Taylor & Francis
Taylor & Francis Group

Minimally Invasive Therapy & Allied Technologies

ISSN: 1364-5706 (Print) 1365-2931 (Online) Journal homepage: https://www.tandfonline.com/loi/imit20

Measuring fatigue and stress in laparoscopic surgery: validity and reliability of the star-track test

Kim Platte, Chantal C.J. Alleblas, Joanna Inthout & Theodoor E. Nieboer

To cite this article: Kim Platte, Chantal C.J. Alleblas, Joanna Inthout & Theodoor E. Nieboer (2019) Measuring fatigue and stress in laparoscopic surgery: validity and reliability of the star-track test, Minimally Invasive Therapy & Allied Technologies, 28:1, 57-64, DOI: 10.1080/13645706.2018.1470984

To link to this article: https://doi.org/10.1080/13645706.2018.1470984

Published online: 15 May 2018.

MINIMALLY INVASIVE THERAPY & ALLIED TECHNOLOGIES
2019, VOL. 28, NO. 1, 57–64
https://doi.org/10.1080/13645706.2018.1470984

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

OPEN ACCESS Check for updates

Measuring fatigue and stress in laparoscopic surgery: validity and reliability of the star-track test

Kim Platte[a], Chantal C.J. Alleblas[a], Joanna Inthout[b] and Theodoor E. Nieboer[a]

[a]Department of Obstetrics and Gynaecology, Radboud University Medical Center, Nijmegen, The Netherlands; [b]Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands

ABSTRACT

Background: The star-track test has been assessed as valid and reliable to measure manual dexterity in the context of open surgery. We aimed to determine the construct validity and test-retest reliability of the star-track test for manual dexterity in a laparoscopic setting.

Material and methods: The star-track test was performed in a laparoscopic box trainer. To determine construct validity an open-label, randomized four-period crossover trial was conducted. Alongside a baseline (non-interventional) measurement, interventions involved: physical fatigue, mental stress and a combination of these. The test-retest trial involved two separate (non-interventional) measurements. The primary outcome measures were accuracy, speed and manual dexterity (the integrated measure of accuracy and speed).

Results: Participants made significantly more errors when physically fatigued, whereas participants performed the test significantly slower when mentally stressed. Manual dexterity was significantly affected in the case of combined intervention. High test-retest reliability was found for errors (ICC = 0.90) and completion time (ICC = 0.64). Fair test-retest reliability for the integrated measure was found (ICC = 0.37).

Conclusion: The star-track test is a valid and reliable tool to evaluate the effect of physical fatigue and/or mental stress on the characteristics of manual dexterity in a laparoscopic setting.

ARTICLE HISTORY
Received 5 July 2017
Accepted 30 March 2018

KEYWORDS
Laparoscopy; surgical skills; simulation; performance; physical workload; mental stress

Introduction

Performing laparoscopic surgery requires proficiency in complex motor skills, especially manual dexterity. This involves the ability to use the hands to perform an action accurately, quickly and skillfully and concerns e.g., hand-eye-coordination. Laparoscopic surgery demands a great deal of a surgeon's mental and physical capacities [1–4]. This commonly leads to physical fatigue or musculoskeletal disorders [3–5]. It is known that these physical conditions affect accuracy in precision tasks [6–8], whereas preservation of a surgeon's technical skills is highly important for patient safety [9].

Nowadays, advanced tracking systems and software are used to evaluate psychomotor skills in laparoscopy [10]. Montanari et al. [11] stated that in comparison with these box trainers, the use of a low-cost box trainer is equally effective in terms of basic laparoscopic skill acquisition. The implementation of box trainers or tracking systems in surgical simulators for surgical training programs enables objective assessment of those laparoscopic psychomotor skills, by means of force and motion-related metrics [12,13]. To date those systems are mainly used to assess surgical competence level and monitor progression of surgeons' skills during training [14,15]. The Fundamentals of Laparoscopic Surgery (FLS) program is widely used to train and assess basic laparoscopic surgery [13]. However, subjectivity of interpretation, lack of immediate scoring and feedback, and costs have been listed as disadvantages of this program [16]. A certain extent of variability in task performance remains unnoticed while a predefined path allows for exact registration of completion time and error, which is especially relevant in measuring manual dexterity. Also, to add clinical relevance in analyzing laparoscopic skills, task completion time and predefined errors should be assessed in addition to detailed motion metrics [12,15].

The star-track test is a rather simple device, used to objectively assess the impact of stressors on manual

CONTACT Kim Platte kim.platte@radboudumc.nl Radboud University Medical Center, Department of Obstetrics and Gynaecology, PO Box 9101, 6500 HB Nijmegen, The Netherlands

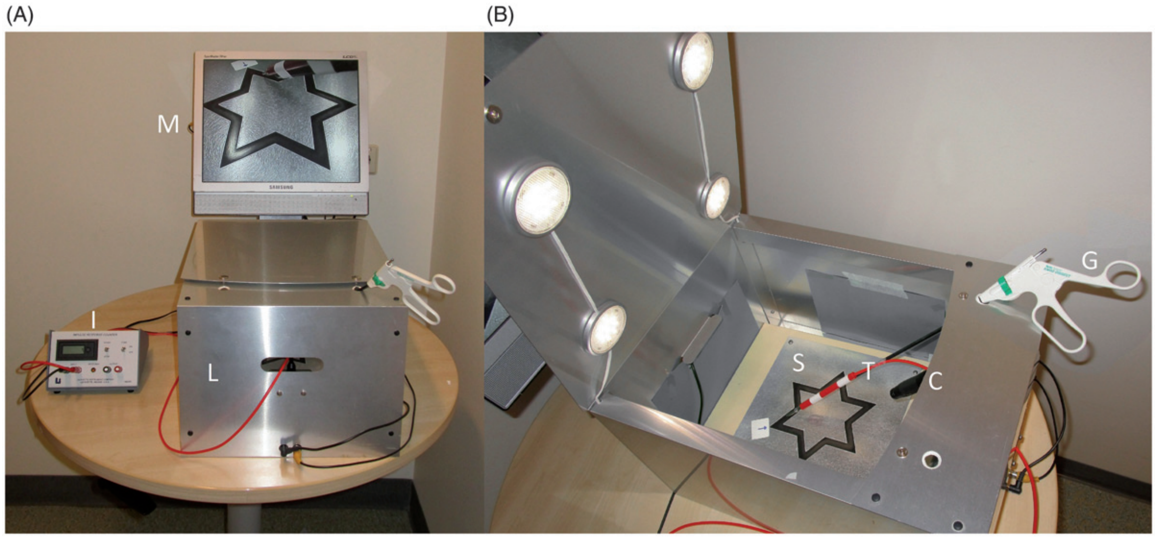

Figure 1. (A) Front view of the experimental setup including the laparoscopic box trainer (L), monitor (M) and impulse counter (I). (B) Inside view of the box including a camera (C), the aluminum plate (22 × 22 cm) with the non-conducting 0.9 cm wide black star pattern (point to point distance: 15.3 cm) (S) anodized into the surface. The metallic tracing stylus (T) was attached to the laparoscopic grasper (G).

dexterity (defined as the integrated measure of number of errors and completion time). Dorion et al. [17] used the star-track test to measure and compare the surgeons' manual dexterity after performing open surgery. The device consisted of a metal plate with a star-shaped track anodized into the surface. The star shape had to be followed with a conducting stylus and when the stylus moved beyond the border of the star shape, an error was counted. They studied the effect of 20 s long intra-operative micro-breaks every 20 min during surgery. Subsequently, Kildebro et al. [18] further investigated the device. It was concluded that the device could be used in conventional open surgery to discriminate between a subject's baseline manual dexterity and a subject's manual dexterity after exposure to physical fatigue or mental stress.

To date, little is known about the potential deterioration of the surgeon's manual dexterity during laparoscopy as a consequence of intra-operative occurrence of mental stress or physical complaints. It is, however, of interest to study the effect of intra-operative induction of mental stress and physical strain on the manual dexterity of laparoscopic surgeons. Based on the results of Kildebro et al. [18], it was hypothesized that the star-track test would also enable distinction between a subject's manual dexterity at baseline and after exposure to physical fatigue or mental stress in laparoscopic surgery. The purpose of this study was to investigate the construct validity of the star-track test and to determine its test-retest reliability in the context of laparoscopic surgery.

Material and methods

Experimental setup and procedure

The star-track test (Figure 1) consists of a tracing task in a laparoscopic box. For this study, the Automatic Mirror Tracer Model 58024A*C (Lafayette Instrument Co. Europe, Loughborough, UK) was used. After the mirror and the shield had been dismounted from the device, the aluminum plate with a non-conducting star-shaped track was placed in a laparoscopic box trainer to create a laparoscopic setup. Furthermore, the device consists of a metal-tipped stylus and an impulse counter that registers every error, defined as any contact between the aluminum plate and the stylus. A laparoscopic grasper was inserted through the upper side of the laparoscopic box (on the right side for right-handed participants and on the left side for left-handed participants) and the metallic-tracing stylus was attached to the end of the grasper's shaft. A camera inside the box reproduced a live video of the star-track on the monitor. Table height was set to an ergonomically sound position for each participant. Visual and auditory distractors were eliminated from the experimental room. Figure 1 shows the complete overview, respectively the front view (Figure 1(A)) and the inside view (Figure 1(B)) of the experimental setup.

Participants were instructed to control the laparoscopic grasper with their dominant hand and to follow the star-track ten times consecutively: five times clockwise followed by five times counterclockwise.

Constant contact between the stylus tip and the star-track surface was required. The goal was to complete this task as fast and accurately as possible. It was emphasized that both the time to complete the task and the amount of errors were equally important parameters for the overall test score. After this instruction and before performing the test, participants gave their written informed consent for study purposes. Subsequently, participants were instructed to try one round clockwise and one round counter-clockwise. This enabled each participant to become familiarized with the task and to customize the experimental set-up (table height and monitor position) to make sure that poor ergonomics would not hamper task performance [19]. While the participant performed the test, the examiner scored all errors per round by registering all buzzers produced by the impulse counter. The time to complete each round and to complete the whole test was measured by using a stopwatch.

Construct validity

To establish the construct validity of the star-track test with regard to its ability to distinguish between a subject's normal manual dexterity and his or her dexterity while being physically fatigued and/or mentally stressed, a randomized four-period crossover trial with a control measurement and three active interventions was conducted. Therefore we aimed to include 30 subjects. Subjects were medical interns of the Radboud University Medical Center from the Department of Obstetrics and Gynecology and medical students from the Radboud University, Nijmegen, The Netherlands.

Participants were instructed to perform the star-track test four times, each time separated by two days [18,20]. A baseline measurement and three interventions were randomized in order for each participant. The first intervention involved physical fatigue. Since mainly a surgeon's shoulders are susceptible to fatigue during laparoscopy, shoulder fatigue was induced by an isometric exercise lifting dumbbells towards a 90° shoulder abduction and maintain this position for as long as possible [18,21,22]. Prior to lifting, participants could choose between dumbbells of 1.5, 2 and 2.5 kg. The weight was chosen based on the participant's physical capacities in consultation with the experimenter. Participants verbally indicated their intensity of shoulder fatigue and when participants indicated that they had almost reached maximum fatigue, the experimenter encouraged to continue the exercise as long as possible [23]. Subjects had to start performing the star-track test within ten seconds after the point of maximal shoulder fatigue was reached [18]. The second intervention was the addition of mental stress. The participants were asked to solve mental calculation, while performing the test [24]. The third intervention involved the combination of intervention one and two.

Test-retest reliability

To determine the reliability of the star-track test for manual dexterity in laparoscopic surgery, its consistency over time was assessed in a separate trial. We aimed to include 12 participants: six gynecologists who often perform laparoscopic surgery and six gynecology residents with moderate experience in laparoscopy. Participants were enrolled by a request for participation through email. All were to perform the star-track test two times, separated by two days without any interventions.

Statistical analysis

The sample size calculation was based on the results for the integrated measure as found by Kildebro et al. [18] and additional information provided by the author. For the construct validity trial, 30 participants were needed to find with 80% power a significant difference of 34 in the integrated measure, assuming a standard deviation of 55, a within-subject correlation of 0.5, and a two-sided Bonferroni-adjusted significance level of 0.0167 (i.e., 0.05 divided by 3). A total of 30 participants were also sufficient to detect a difference of 2.8 errors, assuming an SD of 4.5 errors, or a difference of ten seconds in completion time, assuming an SD of 16.4 s, with 80% power, a within-subject correlation of 0.5, and a two-sided significance level of 0.0167. For the reliability trial, based on a sample size calculation providing 80% power, we needed 12 participants to detect an intra-class correlation coefficient (ICC) > 0.6 assuming an actual ICC of 0.9, if using a one sided 95% confidence interval and two measurements.

We evaluated manual dexterity as the combined outcome of errors and time. Since these parameters have different units of measurement, we used an integrated measure for the number of errors and completion time [25]. To calculate this integrated measure, first the results on each outcome (the number of errors or the completion time) were converted into ranks. Then the difference between the individual ranks and the median rank were calculated per outcome and divided by the median rank, resulting in fractional ranks on a scale from −100 to 100. A negative score indicates a better performance, because, when compared to the median rank (with score 0 on the scale), fewer errors were made or less time to complete the task was needed. The fractional ranks of errors and time were then added on a per-subject basis to form the integrated measure.

To determine whether manual dexterity was affected by the interventions, a linear mixed model analysis, taking into account the crossover design, was performed for each outcome. The model included as fixed factors the intervention and the period, and as random factor the participant, with a compound symmetry structure reflecting the correlation between the repeated measurements of a participant. A Bonferroni correction was applied to correct for the three pairwise comparisons per outcome. Two-sided p values $\leq .05$ were regarded as statistically significant.

The single measure ICC was used for test-retest reliability analysis [26,27]. The (single measure) ICC for absolute agreement was estimated by means of a two-way mixed model, with the timing as fixed factor and participant as random factor. The Pearson correlation coefficient (r) was used to evaluate the correlation between the time needed to complete a round and the number of errors made each round. We interpreted the correlation outcomes as follows: ICC or $r < 0$ reflects 'poor', 0 to 0.20 'slight', 0.21 to 0.4 'fair', 0.41 to 0.60 'moderate', 0.61 to 0.8 'substantial', and above 0.81 'almost perfect' agreement [26]. The statistical analysis was performed using IBM SPSS Statistics version 22.0 (SPSS, Chicago, IL, USA).

Results

Construct validity

Thirty participants, i.e., 12 medical interns and 18 medical students, were enrolled in this study, involving 21 females and nine males with a median age of 22 years (range 20–27). Two participants were left-handed. All participants gave their written informed consent prior to participation. Summary data of the three outcome measures after baseline and the three interventions are presented in Table 1. The mean scores for each intervention on day 1, 4, 7 and 10 were calculated and presented in Figure 2, respectively mean errors (Figure 2(A)), task completion time (Figure 2(B)) and integrated measures (Figure 2(C)) per period.

Table 2 presents the estimated differences between the baseline scores and the scores after the interventions. Compared to baseline, a significantly increased number of errors was found when participants were exposed to the fatigue intervention ($p = .014$), whereas no notable difference in errors was found after the stress intervention ($p = 1.0$) or combined intervention ($p = .774$). Task completion time was not affected by the fatigue intervention ($p = 1.0$) while both stress ($p < .001$) and the combination of fatigue and stress ($p < .001$) resulted in a significantly increased time in completing the task. Eventually, only the combined intervention resulted in a statistically significant inferior score on the integrated measure ($p = .014$).

Table 1. Summary data of the test scores after baseline and intervention measurements*.

	Number of errors	Task completion time (seconds)	Integrated measure
Baseline	12.5 (9.3; 15.7)	177 (154; 200)	−24.0 (−44.3; −3.6)
Fatigue	17.0 (13.8; 20.1)	176 (153; 199)	2.9 (−17.4; 23.3)
Stress	12.1 (8.9; 15.3)	222 (199; 245)	4.6 (−15.8; 25.0)
Combined	14.2 (11.1; 17.4)	218 (196; 241)	16.3 (−4.1; 36.7)

*Data presented as estimated mean (95% Confidence Interval). A higher score indicates inferior performance.

It was observed that participants improved their speed whilst performing the test. The more rounds participants performed, the less time they needed to complete a round. This was observed for both clockwise and counterclockwise, while no correlation between direction of movement and speed was found ($r = -0.05$). A negative correlation ($r = -0.22$; $p < .01$) was found between time needed to complete a round and errors per round, i.e., significantly more errors are made when a participant moves faster. Table 3 presents the Pearson Correlations of the time and errors per round for each intervention.

Test-retest reliability

For the test-retest reliability trial, 16 subjects were included: four gynecologists, one general surgeon, six residents and five medical interns. Eleven participants were female. Two participants were left-handed. The mean scores for each test day are presented in Table 4. There was an almost perfect correlation between the number of errors of the two testing days (ICC = 0.90 (95% CI 0.74 to 0.96)). Substantial reliability was found for task completion time (0.64 (95% CI 0.06 to 0.88)). Fair reliability was found when the integrated measure was used to evaluate the results (0.37 (95% CI −0.08 to 0.71)).

Discussion

The purpose of this study was to examine the construct validity and test-retest reliability of the

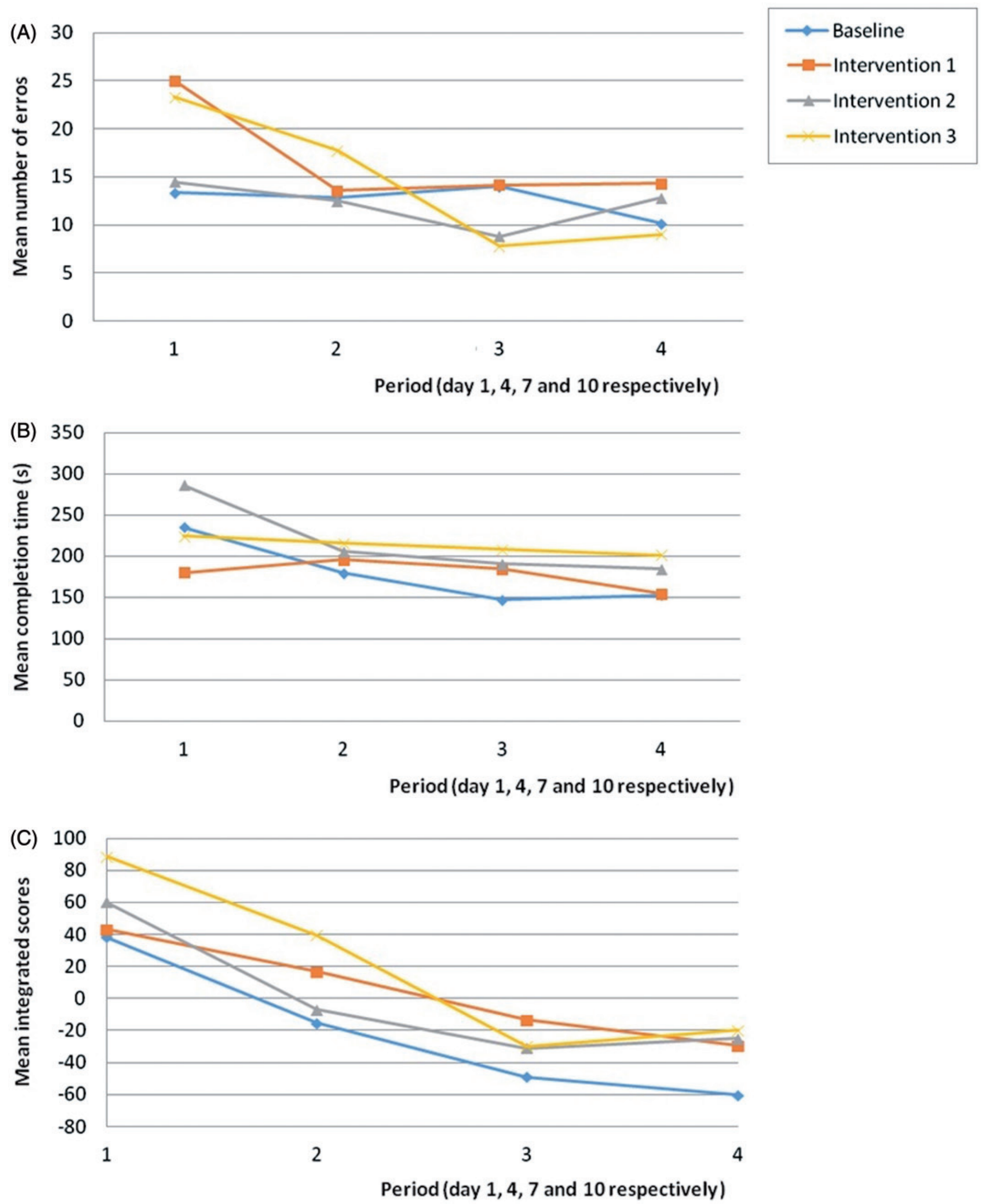

Figure 2. Mean scores for each intervention on day 1, 4, 7 and 10 respectively. (A) Mean errors per period. (B) Mean task completion time per period. C: Mean integrated measures per period.

Table 2. Pairwise comparisons; interventions versus baseline measurement*.

	Number of errors	Task completion time	Integrated measure
Fatigue vs. Baseline	4.5 (0.7; 8.2); $p = .014$	−1.2 (−24.1; 21.7); $p = 1.0$	26.9 (−6.9; 60.7); $p = .166$
Stress vs. Baseline	−0.4 (−4.1; 3.3); $p = 1.0$	45.2 (22.4; 68.0); $p < .001$	28.6 (−5.1; 62.2); $p = .124$
Combined vs. Baseline	1.7 (−2.0; 5.5); $p = .774$	41.1 (18.3; 64.0); $p = .000$	40.2 (6.5; 74.0); $p = .014$

*Data presented as mean difference (95% Confidence Interval for difference). A positive mean difference indicates inferior performance during the respective intervention compared to baseline performance and vice versa. *p*: Bonferroni corrected *p* values.

star-track test for manual dexterity in the context of laparoscopic surgery. The results showed that the star-track test enables distinction between a subject's baseline manual dexterity and a subject's manual dexterity after exposure to fatigue or mental stress, by significantly affecting accuracy and speed respectively. Almost perfect test-retest reliability for number of errors and substantial test-retest reliability for task completion time was found. The integrated measure of accuracy and speed, for which fair test-retest

Table 3. Pearson Correlation between time needed to complete a round and errors in each round.

	r	Sig.*
Baseline	−0.19	<0.01
Fatigue	−0.28	<0.01
Stress	−0.20	<0.01
Combined	−0.23	<0.01

*Correlation is significant at the 0.01 level.

Table 4. Mean outcomes of day 1 and 2 of the test-retest trial*.

	Number of errors	Task completion time (seconds)	Integrated measure
Test day 1	6.7 (4.1; 9.3)	177 (160; 194)	16.9 (−17.4; 51.1)
Test day 2	7.4 (3.9; 10.9)	158 (142; 173)	−16.9 (−47.9; 14.2)

*Data presented as mean (95% Confidence Interval).

reliability was found, was significantly affected by the combined intervention involving both physical fatigue and mental stress.

Laparoscopic surgeons commonly suffer from physical strain and mental stress due to highly demanding laparoscopic procedures [28]. It is assumed that these demanding conditions affect their manual dexterity and may deteriorate surgical performance [29]. The results of the construct validity trial indicate that physical fatigue results in an increased level of errors, whereas mental stress affects the time required to complete the task. However, both interventions did not affect the integrated measure of errors and time. The combination of exposure to both mental stress and physical fatigue did affect the integrated measure of errors and completion time. From these results, it can be concluded that the underlying characteristics of manual dexterity are affected by different types of load (mental or physical). Therefore, both error and time are equally important as the integrated measure for performance analysis.

Box trainers and surgical simulators are used in surgical training programs to assess psychomotor laparoscopic skills. These devices are mainly used to assess surgical competence level by motion-related metrics [10,12,14,15]. However, for clinical implementation, quality metrics, such as errors, should be assessed additionally to analyze laparoscopic performance [12,15]. Whereas devices such as the ICSAD (Imperial College Surgical Assessment Device), the ADEPT (Advanced Dundee Endoscopic Psychomotor Trainer), the ProMIS^{TM} (Haptica, Dublin, Ireland), the HUESAD (Hiroshima University Endoscopic Surgical Assessment Device) and the TrEndo Tracking System analyze technical performance by motion related metrics such as time, number of movements and path length, the star-track test is particularly capable of measuring errors in addition to completion time in a relatively low-cost and simple setting with realistic laparoscopic features including graspers [30,31].

The reliability of the star-track test in a non-laparoscopic setting was reported strong, after investigation in previous studies [17,18]. Respectively Pearson's correlations of $r = 0.955$ (17) and $r = 0.90$ (18) were found. We used the ICC to evaluate reliability, because unlike Pearson's *r*, the ICC accounts for both consistency of performances from test to retest (within-subject range) as well as change in average performance of participants as a group over time. The test showed high test-retest reliability for errors (ICC = 0.897) and completion time (ICC = 0.64). Fair test-retest reliability was found for the integrated measure (ICC = 0.367), which might have been the result of the high sensitivity of ranks (used in the calculation of the integrated measure) to small changes in the original outcomes. For example, a difference as small as one second in task completion time may result in a different 'fractional rank' whereas it could be questioned whether a difference of one second is clinically relevant, given the large range of the completion times. We did use Pearson's *r* to evaluate the correlation between errors and time within rounds, while a linear correlation between two variables was to be proven. This showed that when participants moved faster, more errors were made ($r = -0.22$). The fatigue intervention resulted in the highest correlation between completion time per round and errors per round ($r = -0.28$), though similar results were found for the baseline and other interventions.

This study had some potential limitations. The medical students and interns who participated in the construct validity trial had no or very little experience with handling laparoscopic equipment. This may have led to improvement in performance in the course of time as a consequence of learning. Figure 2 implies that familiarization with the test results in a better performance during stressful and fatigued conditions. Even though a learning effect seems to be present, it did not affect our final outcomes because a randomized cross-over design was used to prevent any interference of learning effects on construct validity outcomes. Nevertheless, when the star-track test is to be used in further research, participants should get the opportunity to first familiarize themselves with the task to completely avert the influence of learning effects on study outcomes. During this study participants were to use solely their dominant hand to perform the task. In further research the effect of physical fatigue and mental stress on the manual dexterity of both hands should be investigated, while overall manual dexterity improves by opposite hand training due to interlimb transfer of motor learning [32,33]. Furthermore, the condition in which a participant performs the task, such as quantity of physical effort prior to testing, should be standardized when the star-track test is used as an evaluative tool in research. This could interfere with a participant's performance, even though this was not the case in this study. At last, the impulse counter automatically provides audio feedback and registers errors when there is contact between the stylus and the non-anodized part of the aluminum plate. However, since the impulse counter had the tendency to register multiple errors due to on and off contact between the stylus and the plate as a consequence of the subject's effort to get back on the star-track. We let a researcher monitor and count errors to make sure the correct amount of errors was registered.

In conclusion, the star-track test for manual dexterity is a valid and reliable tool to evaluate physical fatigue and mental stress in a laparoscopic setting. An integrated measure can be used to get a complete overview of the impact of these influences on a laparoscopic surgeon's manual dexterity. However, the separate evaluation of the difference in number of errors and completion time should be preferred whereas this eases interpretation of outcomes. The star-track test is capable of measuring the characteristics of manual dexterity after exposure to fatigue and/or stress in a fast and easy assessable way, and can therefore be used in further research to reflect on a surgeon's accuracy and manual dexterity in laparoscopic surgery.

Declaration of interest

Kim Platte, Chantal Alleblas, Joanna IntHout and Theodoor Nieboer have no conflict of interest or financial ties to disclose.

ORCID

Kim Platte http://orcid.org/0000-0002-4968-0647
Joanna Inthout http://orcid.org/0000-0002-6127-0747
Theodoor E. Nieboer http://orcid.org/0000-0002-4227-1461

References

[1] Arora S, Sevdalis N, Nestel D, et al. The impact of stress on surgical performance: a systematic review of the literature. Surgery. 2010;147:318–330.

[2] Berguer R, Smith WD, Chung YH. Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. Surg Endosc. 2001;15:1204–1207.

[3] Wauben LS, van Veelen MA, Gossot D, et al. Application of ergonomic guidelines during minimally invasive surgery: a questionnaire survey of 284 surgeons. Surg Endosc. 2006;20:1268–1274.

[4] Park A, Lee G, Seagull FJ, et al. Patients benefit while surgeons suffer: an impending epidemic. J Am Coll Surg. 2010;210:306–313.

[5] Alleblas CC, de Man AM, van den Haak L, et al. Prevalence of Musculoskeletal Disorders Among Surgeons Performing Minimally Invasive Surgery: A Systematic Review. Ann Surg. 2017;266:905–920.

[6] Huysmans MA, Hoozemans MJ, van der Beek AJ, et al. Fatigue effects on tracking performance and muscle activity. J Electromyogr Kinesiol. 2008;18: 410–419.

[7] Daruwalla J, Marlow N, Field J, et al. Effect of fatigue on laparoscopic skills: a comparative historical cohort study. ANZ J Surg. 2014;84:137–142.

[8] Chandra S, Hayashibe M, Thondiyath A. Dominant component in muscle fatigue induced hand tremor during laparoscopic surgical manipulation. Conf Proc IEEE Eng Med Biol Soc. 2014;2014:6539–6542.

[9] Rodrigues SP, Ter Kuile M, Dankelman J, et al. Patient safety risk factors in minimally invasive surgery: a validation study. Gynecol Surg. 2012;9: 265–270.

[10] Oropesa I, Sanchez-Gonzalez P, Lamata P, et al. Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery. J Surg Res. 2011;171:e81–95.

[11] Montanari E, Schwameis R, Louridas M, et al. Training on an inexpensive tablet-based device is equally effective as on a standard laparoscopic box trainer: A randomized controlled trial. Medicine. 2016;95:e4826.

[12] Oropesa I, Chmarra MK, Sanchez-Gonzalez P, et al. Relevance of motion-related assessment metrics in laparoscopic surgery. Surg Innov. 2013;20:299–312.

[13] Derossis AM, Fried GM, Abrahamowicz M, et al. Development of a model for training and evaluation of laparoscopic skills. Am J Surg. 1998;175:482–487.

[14] Chmarra MK, Grimbergen CA, Jansen FW, et al. How to objectively classify residents based on their psychomotor laparoscopic skills? Minim Invasive Ther Allied Technol. 2010;19:2–11.

[15] Chmarra MK, Jansen FW, Grimbergen CA, et al. Retracting and seeking movements during laparoscopic goal-oriented movements. Is the shortest path length optimal? Surg Endosc. 2008;22:943–949.

[16] Zendejas B, Ruparel RK, Cook DA. Validity evidence for the Fundamentals of Laparoscopic Surgery (FLS) program as an assessment tool: a systematic review. Surg Endosc. 2016;30:512–520.

[17] Dorion D, Darveau S. Do micropauses prevent surgeon's fatigue and loss of accuracy associated with prolonged surgery? An experimental prospective study. Ann Surg. 2013;257:256–259.

[18] Kildebro N, Amirian I, Gogenur I, et al. Test re-test reliability and construct validity of the star-track test of manual dexterity. Peer J. 2015;3:e917.

[19] Sari V, Nieboer TE, Vierhout ME, et al. The operation room as a hostile environment for surgeons: physical complaints during and after laparoscopy. Minim Invasive Ther Allied Technol. 2010;19:105–109.

[20] Aaron DH, Jansen CW. Development of the Functional Dexterity Test (FDT): construction, validity, reliability, and normative data. Hand Ther. 2003;16:12–21.

[21] Piscione J, Gamet D. Effect of mechanical compression due to load carrying on shoulder muscle fatigue during sustained isometric arm abduction: an electromyographic study. Eur J Appl Physiol. 2006;97: 573–581.

[22] Parr J, Borsa P, Fillingim R, et al. Psychological influences predict recovery following exercise induced shoulder pain. Int J Sports Med. 2014;35: 232–237.

[23] Wang DM, Wang J, Ge LZ. [Relationship between surface electromyographic signal (sEMG) changes and subjective assessment of muscle fatigue during isometric contractions]. Space Med Med Eng (Beijing). 2004;17:201–204.

[24] Goodell KH, Cao CG, Schwaitzberg SD. Effects of cognitive distraction on performance of laparoscopic surgical tasks. J Laparoendosc Adv Surg Tech Part A. 2006;16:94–98.

[25] Silverman DG, O'Connor TZ, Brull SJ. Integrated assessment of pain scores and rescue morphine use during studies of analgesic efficacy. Anesth Analg. 1993;77:168–170.

[26] Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics. 1977;33:363–374.

[27] Koch GG. Intraclass Correlation Coefficient. Encyclopedia of Statistical Sciences: John Wiley & Sons, Inc.; 2004.

[28] Dalager T, Sogaard K, Bech KT, et al. Musculoskeletal pain among surgeons performing minimally invasive surgery: a systematic review. Surg Endosc. 2017;31:516–526.

[29] Fargen KM, Turner RD, Spiotta AM. Factors That Affect Physiologic Tremor and Dexterity During Surgery: A Primer for Neurosurgeons. World Neurosurg. 2016;86:384–389.

[30] Francis NK, Hanna GB, Cuschieri A. The performance of master surgeons on the Advanced Dundee Endoscopic Psychomotor Tester: contrast validity study. Arch Surg. 2002;137:841–844.

[31] van Hove PD, Tuijthof GJ, Verdaasdonk EG, et al. Objective assessment of technical surgical skills. Br J Surg. 2010;97:972–987.

[32] Nieboer TE, Sari V, Kluivers KB, et al. A randomized trial of training the non-dominant upper extremity to enhance laparoscopic performance. Minim Invasive Ther Allied Technol. 2012;21:259–264.

[33] Middleton KK, Hamilton T, Tsai PC, et al. Improved nondominant hand performance on a laparoscopic virtual reality simulator after playing the Nintendo Wii. Surg Endosc. 2013;27:4224–4231.