

# Development and implementation of the Structured Training Trainer Assessment Report (STTAR) in the English National Training Programme for laparoscopic colorectal surgery

Susannah M. Wyles<sup>1</sup> · Danilo Miskovic<sup>2</sup> · Zhifang Ni<sup>3</sup> · Ara W. Darzi<sup>1</sup> · Roland M. Valori<sup>4</sup> · Mark G. Coleman<sup>5</sup> · George B. Hanna<sup>6</sup>

Received: 17 March 2015 / Accepted: 19 May 2015 / Published online: 24 June 2015  
© Springer Science+Business Media New York 2015

## Abstract

**Background** There is a lack of educational tools available for surgical teaching critique, particularly for advanced laparoscopic surgery. The aim was to develop and implement a tool that assesses training quality and structures feedback for trainers in the English National Training Programme for laparoscopic colorectal surgery.

**Methods** Semi-structured interviews were performed and analysed, and items were extracted. Through the Delphi process, essential items pertaining to desirable trainer characteristics, training structure and feedback were determined. An assessment tool (Structured Training Trainer Assessment Report—STTAR) was developed and tested for feasibility, acceptability and educational impact.

**Results** Interview transcripts (29 surgical trainers, 10 trainees, four educationalists) were analysed, and item lists

created and distributed for consensus opinion (11 trainers and seven trainees). The STTAR consisted of 64 factors, and its web-based version, the mini-STTAR, included 21 factors that were categorised into four groups (training structure, training behaviour, trainer attributes and role modelling) and structured around a training session timeline (beginning, middle and end). The STTAR (six trainers, 48 different assessments) demonstrated good internal consistency ( $\alpha = 0.88$ ) and inter-rater reliability (ICC = 0.75). The mini-STTAR demonstrated good inter-item reliability ( $\alpha = 0.79$ ) and intra-observer reliability on comparison of 85 different trainer/trainee combinations ( $r = 0.701$ ,  $p = <0.001$ ). Both were found to be feasible and acceptable. The educational report for trainers was found to be useful (4.4 out of 5).

**Conclusions** An assessment tool that evaluates training quality was developed and shown to be reliable, acceptable and of educational value. It has been successfully implemented into the English National Training Programme for laparoscopic colorectal surgery.

✉ George B. Hanna  
g.hanna@imperial.ac.uk

Susannah M. Wyles  
susannah.wyles@imperial.ac.uk

<sup>1</sup> Department of Surgery and Cancer, St Mary's Hospital, Imperial College London, London, UK

<sup>2</sup> St James University Hospital, Leeds, UK

<sup>3</sup> National Training Programme, Department of Surgery and Cancer, St Mary's Hospital, Imperial College London, London, UK

<sup>4</sup> Department of Gastroenterology, Gloucestershire Hospitals NHS Foundation Trust, Gloucester, UK

<sup>5</sup> National Training Programme, Derriford Hospital, Plymouth, UK

<sup>6</sup> National Training Programme, Department of Surgery and Cancer, St Mary's Hospital, Imperial College London, Praed St, London W2 1NY, UK

**Keywords** Training · Assessment · Surgery · Education · Feedback · Teaching

Surgeons execute and teach advanced laparoscopic surgery in different ways. There is no clear consensus about the “right” approach to structure the training episodes [1]. However, the provision of high-quality teaching is imperative to maximise the benefits from training opportunities as trainers have been shown to impact differently trainees' performance [2]. This is of particular importance with the current challenges facing surgical training. Teaching surgical techniques requires non-technical skills as communication, and teamwork have been

shown to influence surgical performance [3–5]. For a trainee to learn a practical skill effectively, the trainer needs to ensure that the trainee maintains motivation, continues to progress and does not suffer from cognitive overload whilst remaining mindful of patient safety and theatre productivity [6]. However, there are limited mechanisms for a trainee to provide feedback on training characteristics [7–9] in either simulation courses or clinical environment. Within some specialties of medicine, assessment of teaching quality has been attempted; however, none have been designed for peers or senior trainees to utilise within surgery [10–12].

Laparoscopic colorectal surgery is technically demanding. The long learning curve prohibits a self-taught approach with evidence that supervised teaching promotes a better patient outcome [13]. The National Training Programme in Laparoscopic Colorectal Surgery of England (NTP) was a multi-centre educational scheme with the aim to shorten the learning curve whilst minimising patient morbidity and mortality. It provided a competency-orientated, structured and supervised laparoscopic training to colorectal specialists in England [14, 15]. The trainers within the programme were experienced surgeons deemed competent in the operative techniques and appointed through an application process and peer review. There was no known prerequisite or method for the objective assessment of the quality of the education and training provided by these individuals [16]. Hence, this educational environment provided an ideal model to investigate methods to evaluate training quality for advanced laparoscopic surgery.

The overall aim of the present study was to develop and implement a tool that assesses training quality and structures feedback for trainers in the English National Training Programme in laparoscopic colorectal surgery. In order to achieve this aim, the specific objectives were to (1) identify the characteristics deemed important in a good surgical trainer, (2) determine a structure for each advanced laparoscopic surgery training session, (3) create a structured formative trainer assessment tool, (4) study the reliability of this tool, (5) implement the tool in the national training programme and (6) determine its feasibility, acceptability and educational impact.

## Materials and methods

Given the lack of evidence on the subject, an approach based on grounded theory was taken, i.e. minimal assumptions were made, and the framework was determined from the data through an iterative process [17]. The methodology for this study can be divided into the following stages.

### **Determining the characteristics of a good surgical trainer and components of the training structure in advanced laparoscopic surgery**

Two interviewers (SMW and DM) performed semi-structured interviews with surgical trainers, trainees and educationalists to determine the characteristics deemed important for a good surgical trainer, for a training session and for feedback. Both interviewers transcribed and analysed the data, and derived themes. Using NVivo, free nodes were created, and using a constant comparative method of analysis, a framework was developed. The analysis process was verified by two psychologists trained and experienced in the use of NVivo (NVivo, QSR International, Cheshire, UK) [18]. Item lists were developed using those nodes from the analysis that pertained to good trainer characteristics, training structure and reflection. Twenty surgeons were invited to participate in the consensus process via e-mail communication, aiming for 10–15 trainers and 5–10 trainees from different training regions throughout England, in order to achieve a balanced, broad perspective on the selection of items. Consensus was sought through the application of the Delphi technique [19]. The items were listed in alphabetical order to prevent any unintentional interpretable ranking of importance. There were also free text boxes next to each item to allow for the study participants to provide additional input regarding that item. Opinion regarding how important the responder felt each item was for each training session was ranked using a seven-point Likert scale: 1 = very unimportant, 2 = moderately unimportant, 3 = unimportant, 4 = undecided, 5 = important, 6 = moderately important and 7 = very important. Items scoring six or higher were considered to be “essential”, and those >5 “desirable”.

### **Development of the detailed Structured Training Trainer Assessment Report (STTAR)**

Having determined the importance of the different items, an assessment tool was created using items with a score >5. Those scoring <4 were omitted. Items with scores between four and five were reviewed with a low threshold for initial inclusion with potential subsequent exclusion with future iterations of the form. The separate items were printed and organised through a “brown paper” technique—a process-mapping technique that allows a group of people to interact and order separate components of a process—to enable a logical timeline to be applied to the structure of the form to help facilitate ultimate prospective completion [20]. The format and content of the form, the scoring system and the accompanying instruction sheet and “dictionary” were piloted in both the clinical and course environment, and each iteration of the form was reviewed by an expert panel

(two psychologists, two surgeons, two surgeons with educational degrees, two educationalists). Having agreed the overall structure, the form was then piloted more formally in order to test different scoring systems. Items were altered if they were impossible to score and combined where they appeared to be overlap.

### **Reliability and feasibility of using STTAR in “Train the Trainer laparoscopic colorectal surgery course”**

Reliability and feasibility of STTAR took place during the NTP “Train the Trainer laparoscopic colorectal surgery course” (Lapco TT). During day two of this course, course delegates (LCS trainers) teaching LCS to a trainee are observed directly by another course delegate (assessor) over a period of twenty minutes, who then, with faculty members, provides feedback on their teaching performance [21]. The training episode is also observed indirectly by the other course participants and faculty who can provide additional feedback. This setting was chosen as it was as controlled an environment as possible with respect to clinical variables: i.e. the theatre and staff were prepared and set up for training; the same trainee was trained by several different delegates; for a fixed amount of time on the same cases; and in the same theatre using the same equipment. All course delegates completed a short questionnaire regarding prior experience and demographic data. All those who observed the training episode, either directly or indirectly, including the assessor and faculty members, completed a STTAR for each training episode.

### **Development of a web-based evaluation system, the “mini-STTAR”**

Using the data collected from the interview and Delphi process, all essential items (those scoring >6) were selected. A formative assessment form was then devised that would allow trainees to give trainers feedback immediately after a training episode. This form was reviewed by the same expert panel and two trainee surgeons who were experienced in laparoscopic surgery. Items were then developed within the form and pretested in a pilot study in both a clinical and course setting. Those items that were not understood or duplicates were re-written or removed. A five-point Likert scale that pertained to the degree that the trainee felt each item occurred was devised (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree). The order of some of the items was altered to improve accuracy of completion and to prevent the same score from being given blindly to each item [22]. The final version was reviewed and approved by the educational training committee, and the approved iteration was transformed into a web-based electronic form.

### **Implementation of the mini-STTAR into a national training programme**

Trainees completed the electronic form after each NTP training case within NTP-approved training environments over a period of 12 months. Feasibility was assessed through the mean time taken to complete a form, and acceptability through analysis of a Likert scale (1 = most unacceptable, 2 = unacceptable, 3 = neutral, 4 = acceptable, 5 = most acceptable [(median/interquartile range (IQR)]. In addition, each trainer who had received more than four assessments was provided with his/her educational report that represented the evaluation of training quality during this period. Trainers were asked about the usefulness of this report (1 = not at all useful, 2 = quite useful, 3 = undecided, 4 = useful, 5 = very useful) and how likely the report would change their training behaviour and strategy.

### **Reliability of STTAR and mini-STTAR**

Inter-rater reliability was calculated using intra-class correlation coefficient (ICC), and inter-item reliability with Cronbach’s alpha. For the mini-STTAR, test–retest and intra-rater reliability were possible to analyse where the trainee and trainers performed more than one case together (intra-class correlation coefficient and Pearson correlation). A lack of a prior gold standard meant criterion validity could not be assessed for either assessment.

## **Results**

### **Determining the characteristics of a good surgical trainer and components of the training structure in advanced laparoscopic surgery**

Forty-five participants were invited to take part via e-mail. Of the 43 who participated, semi-structured interviews took place over a 10-month period, between 29 surgical trainers from different surgical specialties, six NTP consultant trainees, four senior surgical trainees who were all experienced in advanced laparoscopic surgery, four educationalists and two researchers trained in interview techniques. The participants were from centres across England and included two Canadian surgeons who had extensive experience in devising training programmes and in teaching advanced laparoscopic surgery [1]. The interviews in the majority were conducted over the telephone at a time of the participant’s convenience. Some field notes were taken during the interviews to act as reminders for the interviewer to draw on interesting points made by the participant at an opportune moment, rather than immediately, so

as not to interrupt the flow of the interview [23]. The questions were therefore developed as analysis cycles took place and emergent themes became evident. Sampling continued until “saturation” was achieved [17]. The mean length of interview was 17 min (6 min 08 s–42 min 28 s), getting progressively shorter over the duration of the study as less new information was produced. There was excellent inter-rater agreement between the interviewers item and theme analysis (Cohen’s  $\kappa = 0.92$ ).

From the interview analysis, 66 items were created that related to “good trainer/characteristics”, for “reflection” there were 26 different items and 96 for “training technique”. The list was distributed to 11 surgical trainers and seven senior trainees, who were all able to perform both open and advanced laparoscopic surgery (uptake rate 90 %). Consensus was reached after two rounds of the Delphi process, with an overall reduction in the SD between round 1 and 2. When comparing the ratings given by trainers and trainees, none of the good trainer characteristics or reflection items had >1 point of difference between the standard deviations of the perceived importance. For the training technique items, however, there were many training technique items that had >1 point of difference between the standard deviations, but this reduced for all items between round 1 and 2 (“Appendix”).

### Development of detailed Structured Training Trainer Assessment Report (STTAR)

This process involved 23 iterations of the assessment form. The initial list included 150 items: good trainer/expert characteristics 46 items, 12 for reflection and 76 for training episode. These were allocated loosely into three groups according to the three periods of a training episode (“set”, “dialogue” and “closure”) [6]. Set referred to the beginning of a training episode where a conversation takes place between the trainer and trainee and an agreement made as to the allocation of the different parts of the practical part of the training. Dialogue refers to the actual training—i.e. the interactions between trainer and trainee during the procedure. Closure describes the feedback, reflection and conclusion of the episode and formulation of learning points. For practical reasons, dialogue was further separated into a subsection for challenging situations within the case, in order to monitor the adjustment in the training if any. The different items within the three groups were then further subdivided into “structure”, “teaching behaviour”, “attributes” and “role model”. A further eight items were added from the <5 groups: two for reflection (one because of a double negative in the wording and scoring system) and six for training technique. The expert panel then analysed the items, and where possible, more coherent summary words were created for those items with

overlapping meanings. An assessment form of four groups of 16 different items making a total of 64 items was developed. A “dictionary” to explain the meaning of each of the different points was created to minimise misinterpretation of the summary terms and to aid the assessors using the form (“Appendix”). A further addition was to change one of the attribute items to being “comfortable in silence”. This item was added after re-reading the interview transcripts, and through piloting the forms within the operating theatre, as it became apparent that during the training episode, it was not always necessary for the trainer to speak [24].

In the development of the scoring system to measure the extent that each item was demonstrated, a binary process of “yes” and “no” was not acceptable as on piloting, observers felt that there were different “shades” of yes and no. Thus, a seven-point Likert scale was developed (1 = never, 2 = hardly ever, 3 = a bit, 4 = neutral, 5 = occasionally but could have done more, 6 = almost as much as possible, 7 = every opportunity, N/A = not applicable). The form also had a space under each item where the user could tally down each time the event happened to enable the form to be completed accurately prospectively. The final iteration also used a visual analogue scale to enable the rater to mark down where they felt the observer trainer scored in the attribute and role model sections as these were the sections that others found difficult to rate (Fig. 1). A cover sheet was developed to enable baseline information to be gleaned regarding the training partnership and the specific case: procedure, number performed in the past, number performed with this trainer, case difficulty (1–6, with 1 being very easy and 6 very difficult which was in keeping with the other NTP forms) and brief anonymous details regarding the patient. An instruction sheet was also developed to aid completion of the form by the users. Piloting demonstrated that this was sufficient and that no formal training session was required.

### Reliability and feasibility of STTAR

Six male NTP trainers took part in a Bradford training centre Lapco TT course. They were aged between 41 and 54. All had performed >100 laparoscopic colorectal resections. Forty-eight STTAR forms were completed over the day, with the range of scores being 19.62–22.41, of a possible 28 (Table 1). The different scores obtained by the six trainers in “structure”, “training behaviour”, “attributes” and “role model” demonstrated a globally lower score in the role model section (Table 1). Inter-item reliability had a Cronbach’s  $\alpha$  of 0.88, which did not change if items were removed. There was good inter-rater reliability [ICC 0.75 (95 % CI 0.63, 0.841),  $F$  4.001,  $p < 0.000$ ]. The

		Structure		Teaching Behaviour		Attributes		Role Model					
		Training Structure		Training behaviour during case		Characteristics demonstrated		Technical and non-technical skills					
		1-7	N/A	1-7	N/A	1-7	N/A	1-7	N/A				
SET		Contextual conversation	<input type="checkbox"/>	<input type="checkbox"/>	Ground rules	<input type="checkbox"/>	<input type="checkbox"/>	Motivated	<input type="checkbox"/>	<input type="checkbox"/>	Communication with team	<input type="checkbox"/>	<input type="checkbox"/>
		Define aims	<input type="checkbox"/>	<input type="checkbox"/>	Knowledge	<input type="checkbox"/>	<input type="checkbox"/>	Confident	<input type="checkbox"/>	<input type="checkbox"/>	Takes control	<input type="checkbox"/>	<input type="checkbox"/>
		Align agendas	<input type="checkbox"/>	<input type="checkbox"/>	Concerns	<input type="checkbox"/>	<input type="checkbox"/>	Insight into ability	<input type="checkbox"/>	<input type="checkbox"/>	Ensures patient safety	<input type="checkbox"/>	<input type="checkbox"/>
		Environment preparation	<input type="checkbox"/>	<input type="checkbox"/>	Case-specific	<input type="checkbox"/>	<input type="checkbox"/>	Non-threatening	<input type="checkbox"/>	<input type="checkbox"/>	Foresight	<input type="checkbox"/>	<input type="checkbox"/>
DIALOGUE	Standard part of procedure	Aims focused	<input type="checkbox"/>	<input type="checkbox"/>	Guiding verbal input	<input type="checkbox"/>	<input type="checkbox"/>	Approachable	<input type="checkbox"/>	<input type="checkbox"/>	Competence	<input type="checkbox"/>	<input type="checkbox"/>
		Ability matched task	<input type="checkbox"/>	<input type="checkbox"/>	Questioning/ option generation	<input type="checkbox"/>	<input type="checkbox"/>	Articulate	<input type="checkbox"/>	<input type="checkbox"/>	Strategic	<input type="checkbox"/>	<input type="checkbox"/>
		Deconstruction	<input type="checkbox"/>	<input type="checkbox"/>	Encouraging, positive reinforcement	<input type="checkbox"/>	<input type="checkbox"/>	Listens	<input type="checkbox"/>	<input type="checkbox"/>	Knowledgeable	<input type="checkbox"/>	<input type="checkbox"/>
		Accessible demonstration	<input type="checkbox"/>	<input type="checkbox"/>	Corrective feedback	<input type="checkbox"/>	<input type="checkbox"/>	Patient	<input type="checkbox"/>	<input type="checkbox"/>	Patient-focused	<input type="checkbox"/>	<input type="checkbox"/>
	Particularly in difficult part	Stretch (allow to struggle)	<input type="checkbox"/>	<input type="checkbox"/>	Warning verbal input	<input type="checkbox"/>	<input type="checkbox"/>	Calm	<input type="checkbox"/>	<input type="checkbox"/>	Excellent decision making	<input type="checkbox"/>	<input type="checkbox"/>
		Informing	<input type="checkbox"/>	<input type="checkbox"/>	Strategy justification	<input type="checkbox"/>	<input type="checkbox"/>	Comfortable in silence	<input type="checkbox"/>	<input type="checkbox"/>	Leader	<input type="checkbox"/>	<input type="checkbox"/>
		Take over when appropriate	<input type="checkbox"/>	<input type="checkbox"/>	Directing verbal input	<input type="checkbox"/>	<input type="checkbox"/>	Supportive/rescuing	<input type="checkbox"/>	<input type="checkbox"/>	Team skills	<input type="checkbox"/>	<input type="checkbox"/>
		Active assistance/facilitating	<input type="checkbox"/>	<input type="checkbox"/>	Controlling verbal input (stop)	<input type="checkbox"/>	<input type="checkbox"/>	Emotionally intelligent	<input type="checkbox"/>	<input type="checkbox"/>	Patient-focused	<input type="checkbox"/>	<input type="checkbox"/>
CLOSE		Ask trainee's opinion	<input type="checkbox"/>	<input type="checkbox"/>	Encourage self-reflection	<input type="checkbox"/>	<input type="checkbox"/>	Honest	<input type="checkbox"/>	<input type="checkbox"/>	Excellent teacher	<input type="checkbox"/>	<input type="checkbox"/>
		Appropriate use of materials	<input type="checkbox"/>	<input type="checkbox"/>	Positive and negative reinforcement	<input type="checkbox"/>	<input type="checkbox"/>	Non-threatening	<input type="checkbox"/>	<input type="checkbox"/>	Professionalism	<input type="checkbox"/>	<input type="checkbox"/>
		Performance critique	<input type="checkbox"/>	<input type="checkbox"/>	Analytical	<input type="checkbox"/>	<input type="checkbox"/>	Self-reflects	<input type="checkbox"/>	<input type="checkbox"/>	Excellent communicator	<input type="checkbox"/>	<input type="checkbox"/>
		Learning point agreement	<input type="checkbox"/>	<input type="checkbox"/>	Approachable (allows discussion)	<input type="checkbox"/>	<input type="checkbox"/>	Inspirational	<input type="checkbox"/>	<input type="checkbox"/>	Seeks feedback	<input type="checkbox"/>	<input type="checkbox"/>
Sub-total scores:												TOTAL	
Comments:													

Fig. 1 Structured Training Trainer Assessment Report (STTAR)

Table 1 Overall and breakdown of trainers STTAR scores

Trainer	STTAR	Set	Dialogue	Closure	S	T	A	R	(Overall)
F	19.62	4.82	5.00	4.83	4.60	4.72	5.80	4.96	4.67
B	20.46	4.94	5.35	5.14	4.93	5.16	5.91	4.90	5.3
A	20.80	5.31	5.14	5.38	5.18	5.13	5.65	4.84	5.25
D	21.07	4.91	5.56	5.60	5.31	5.42	5.76	5.18	5.6
C	22.21	5.74	5.53	5.98	5.50	5.67	6.04	5.35	5.7
E	22.41	6.04	5.66	5.76	5.88	5.81	5.58	5.39	5.5

S structure, T training behaviour, A attributes, R role modelling

STTAR forms were completed during the training episode without problem, and there was a good inter-rater reliability between the tally scores marked down by each observer during the case for each different item within STTAR (Cronbach's  $\alpha = 0.84$ , ICC 0.72). The form was deemed feasible as the majority of it was filled during the training episode, with an additional average of 5 min for completion of scores [mean 5 (range 3–6)]. All forms were completed before the start of the next case. They were also thought to be acceptable with an average Likert scale of four (range 3–5).

**Development of the web-based evaluation, the mini-STTAR**

Fifty items scored more than six on the Likert scale (moderately important), in the Delphi process. These were used to develop the mini-STTAR (Fig. 2). Two authors created the assessment form using the different items, and the mini-STTAR was adjusted and improved through review from both the focus group and the educational committee of the NTP. The form was also piloted both in theatre and during a two-day cadaveric course. Different

### Mini-STTAR: Trainee evaluation of trainer

Trainer:	Trainee:	Level:
Procedure:	Previous number of specific procedure:	
Total number of cases with this trainer:		Hospital:

This trainer:

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	N/A
Had a structured approach to the training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Agreed clear aims for this training episode	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Adjusted training appropriately to level of trainee	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Was encouraging	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Was non-threatening	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Was patient	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Provided opportunities to ask questions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Communicated well	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Took over procedure when appropriate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Provided too much verbal input (e.g. difficult to concentrate on procedure)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Provided too little verbal input (e.g. didn't always give guidance when required)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Provided too much physical input (e.g. didn't stretch trainee's abilities)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Provided too little physical input (e.g. trainee's abilities over-stretched)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Provided corrective critique during procedure (e.g. criticised but with explanation)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Provided positive critique during procedure (e.g. praised but with explanation)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Encouraged team awareness	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Was patient-focused	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Encouraged self-reflection on performance	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Derived and agreed learning points from the case	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Is a good role model with respect to their attitude and behaviour (for trainees in general)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Overall is an excellent teacher	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Overall, please indicate the extent to which the training met your expectations:	Below <input type="checkbox"/>		Met <input type="checkbox"/>		Exceeded <input type="checkbox"/>	

Further comments about trainer and/or specific details about case:

	Extremely relevant	Relevant	Neutral	Irrelevant	Extremely irrelevant
Overall, how relevant did you find this form?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
How long did it take you to complete it?	<input type="checkbox"/> <input type="checkbox"/> minutes				

**Fig. 2** Mini-Structured Training Trainer Assessment Report (mini-STTAR). Items can be split into different groups “structure” (1–3, 19), “training behaviour” (9–15, 18), “attributes” (4–8), “role model” (16, 17, 20)

iterations of the mini-STTAR were trailed, and the fifth draft had twenty-one statements for the trainee to rate the degree to which they thought the item occurred during the

teaching session, using a five-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = undecided, 4 = agree and 5 = strongly agree). These items could be

grouped into “structure” (items 1–3, 19), “training behaviour” (items 9–15, 18), “attributes” (items 4–8) and “role model” (items 16, 17 and, 20) (hence mirroring the structure of the detailed STTAR—the observer form) (Fig. 2). There was also a free text box for any additional comments, plus an opportunity for the trainee to indicate the degree to which the training episode met their expectations. Some of the questions were “negative” to minimise responding bias, but data were adjusted in the analysis process to ensure the scores were unidirectional.

**Reliability, feasibility and acceptability of mini-STTAR in the NTP**

Over the 12-month period, 459 training episodes delivered by 44 NTP trainers (42 male and 2 female), in 10 NTP training centres, were rated by 74 trainees (64 male and 10 female). The mean score for structure, training behaviour, attributes and role model was 4.37, 4.31, 4.55 and 4.41, respectively (Fig. 3). Questions 10–13 scores were reversed to ensure unidirectional questions and scores. Inter-item reliability between the 21 different questions was good (Cronbach’s  $\alpha$  0.79), with no significant improvement with the removal of any of the questions (highest achievable 0.81). There were 85 different trainer/trainee combinations where at least two forms were registered, and the intra-observer reliability using Pearson correlation was significant ( $r = 0.701, p = <0.001$ ). For 56 different trainer/trainee combinations where at least

three forms were completed, Pearson correlation remained significant ( $p = <0.001$ ).

The mean and median time for completion was 5.7 and 5.0 (interquartile range 2.0–7.0) minutes, respectively. The median score for the usefulness of the form was 4.0 (IQR 3–4: 1 = extremely irrelevant, 5 = extremely relevant). In terms of trainees’ perception of training, only five out of 459 training episodes failed to meet trainees’ expectations, whilst 282 met and 152 exceeded expectations.

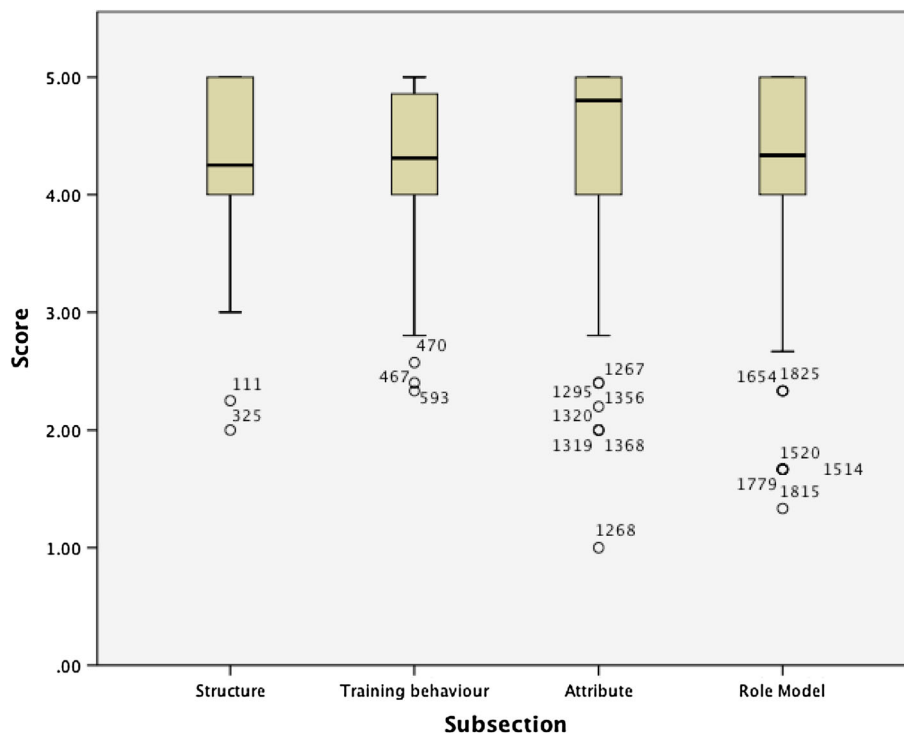
**Educational impact of mini-STTAR**

Thirty-eight trainers who had taught at least four cases within the NTP were provided with an educational report with an average of 17 cases per trainer (IQR 21.3). The average trainers score for the usefulness of being provided with a training report was 4.4 (max 5, range 2–5), with 16 out of the 17 who responded to the questionnaire stating that it would help them reflect and impact their training behaviour for further training episodes.

**Discussion**

In this study, we developed and established the reliability of STTAR as a tool for the evaluation of training quality. The STTAR and its shortened electronic format the mini-STTAR were implemented in the NTP in England and were found to be feasible and acceptable by trainers. The

**Fig. 3** Comparison of mean scores for different subsections (structure, training behaviour, attributes and role model) within mini-STTAR—no significant difference found ( $p = 0.07$ )



tool was used to compose an educational report for each trainer in order to provide a framework for reflection on training characteristics. The STTAR and mini-STTAR allow detailed analysis of training characteristics and non-technical skills relevant to training from both an observer and a trainee's viewpoint.

Previously described characteristics of a medical teacher include enthusiasm, clear, well-organised presentation of instructional material, skill in interaction with students/residents and group settings, involvement of the learner in the teaching process, a humanistic orientation, content knowledge of the subject and use of case-based teaching scripts [25]. These factors are not that dissimilar to those found in this research. A more recent study described the necessary characteristics of a good surgical trainer, and there is also certainly some overlap with these items [26]. The most important reflection factor was thought to be to "have a feedback session". Given that feedback has been demonstrated to improve performance, this should be a fundamental part of training advanced laparoscopic surgery, despite it being a shift from the more traditional method of surgical teaching [27].

For the structure of training episodes, there was a difference in opinion between trainers and trainees. This was not much greater than one point, but often the trainees felt more strongly than the trainers. Interestingly, trainees did not want to be taken out of their comfort zone, but they did want to be allowed to struggle, whereas trainers wanted the contrary. This disparity is likely to demonstrate a difficulty with the use of language within the questionnaire for these particular items. For the trainer, some of the free comments that came back from round one were that they did not like the idea that the trainee could potentially be putting the patient at risk due to "struggle", whereas "to be taken out of the comfort zone" simply suggests that the trainee is being pushed, which is an important part of "adult" learning and is in concordance with Vygotsky's theory of ZPD [28]. The other differences occurred in case selection where trainees give the opinion that they want to operate on a full case mix from the start. This is likely to reflect the enthusiasm of most trainees to gain operative experience in a supervised environment. The Lapco risk prediction score supports case selection during the learning curve and guides the suitability of cases for training [29, 30].

On using the observer version of STTAR, although trainers generally scored highly, there were lower scores within training behaviour. When looking at individual scores, this can be explained partly by many trainers failing to "set ground rules". Since in the Lapco TT setting, trainer pairs take it in turns to teach the same trainee, the re-setting of ground rules may have failed to take place for fear of repetition [21]. In any case, from the interviews it is described as being the key process to being able to control

a training relationship—especially when the power gradient is reversed (i.e. more junior trainer with senior trainee). The trainers were also poor at seeking feedback from the trainees, at self-reflecting and at encouraging self-reflection in the trainees. This was again demonstrated in the low scores for closure. These factors are the most unnatural to the usual "master-apprentice" model of teaching and suggest that the trainer has to now make what used to be implicit, explicit, in order to maximise the limited training time. Using the STTAR as a feedback tool will help the trainers improve their feedback technique, and once they are in the routine of performing feedback, this may aid their reflection abilities, as feedback given constructively has been shown to improve learning [27]. Reflecting, or focusing on a key point from the training episode, develops "mental practice" which has been shown to improve operative performance and may well affect training ability in the same way [31, 32].

The mini-STTAR mean score was 4.6 of a possible 5. This is surprisingly high as the trainees within the NTP were potentially the harshest critics as they themselves were trainers who were challenging to teach for several reasons—age difference, expectations, limited time, consultant surgeons and the usual hierarchy were not in place, which ordinarily might hinder response candour if trainees feared of a lack of anonymity, or alter the response due to a predetermined conception of the trainer (Halo effect) [22]. The alteration in hierarchy has in fact been shown not to be a significant factor in the appreciation of the quality of training [15]. It is reassuring for the NTP that the trainers were perceived to be fulfilling their role.

Several biases in responding to questionnaires have been described. The current study addressed "satisficing bias", where responders give a response just for the sake of giving a response by occasionally switching the direction of the scale. This also helped to reduce "acquiescence bias", where the responder simply ticks the positive response [22]. For the STTAR, the Likert scale was a bipolar odd-numbered seven-item labelled scale with a central point of neutrality in order to reduce "central-tendency" bias, where the responders tend to avoid the end of the scales [22]. For the mini-STTAR, a five-point scale was used as the seven-point scale was found to be too cumbersome for the trainees to use which may have caused some bias. Although the scales were Likert, it was presumed that the gaps between each value were equal. This may have allowed for some inaccuracies to occur. One of the benefits of interviewing either face to face or over the telephone was that it was possible for the interviewer to rephrase the questions should it become clear that the interviewee did not understand. There is a possible risk of introducing bias at this stage given that the interviewer may inadvertently influence the answers, and furthermore, despite these



measures, the interviewee still may not completely understand [22]. All of the participants in this study were fluent in English, however, and the interviewers were trained in interviewing techniques, which should have minimised this bias. The STTAR Likert scale that measures the extent that something happened has some limitations as a high score for some items can be good, or bad, depending on the context. This is particularly difficult for the “attributes”. Although the scale works and gives the observer a relatively easy way of assessing, it could in fact be more analytical.

Limitations of the study design include the fact that telephone interviews can limit the quality of data collected through the brevity of responses and the lack of control over participant focus. A semi-structured approach was adopted given that so little was known about the subject, and the questions were formulated from the aims of what was needed to be determined for the NTP. Furthermore, no anaesthetists, theatre nurses, ward nurses or patients were invited to participate in the study. Although each of these groups is involved during the training episode, it is an indirect interaction and after careful consideration, they were excluded. This presumption may have been incorrect. Other biases within the study methodology have already been addressed. For the observer version of STTAR, the participants within the pilot study were a very specific group of people fundamentally motivated to teach and learn evidenced by the fact that they had enrolled in a non-compulsory course designed to focus on training. Likewise, for the web-based mini-STTAR, used after actual NTP training sessions, both trainers and trainees have voluntarily invested time in training episodes, and thus, it is expected that they would be very motivated. It should be acknowledged that there is likely to be a slightly positive skew to these results. The training environment within the Lapco TT course and the NTP itself may not be truly generalisable to the “normal” training set up. In addition, it was not possible to determine criterion validity, as there is no current gold standard to compare with.

In conclusion, STTAR has been developed and shown to be a reliable, feasible and acceptable evaluation tool to both trainers and trainees in a national training programme. The context of using STTAR is of particular importance as trainees were specialists undergoing training in advanced laparoscopic skills and thus representing a challenging training environment. The mini-STTAR enables trainees to construct a feedback for trainers after every training episode. This will then provide trainers with an independent detailed analysis of their teaching skills and highlight those areas of excellence and those in need of development. The tools have been implemented throughout the national training programme and the Lapco “Training the Trainer” course. The STTAR is a powerful evaluation tool that can be used in further training programmes to ensure the quality of training in any practical skill.

**Acknowledgments** This work was funded by the National Cancer Action Team, Department of Health UK. Thanks go to Mrs. Laura Langsford for enabling the form to be piloted within the Lapco TT course and to Mr. Andy McMeeking at the National Cancer Action Team for supporting some of the funding. Also to Dr. John Anderson and Dr. Chris Wells for their invaluable advice and input to both the methodology and the form structure. To the National Training Programme and all trainers and trainees involved in this. Also specifically to the following surgeons and their trainees and theatre staff: Mr. Austin Acheson, Mr. Chris Cunningham, Mr. John Griffiths, Mr. Mark Gudgeon, Mr. Alan Horgan, Mr. Roel Hompes, Mr. Ian Jenkins, Mr. Mark Katory, Mr. Ian Lindsey, Mr. Charles Maxwell-Armstrong, Professor Robin Kennedy, Professor Roger Motson and Professor Timothy Rockall, who all readily welcomed the researchers into their operating theatres in the early piloting phases of the STTAR and mini-STTAR.

**Disclosures** Dr. Valori is the director of a consultancy called “Quality Solutions for Healthcare LLP” which specialises in quality improvement and training. Drs. Wyles, Miskovic, Ni, Coleman and Professors Darzi and Hanna have no conflicts of interest or financial ties to disclose.

## Appendix

See Table 2.

**Table 2** Training technique items with >1 difference in consensus, with all becoming more extreme with a shift to the right of the seven-point scale

Training item	Round 1	S.D. 1	Round 2	S.D. 2	S.D. 1–S.D. 2
Discuss with trainee what they want to achieve	5.61	2.2	6.5	0.7	1.5
Highlight specific potential problems of case	5.56	2.2	6.17	0.7	1.5
Alter the level of supervision according to trainee’s level	5.28	2.27	6.28	0.82	1.45
Ensure patient safety is a priority	6.61	1.65	6.94	0.23	1.42
Get trainee to do one thing at a time	5.06	2.53	5.56	1.14	1.39
Identify trainee’s concerns	5.39	2.11	6.22	0.73	1.38
Demonstrate the task in an accessible way for the trainee	5.72	2.21	6.11	0.83	1.38

**Table 2** continued

Training item	Round 1	S.D. 1	Round 2	S.D. 2	S.D. 1–S.D. 2
Identify trainee's anxieties	5.33	2.14	6.17	0.78	1.36
Explain why you are doing what you are doing when you take over	5.67	2.22	6.33	0.9	1.32
Establish a training environment	4.89	2.29	5.39	0.97	1.32
Ensure assistant is senior	4.56	2.33	4.28	1.01	1.32
Establish roles	4.94	2.5	5.61	1.19	1.31
Informal discussion with trainee in theatre	5.33	2.27	5.67	1.02	1.25
Identify trainee's level of confidence	5.33	2.14	6.11	0.9	1.24
Encourage trainee throughout the procedure	5.17	2.22	5.61	1.03	1.19
Reduce the amount of verbal input as trainee progresses	5.39	2.33	5.83	1.2	1.13
Reduce the amount of physical input as trainee progresses	5.83	1.79	6.28	0.66	1.13
Minimise interruptions in theatre	4.83	1.97	5.39	0.85	1.12
Establish ground rules	5.39	2.11	5.94	0.99	1.12
With the trainee discuss and agree on potential solutions for challenges within the case	5.89	1.71	6.39	0.6	1.11
Reduce number of operations on the list	5.11	2.19	5.44	1.09	1.1
Provide specific top tips for case	5.89	1.67	6.39	0.6	1.07
Have specific designated training list	4.39	2.2	5.17	1.15	1.05
Establish realistic expectations	6.22	1.66	6.56	0.61	1.05
Encourage mental practice of procedure, i.e. trainee takes time to think through case	5.11	2.05	5.67	1	1.05
Align agendas, i.e. what you expect/what they expect from the day	5.28	1.96	5.61	0.91	1.05
Identify trainee's needs	6	1.64	6.5	0.61	1.03
Check understanding of procedure	5.78	1.7	6	0.68	1.02
Accentuate the important steps	5.94	1.62	6.17	0.61	1.01
Stop the trainee and discuss the options of how to proceed	5.56	2.06	6.06	1.05	1.01
Encourage trainee's awareness of the team	5.44	1.85	5.83	0.85	1
Discuss with trainee what wants to achieve	5.17	1.94	5.78	0.94	1
Allow enough time for training	6.06	1.66	6.28	0.66	1

## References

- Birch DW, Bonjer HJ, Crossley C et al (2009) Canadian consensus conference on the development of training and practice standards in advanced minimally invasive surgery: Edmonton, Alta., Jun. 1, 2007. *Can J Surg* 52(4):321–327
- Blue AV, Griffith CH 3rd, Wilson J et al (1999) Surgical teaching quality makes a difference. *Am J Surg* 177(1):86–89
- Hull L, Arora S, Aggarwal R et al (2012) The impact of non-technical skills on technical performance in surgery: a systematic review. *J Am Coll Surg* 214(2):214–230
- Sevdalis N, Davis R, Koutantji M et al (2008) Reliability of a revised NOTECHS scale for use in surgical teams. *Am J Surg* 196(2):184–190
- Haynes AB, Weiser TG, Berry WR et al (2009) A surgical safety checklist to reduce morbidity and mortality in a global population. *New Engl J Med* 360(5):491–499
- Peyton J (1998) Teaching and learning in medical practice. Manticore Europe Limited, Rickmansworth
- Skeff KM, Stratos GA, Berman J et al (1992) Improving clinical teaching. Evaluation of a national dissemination program. *Arch Intern Med* 152(6):1156–1161
- Copertino N, Blackham R, Hamdorf JM (2010) A short course for surgical supervisors and trainers: effecting behavioural change. *ANZ J Surg* 80(12):896–901
- London Deanery ARCP <http://www.londondeanery.ac.uk/specialty-schools/surgery>. (Accessed 22 Sept 2014)
- Copeland HL, Hewson MG (2000) Developing and testing an instrument to measure the effectiveness of clinical teaching in an academic medical center. *Acad Med* 75(2):161–166
- Cohen R, MacRae H, Jamieson C (1996) Teaching effectiveness of surgeons. *Am J Surg* 171(6):612–614
- Beckman TJ, Ghosh AK, Cook DA et al (2004) How reliable are assessments of clinical teaching? A review of the published instruments. *J Gen Intern Med* 19(9):971–977
- Miskovic D, Wyles SM, Ni M et al (2010) Systematic review on mentoring and simulation in laparoscopic colorectal surgery. *Ann Surg* 252(2):943–951
- Coleman MG, Hanna GB, Kennedy R (2011) The National Training Programme for laparoscopic colorectal surgery in England: a new training paradigm. *Colorectal Dis* 13(6):614–616
- Wyles SM, Miskovic D, Ni M et al (2012) "Trainee" evaluation of the English National Training Programme for laparoscopic colorectal surgery. *Colorectal Dis* 14(6):352–357
- Lapco. National Training Programme in laparoscopic colorectal surgery. <http://www.lapco.nhs.uk>. (Accessed 22 Sept 2014)
- Kennedy TJ, Lingard LA (2006) Making sense of grounded theory in medical education. *Med Educ* 40(2):101–108
- NVIVO [http://www.qsrinternational.com/FileResourceHandler.ashx/RelatedDocuments/DocumentFile/456/NVivo\\_8\\_brochure.pdf](http://www.qsrinternational.com/FileResourceHandler.ashx/RelatedDocuments/DocumentFile/456/NVivo_8_brochure.pdf). (Accessed 22 Sept 2014)
- Britten N (1995) Qualitative interviews in medical research. *BMJ* 311(6999):251–253

20. Brownpaper-technique [http://www.fpxhn.net/ttsubpages/1week\\_comtem/process-mapping.pps](http://www.fpxhn.net/ttsubpages/1week_comtem/process-mapping.pps). (Accessed 17th Mar 2013)
21. Mackenzie H, Cuming T, Miskovic D et al (2013) Design, delivery, and validation of a trainer curriculum for the national laparoscopic colorectal training program in England. *Ann Surg* 261(1):149–156
22. Streiner N, Norman G (2008) *Health measurement scales: a practical guide to their development and use*, 4th edn. Oxford University Press, Oxford
23. Whyte WF (1982) *Interviewing in field research*. George Allen and Unwin, London
24. Lingard L, Regehr G, Espin S et al (2006) A theory-based instrument to evaluate team communication in the operating room: balancing measurement authenticity and reliability. *Qual Saf Health Care* 15(6):422–426
25. Irby DM (1978) Clinical teacher effectiveness in medicine. *J Med Educ* 53(10):808–815
26. Nisar PJ, Scott HJ (2011) Key attributes of a modern surgical trainer: perspectives from consultants and trainees in the United Kingdom. *J Surg Educ* 68(3):202–208
27. Rolfe I, McPherson J (1995) Formative assessment: how am I doing? *Lancet* 345(8953):837–839
28. Vygotsky L (1978) *Mind in society*. Harvard University Press, Cambridge
29. Miskovic D, Ni M, Wyles SM et al (2012) Learning curve and case selection in laparoscopic colorectal surgery: systematic review and international multicenter analysis of 4852 cases. *Dis Colon Rectum* 55(12):1300–1310
30. Mackenzie H, Miskovic D, Ni M et al (2014) Risk prediction score in laparoscopic colorectal surgery training: experience from the English National Training Program. *Ann Surg* 261(2):338–344
31. Arora S, Aggarwal R, Sevdalis N et al (2010) Development and validation of mental practice as a training strategy for laparoscopic surgery. *Surg Endosc* 24(1):179–187
32. Arora S, Aggarwal R, Sirimanna P et al (2011) Mental practice enhances surgical technical skills: a randomized controlled study. *Ann Surg* 253(2):265–270