in rad^2), where α is the angular position of the instrument in the third DOF and β is the angular position of the instrument in the fourth DOF.
- *Volume (V).* A motion analysis parameter related to the distances between the farthest positions of the instrument while performing a task. The volume is defined as $V = AA * (Max[h] - Min[h])$ (in mm $\times$ rad^2), where h is the length of the part of the instrument inserted into the box trainer (first DOF).

For each participant, the MAPs for the left and right hands were averaged. Hence, a total of 24 MAPs (6 MAPs $\times$ 4 tasks) were obtained for each participant.

Statistical analysis

Explorative statistics on MAPs

Before applying the classification method (LDA), the descriptive statistics of the MAPs were explored. A Kruskal–Wallis test was used to compare all three groups. When

a significant difference between three groups was found, a Wilcoxon test was used to identify statistical differences between each pair of two groups. A *p* value less than 0.05 was considered to be statistically significant. The analysis was done using the Statistics Toolbox of MATLAB 7.

Classification

The study used LDA to automatically determine the threshold level for classifying a resident as experienced, intermediate, or novice according to his or her basic psychomotor skills. The LDA needs a training set consisting of MAPs acquired on participants with known laparoscopic skills levels. The LDA uses the training data to learn the distribution of MAPs for people belonging to each class, namely, experienced, intermediate, and novice. When MAPs of a new resident are provided, the LDA estimates the class to which the resident most likely belongs by comparing the MAPs of the new resident with the previously trained distribution.

Figure 3 illustrates the method for an imaginary case with two MAPs: path length and motion smoothness. The path length is on the horizontal axis, and motion smoothness is on the vertical axis. The E, I, and N symbols in the graph represent the experienced, intermediate, and novice classifications, respectively, in the training set. The distributions of the MAPs for the experienced, intermediate, and novice classifications result in a set of decision boundaries, which indicate the areas in the graph that belong to the experienced, intermediate, and novice categories.

With this information, classification of new residents based on their MAPs becomes straightforward. The asterisk (*) in the graph represents a new resident who did not belong to the training set. The location in the graph defined by its MAPs determines the classification result. In this example, the new resident is classified at the intermediate level.
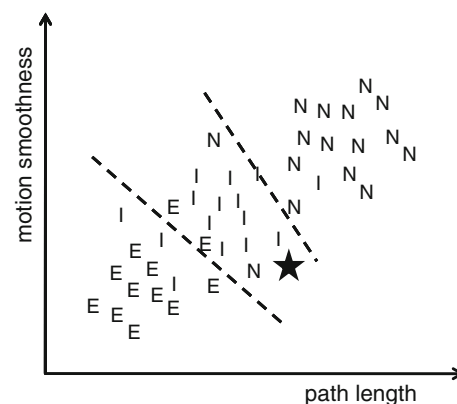
Fig. 3 An imaginary example of linear discriminant analysis (*LDA*) for experienced (*E*), intermediate (*I*), and novice (*N*) residents. The *lines* represent the decision boundaries

The input data of the algorithm are 24 MAPs (6 MAPs × 4 tasks) of the participants in the training set and of a new resident who needs to be classified. The algorithm started by normalizing the MAPs. Next, the number of MAPs was reduced from 24 to 6 by averaging the corresponding MAPs of each task (e.g., an average of the path lengths obtained from the pipe cleaner, rubber band, beads, and circle tasks).

To reduce the number of MAPs even further, we used the well-known principal component analysis (PCA) [18] using two principal components (see Appendix). The LDA was used in two ways: for each task separately and for all four tasks together (average of the corresponding MAPs from those four tasks).

Leave-one-out cross-validation

Performance of the classification methods was examined using a leave-one-out cross-validation [25, 26]. In each leave-one-out validation case, one participant was selected as the test case, and the remaining participants were used as a training set. This was repeated such that each participant was used once as a test case, which resulted in 31 leave-one-out validation cases in this study.

The results of the entire leave-one-out validation can be presented as a confusion matrix. The confusion matrix relates the ground truth classifications to the classifications predicted by the LDA.

Results

Explorative statistics on MAPs

The outcomes of the tests performed by experienced, intermediate, and novice residents are shown in Fig. 4 as box plots, with time, path length, depth perception, motion smoothness, angular area, and volume. The results of statistical analysis showed that there was no significant difference between the MAPs of experienced residents and those of intermediate residents. This was observed for all the tasks and MAPs.

Classification

The results of the LDA performed for each task separately and for all four tasks together are shown in Fig. 5. The best results were obtained for the combination of all the tasks together. Table 1. shows the results of this classification in a confusion matrix. In this matrix, the actual (ground truth) classifications are in the rows, and the classifications predicted by the LDA are in the columns. The values on the

diagonal of the matrix, in bold type, count the correct predictions of the LDA.

Discussion

The proposed classification method correctly classified 23 of 31 participants based on their motor dexterity. This is a good result, especially considering that explorative analysis of the MAPs showed a clear distinction between novices and the remaining participants, whereas the distinction between experienced and intermediate residents was more subtle. No significant differences were found between the MAPs of these two groups. Our method, however, was able to classify 23 participants correctly (7 as experienced, 7 as intermediate, and 9 as novice).

The proposed classification method has the possibility of classifying residents using a single training task as well as a combination of tasks. This means that residents can be classified based on their specific basic laparoscopic skills (e.g., force application) and on a combination of these skills (e.g., force application, cutting, and accuracy). Our results showed that a combination of different tasks produces better classification results.

Tests of the classification method were conducted using a leave-one-out cross-validation, a commonly accepted evaluation method [25, 26]. In this way, the classification method was trained using subjects different from those tested. Tests were performed using a small data set of MAPs for 31 participants. Most of the experienced laparoscopists (80%) who participated in the study performed the four tasks for the first time. This indicates that experienced residents were not necessarily "experts" in the tested tasks.

In contrast, all the intermediate residents had already performed the tasks before participating in the study. It could be expected that lack of experience in tested tasks may have a negative effect on the classification. Our method, however, was able to classify 23 of 31 participants correctly. We believe that this is a promising result because the classification was performed only on a set of psychomotor tasks, and all the results were cross-validated to eliminate overfitting. Moreover, some natural overlap between the skills of the three groups of residents is expected.

The results presented in Fig. 4 show that the equipment used in the study had the ability to discriminate novices from experienced residents, and novices from intermediate residents. There were, however, no significant differences between the MAPs of intermediate residents and experienced residents. Still, our method was sufficiently sensitive to distinguish between experienced and intermediate residents. Therefore, a method that can correctly classify 74% of residents in the three categories is believed to be quite strong and realistic concerning what practitioners should

Fig. 4 Results of the pipe cleaner, rubber band, beads, and circle tests. The results are presented as notched box and whisker plots, in which every box has a line at every quartile, median, and upper quartile value. The whiskers are presented as lines that extend from each end of the box to show the extent of the remaining data. The notches represent the 95% confidence interval for the median. The boxes whose notches do not overlap are significantly different ($p < 0.05$). A few extreme outliers are excluded from the plots to omit excessive compression of the y-axis. The data of the pipe cleaner (task P), rubber band (task R), beads (task B), and circle (task C) are shown separately and after averaging (all tasks). E experienced, I intermediate, N novice
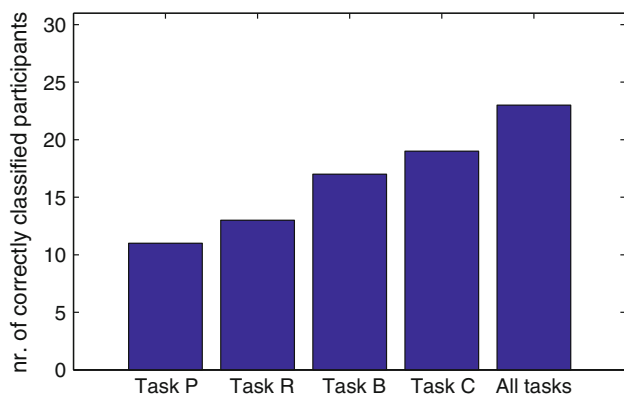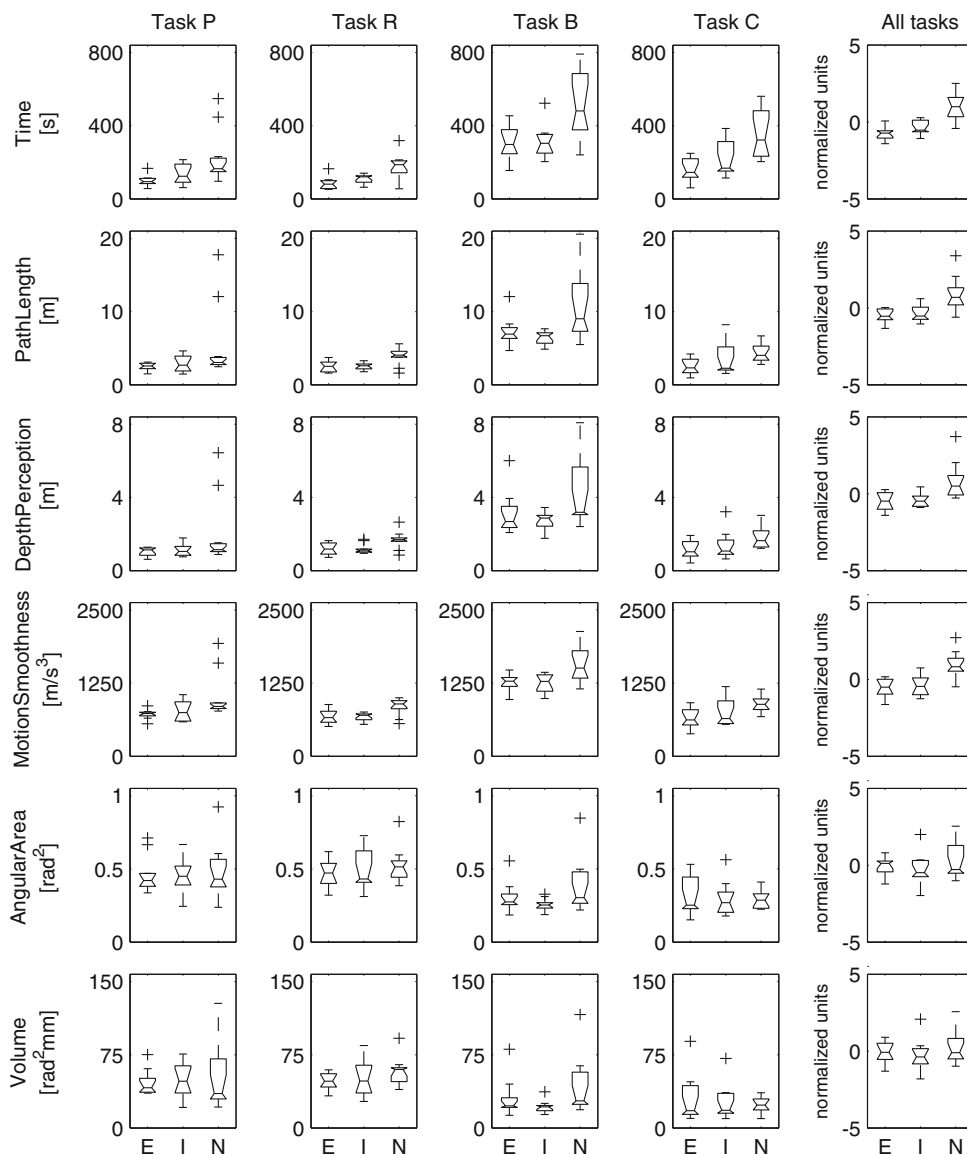
Fig. 5 Results of the classification. Task P (pipe cleaner), task R (rubber band), task B (beads), task C (circle), all tasks (averaged tasks P, R, B, and C)

Table 1 Confusion matrix for linear discriminant analysis (LDA) performed for the combination of the four (averaged) tasks together

		Predicted		
		Experienced	Intermediate	Novice
Actual	Experienced	**7**	3	0
	Intermediate	2	**7**	1
	Novice	1	1	**9**

expect from classification based on psychomotor skills only.

We found no other studies that were able to distinguish between three experience levels with this degree of sensitivity and specificity. Further work on the method should include more validation studies with a larger number of participants,

different tasks, and different MAPs. The availability of a large number of participants would yield a larger training set, which could improve the classification result.

For the clarity of the report, the proposed classification method was tested using only motion analysis–based parameters without any reference to the quality of the task performed (e.g., number of errors). Therefore, a study to investigate whether the addition of quality-based scores improves classification of residents is recommended. Moreover, further studies should consider other surgical skills (e.g., theoretical knowledge) as an input for the classification method.

Defining a threshold level that allows residents to start their clinical program is not trivial because no accepted set of criteria exists to define a "proficient surgeon." Therefore, we investigated whether a classification method could be used to recognize the experience category of residents based on their psychomotor skills. Although experience does not always reflect the level of actual skills and expertise (see the work of Duncan et al. [27] for an analysis in the context of learning to drive a car), we believe that our method represents an important step toward development of the needed (norm-referenced) proficiency criteria.

From the clinical point of view, it is necessary to find appropriate parameters that can measure the quality of actions for an objective evaluation of residents' operative skills. Only with correct parameters and their combination will it be possible to provide information about the level of operative skill.

In this study, we introduced a method for objective classification of residents as experienced, intermediate, and novice surgeons according to their basic laparoscopic skills. With this method, information from various MAPs calculated for different tasks is integrated without the need for any user-defined weighting factors. The classification is based on a set of training examples (residents with known laparoscopic skills), so the problem of establishing the threshold level (passing score) is solved.

The introduced classification method is rather basic, and because of its simplicity, it is used commonly in statistical pattern recognition [17]. For implementation of this method in the training of basic laparoscopic skills, only input data (e.g., motions of the instruments) and a software module are needed. Current virtual reality trainers, for example, record movements of the instruments. Therefore, implementation of a classification method that includes only some changes in the software could easily be included in the software.

Review of the literature shows a lack of methods that distinguish between residents of varying experience, especially when fine gradations in experience must be detected. Cotin et al. [4] introduced a standardized score based on the information from five MAPs. The standardized score was used to classify residents according to their basic laparoscopic skills. The classification was based on the training data of experienced residents only and did not account for the distribution of MAPs for intermediates or novices. Moreover, validation of the standardized score included comparison of only experienced and novice residents.

In contrast to Cotin et al. [4], our method classifies residents based on a training set that includes data of experienced, intermediate, and novice residents. Therefore, our method is able to distinguish between residents with fine gradation in experience (e.g., experienced and intermediate residents).

Another difference between our work and that of Cotin et al. [4] is that our method does not use any user-defined weighting factors, whereas the standardized score does. With our method, the whole process is data driven. The threshold corresponds to the decision boundary, which is automatically and optimally determined based on the training data. This makes our classification method simpler and more general than the standardized score introduced by Cotin et al. [4].

In practice, simpler methods are preferred because of their generalizability and ease of implementation. Fundamentals of Laparoscopic Surgery (FLS) is a program developed by the members of SAGES. With FLS, the technical skills of surgeons are assessed based on the McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) score [28, 29]. This score is calculated for an individual task, taking only time and accuracy (e.g. number of errors) into account [30]. Therefore, the MISTELS score does not provide information about the actual psychomotor laparoscopic skills of a resident, in contrast to our method that classifies residents based on their MAPs.

Any attempt to assess and classify the technical competence of a resident objectively is difficult for two reasons. There is no clear definition of "competence" [15], and operative skill is a combination of a resident's knowledge, judgment, and technical ability [31]. Technical proficiency, nonetheless, seems fundamental to performing surgery safely [15]. This is especially apparent in laparoscopy, which is performed in a limited working area with limited tactile perception and difficult handling of the instruments.

Our classification method can determine the group to which the resident belongs according to his or her basic laparoscopic skills. It can provide an aid for deciding whether the resident is ready to operate on patients, which is a very important aspect when patient safety is considered. Due to the simplicity and generalizability of our method, it should be easy to implement also in current virtual reality trainers. The described classification method could be used for certification and monitoring of a

Table 2 The correlation matrix for the four (averaged) tasks together

	Time	Path length	Depth perception	Motion smoothness	Angular area	Volume
Time	1					
Path length	0.92	1				
Depth perception	0.82	0.96	1			
Motion smoothness	0.95	0.96	0.87	1		
Angular area	0.49	0.63	0.61	0.57	1	
Volume	0.50	0.68	0.69	0.60	0.95	1

resident's progress, with additional motion analysis used to give feedback on the nature of possible resident limitations that need to be improved. An interesting extension of this work involves using the classification framework to identify residents who will not be able to acquire the necessary basic psychomotor laparoscopic skills. This can be done by analyzing patterns of skills acquisition.

Appendix

Principal component analysis

Principal component analysis (PCA) involves calculating an eigenvalue decomposition of the correlation matrix of the MAPs. Given a set of input features (in our case, the six averaged MAPs), PCA extracts a set of "principal components," each of which is a linear combination of the input features:

$$PC = w_0 \times T + w_1 \times PL + w_2 \times DP + w_3 \times MS + w_4 \times AA + w_5 \times V,$$

where $w_0, \ldots, w_5$ serve as weights, T is time, PL is path length, DP is depth perception, MS is motion smoothness, AA is angular area, and V is volume.

The principal components are ordered according to the amount of variance they explain. The first component accounts for as much of the variance in the data as possible, and each succeeding component accounts for as much of remaining variance as possible. Then the cumulative variance explained (CVE) forms a criterion for selecting the number of principal components to be used in the linear discriminant analysis (LDA).

In this study, the CVE showed that the first two principal components accounted for more than 90% of the variance. The LDA was thus performed with the first two principal components.

The correlation matrix for the four (averaged) tasks is shown in Table 2. This correlation matrix was calculated with the data of all 31 participants. The results indicate that time, path length, depth perception, and motion smoothness were positively correlated. They were partially redundant. A positive correlation also existed between angular area and volume. The high correlations indicate that the use of PCA is justified and that the data can be well described by only a few principal components.

References

1. Halvorsen FH, Elle OJ, Fosse E (2005) Simulators in surgery. Minim Invasiv Ther 14:214–223
2. Schijven M, Jakimowicz J (2003) Virtual reality surgical laparoscopic simulators: how to choose. Surg Endosc 17:1943–1950
3. Sutherland LM, Middleton PF, Anthony A, Hamdorf J, Cregan P, Scott D, Maddern GJ (2006) Surgical simulation: a systematic review. Ann Surg 243:291–300
4. Cotin S, Stylopoulos N, Ottensmeyer M, Neumann P, Rattner D, Dawson S (2002) Metrics for laparoscopic skill trainers: the weakest link!. Lect Notes Comput Sci 2488:35–43
5. Martin JA, Regehr G, Reznick R, Macrae H, Murnaghan J, Hutchison C, Brown M (1997) Objective structured assessment of technical skills (OSATS) for surgical residents. Br J Surg 84:273–278
6. Moorthy K, Munz Y, Dosis A, Bello F, Darzi A (2003) Motion analysis in the training and assessment of minimally invasive surgery. Minim Invasiv Ther 12:137–142
7. Park AE, Witzke D (2002) The surgical competence conundrum. Surg Endosc 16:555–557
8. Roberts KE, Bell RL, Duffy AJ (2006) Evolution of surgical skills training. World J Gastroenterol 12:3219–3224
9. Ritchie WP (2004) Basic certification in surgery by the American Board of Surgery (ABS). What does it mean? Does it have value? Is it relevant? A personal opinion. Ann Surg 239:133–139
10. Satava RM (2006) Assessing surgery skills through simulation. Clin Teacher 3:107–111
11. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Papasavas P, Dosis A, Bello F, Darzi A (2007) An evaluation of the feasibility,

validity, and reliability of laparoscopic skills assessment in the operating room. Ann Surg 245:992–999

12. Moorthy K, Munz Y, Dosis A, Bello F, Chang A, Darzi A (2004) Bimodal assessment of laparoscopic suturing skills: construct and concurrent validity. Surg Endosc 18:1608–1612

13. Stefanidis D, Haluck R, Pham T, Dunne JB, Reinke T, Markley S, Korndorffer JR, Arellano P, Jones DB, Scott J (2007) Construct and face validity and task workload for laparoscopic camera navigation: virtual reality versus videotrainer systems at the SAGES Learning Center. Surg Endosc 21:1158–1164

14. Smith CD, Farrell TM, McNatt SS, Metreveli RE (2001) Assessing laparoscopic manipulative skills. Am J Surg 181:547–550

15. Feldman LS, Sherman V, Fried GM (2004) Using simulators to assess laparoscopic competence: ready for widespread use? Surgery 135:28–42

16. Darzi A, Datta V, Mackay S (2001) The challenge of objective assessment of surgical skills. Am J Surg 181:484–486

17. Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition: a review. IEEE Trans Pattern Anal Machine Intell 22:4–38

18. Duda RO, Hart PE, Stork DG (2001) Pattern classification, 2nd edn. Wiley, New York

19. Kolkman W, van de Put MAJ, Wolterbeek R, Trimbos JBMZ, Jansen FW (2008) Laparoscopic skills simulator: construct validity and establishment of performance standards for residency training. Gynecol Surg 5:109–114

20. Kolkman W, van de Put MAJ, van den Hout WB, Trimbos JBMZ, Jansen FW (2007) Implementation of the laparoscopic simulator in a gynecological residency curriculum. Surg Endosc 21:1363–1368

21. Chmarra MK, Bakker NH, Grimbergen CA, Dankelman J (2006) TrEndo, a device for tracking minimally invasive surgical instruments in training setups. Sensor Actuat A-Phys 126:328–334

22. Chmarra MK, Grimbergen CA, Dankelman J (2007) Systems for tracking minimally invasive instruments. Minim Invasiv Ther 16:328–340

23. Chmarra MK, Dankelman J, van den Dobbelsteen JJ, Jansen FW (2008) Force feedback and basic laparoscopic skills. Surg Endosc 22:2140–2148

24. Chmarra MK, Kolkman W, Jansen FW, Grimbergen CA, Dankelman J (2007) The influence of experience and camera holding on laparoscopic instrument movements measured with the TrEndo tracking system. Surg Endosc 21:2069–2075

25. de Bruine M, van Ginneken B, Viergever MA, Niessen WJ (2004) Interactive segmentation of abdominal aortic aneurysms on CTA images. Med Image Anal 8:127–138

26. Fan Y, Rao H, Hurt H, Giannetta J, Korczykowski M, Shera D, Avants BB, Gee JC, Wang J, Shen D (2007) Multivariate examination of brain abnormality using both structural and functional MRI. NeuroImage 36:1189–1199

27. Duncan J, Williams P, Brown I (1991) Components of driving skills: experience does not mean expertise. Ergonomics 34:919–937

28. Ritter EM, Scott DJ (2007) Design of a proficiency-based skills training curriculum for the fundamentals of laparoscopic surgery. Surg Innov 14:107–112

29. Peters JH, Fried GM, Swanstrom LL, Soper NJ, Sillin LF, Schirmer B, Hoffman K (2004) Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. Surgery 135:21–27

30. Vassiliou MC, Ghitulescu GA, Feldman LS, Stanbridge D, Leffondré K, Sigman HH, Fried GM (2006) The MISTELS program to measure technical skills in laparoscopic surgery. Surg Endosc 20:744–747

31. Dankelman J, Chmarra MK, Verdaasdonk EGG, Stassen LPS, Grimbergen CA (2005) Fundamental aspects of learning minimally invasive surgical skills. Minim Invasiv Ther 14:247–256