

# The National Aeronautics and Space Administration-task load index: NASA-TLX: evaluation of its use in surgery

Stephen W. Bell <sup>\*,†</sup> Joseph C. H. Kong <sup>\*,†</sup> David A. Clark <sup>‡</sup> Peter Carne,<sup>\*,§</sup> Stewart Skinner,<sup>\*,§¶||</sup> Stephen Pillinger,<sup>\*\*</sup> Paul Burton<sup>\*,†††</sup> and Wendy Brown <sup>\*,†††</sup>

\*Department of Surgery, Central Clinical School, Monash University, Melbourne, Victoria, Australia

†Department of General Surgery, Alfred Hospital, Melbourne, Victoria, Australia

‡Department of Surgery, St Vincent's Private Hospital Northside, Brisbane, Queensland, Australia

§Department of Surgery, Cabrini Hospital, Melbourne, Victoria, Australia

¶Department of Surgery, Peninsula Private Hospital, Melbourne, Victoria, Australia

||Department of Surgery, Frankston Hospital, Melbourne, Victoria, Australia

\*\*Department of Surgery, Sydney Adventist Hospital, Sydney, New South Wales, Australia and

††Department of Surgery, The Avenue Private Hospital, Melbourne, Victoria, Australia

## Key words

bariatric surgery, NASA TLX, surgical research.

## Correspondence

Stephen W. Bell, Department of Surgery, Central Clinical School, Monash University, Melbourne, VIC, Australia.

Email: [swbell@ccgroup.net.au](mailto:swbell@ccgroup.net.au)

**S. W. Bell** MBBS, FRACS; **J. C. H. Kong** MBChB, FRACS, PhD; **D. A. Clark** MBBS, FRACS, FRCSEd;

**P. Carne** MBBS, FRACS; **S. Skinner** MBBS, FRACS, PhD; **S. Pillinger** MBChB, FRACS;

**P. Burton** MBBS, FRACS, PhD; **W. Brown** MBBS, FRACS, FACS, PhD.

This paper has not been communicated to a meeting or society previously.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Accepted for publication 20 May 2022.

doi: 10.1111/ans.17830

## Introduction

The technical difficulty each individual operation poses for a surgeon is challenging to measure. Each surgeon understands after an operation how difficult it was, but this is necessarily a subjective

## Abstract

**Background:** The technical difficulty an operation creates for a surgeon is difficult to measure. Current measures are poor surrogates. In both research and teaching settings it would be valuable to be able to accurately measure this degree of difficulty. The National Aeronautics and Space Administration Task Load Index (NASA TLX) is a multi-dimensional scale designed to obtain workload estimates relating to a task. This study aimed to evaluate the NASA TLX as an objective measure of technical difficulty of an operation.

**Methods:** Seven surgeons performed 127 pre-defined operations (minimally invasive right hemicolectomy & re-do bariatric surgery) and recorded a NASA TLX score after each operation. These scores were compared to numerous clinical parameters and the score was correlated with the subjective measure of whether the surgeon categorized the operation as “easy”, “moderate” or “difficult”.

**Results:** The NASA TLX score was significantly correlated with operative duration, blood loss, previous abdominal surgery and the surgeons' assessment of difficulty. It did not correlate with intra-operative or post-operative complications, conversion to open surgery or length of stay.

**Conclusions:** The NASA TLX score provides a graded numerical score that correlated significantly with the surgeon's assessment of the technical difficulty, and with operative duration, intra-operative blood loss and previous abdominal surgery. This novel application of this tool could be employed in both research and teaching settings to score surgical difficulty and monitor a trainee's proficiency over time.

assessment. In a research or teaching setting it would be valuable to be able to objectively and quantitatively measure this degree of difficulty. Any new surgical technique or intervention needs to be evaluated for improvements in clinical outcomes and it would also be beneficial to measure any improvement in technical difficulty for

the surgeon. The ability of a trainee to master an operation, and the ability to objectively measure a trainee's progress would help define their individual learning curve.

Current measures of surgical difficulty include: duration of operation, blood loss / transfusion requirement, conversion from minimally invasive surgery, post-operative complications and hospital length of stay. A history of previous abdominal surgery may add to the difficulty of an operation. All are weak surrogate markers with inherent limitations. A recent study from Japan noted that an objective numerical rating index to help surgeons assess the estimated surgical difficulty has not been developed yet. The authors studied predictors of surgical difficulty in total mesorectal excision (TME) for rectal cancer and concluded that studies in this area had reported inconsistent results.<sup>1</sup>

The National Aeronautics and Space Administration Task Load Index (NASA-TLX) is a multi-dimensional scale designed to obtain workload estimates from one or more operators while they are performing a task or immediately afterwards. It was developed by the Human Performance Group at NASA's Ames Research Center.<sup>2,3</sup> The authors developed a technique that involved sub-scale selection and the weighted averaging approach, and this resulted in a tool that has proven to be reasonably easy to use and reliably sensitive to experimentally important manipulations. Its use has spread far beyond its original application (aviation), focus (crew complement), and language (English).

A survey of 550 studies in which NASA-TLX was used or reviewed was published in 2006. It summarized the environments in which it has been applied, the types of activities the raters performed, other variables that were measured that did (or did not) covary, methodological issues, and lessons learned.<sup>4</sup> Most of the early studies using this tool addressed some sort of question about interface design or evaluation: Visual and/or auditory displays (31%), vocal and/or manual input devices (11%), virtual or augmented vision (6%). In addition, these and other studies also examined the impact of underlying systems such as automation and decision aids (26%), digital data link (3%), caution, advisory and warning systems (4%), and new types of information on operator workload. These were performed by military or government organizations. More recently it has been used to study automobile drivers and in the medical field (nursing and anaesthetics).<sup>5-7</sup> The majority of uses have involved individual, definable tasks; and it is not clear if this tool will be useful in the complex setting of a long operation, when multiple smaller tasks are being performed over a period of time.

The aim of this study was to perform an initial evaluation of the NASA TLX score in measuring the technical difficulty of two moderately difficult operations (minimally invasive right hemicolectomy and redo sleeve gastrectomy), and to correlate this score against other commonly used measures of technical difficulty.

## Hypothesis

That the NASA TLX can provide a numerical score that accurately reflects the technical difficulty of a surgical operation.

## Methods

### Study design

A prospective observational study was undertaken in seven tertiary hospitals throughout Australia from 1/11/2020 to 24/12/2021.

Seven experienced colorectal and upper gastro-intestinal surgeons and patient outcomes were studied. The study did not alter the clinical or operative management of the patient. The surgeon completed a 3-min NASA TLX score immediately after the operation, along with operative and clinical outcome parameters. Ethics approval was granted by the Alfred Hospital Office of Ethics & Research Governance (approval number: 61221). The trial was registered with the Australian and New Zealand Clinical Trials Registry (ANZCTR) (Registry Number: 126200001943).

## Participants

Patients were considered eligible if:

Inclusions:

- (1) a) Laparoscopic right hemicolectomy  
b) Robotic right hemicolectomy
- (2) Laparoscopic re-do bariatric surgery

These operations were chosen for this study as they were considered 'moderately difficult', and therefore thought to be able to show differences in degrees of difficulty. They spanned two specialties therefore adding some variation in technique and surgeons.

Exclusions:

- (1) Emergency or urgent surgery.
- (2) Significant comorbidities requiring elective admission to ICU or HDU
- (3) Intra-operative mortality.

## Variables

The NASA TLX was completed immediately after the operation and performed on the official NASA TLX application on a mobile device.<sup>8</sup> Stage one of the evaluation was performed by initially choosing between two paired factors (e.g., Effort versus Mental Demand) depending on which of these factors was more important to the surgeon's experience of workload during the operation ('pairwise comparison'). This process is completed for 15 pairs offered and produces a weighting for each factor.

The six factors are:

Mental demand: How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?

Physical demand: How much physical activity was required? Was the task easy or demanding, slack or strenuous?

Temporal demand: How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?

Overall performance: How successful were you in performing the task? How satisfied were you with your performance?

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Frustration level: How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?

Stage two was to assess each individual factor in the ratings scale. The ratings scale assesses the experience the surgeon had during the operation in six domains. The surgeon gives a numerical score on each of the six visual analogue scales that best matches their experience at the time (0–100).

The final score is then calculated by multiplying the rating scale score by the weighting score, and the total added score is then divided by 15 (the sum of the weights). This produces a final single numerical score ranging from 0 to 600, with higher scores representing a more difficult task. (See supplement 1 for description and pictures for further clarification).

NASA TLX scores were then correlated with the following clinical parameters:

- (1) History of previous abdominal surgery (apart from index bariatric surgery)
- (2) Duration of operation
- (3) Estimated blood loss
- (4) Transfusion requirement
- (5) Intra-operative complications
- (6) Conversion to open surgery
- (7) In-hospital complications (scored using the Clavien-Dindo Classification)<sup>9</sup>
- (8) Length of stay

## Biases

Individual reporting bias was addressed through numerous surgeons in geographically remote sites, across two surgical specialties.

## Sample size

There was limited published data to adequately perform an accurate sample size calculation. A prior non-published small study

analysing the NASA-TLX score evaluated 12 patients undergoing laparoscopic rectal cancer resection. NASA TLX scores in these patients were observably different between a group of seven who received pre-operative Very Low Energy Diet (129), and five who did not (216). It was therefore estimated that a sample size of 20 operations per surgeon would be adequate to show a difference. A total of 140 patients overall were sought, with each surgeon expected to recruit 20 patients.

Appropriate written consent was obtained from both the patients and the surgeons. Data was collected prospectively on the secure and encrypted RedCap online database, with the assistance of the department of Technology, Risk and Information Security at Monash University.

## Statistical analysis

The NASA-TLX scores were analysed according to the surgeons' categorisation of operative difficulty. Pearson's chi-square or Fisher's exact test were used for categorical data and Independent Sample-Median test for continuous data, reported as median and range. Pearson correlation was used to derive the significance of association between NASA TLX score against reported surgeon's perception of individual case difficulty and the other clinical parameters. All statistical analysis was performed using SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp. A *p*-value of <0.05 was considered significant.

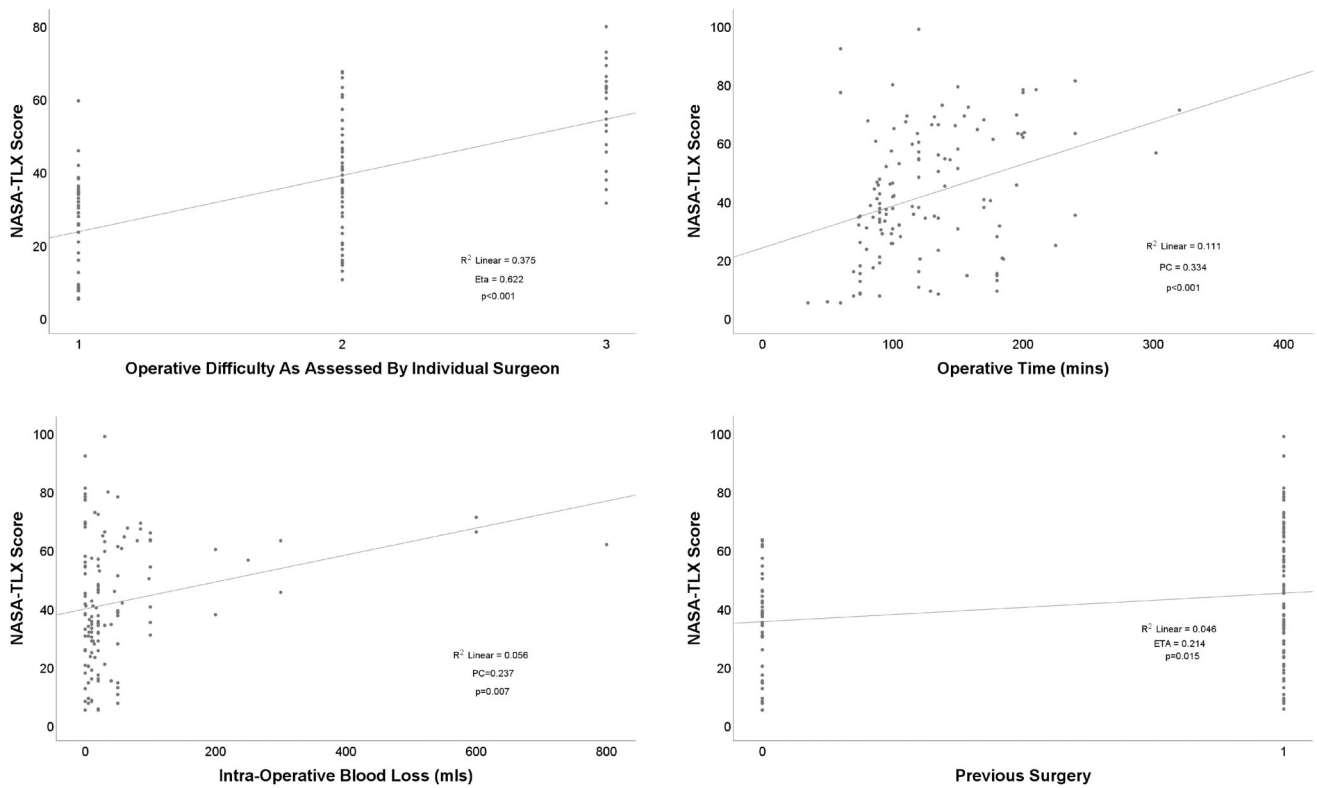
**Table 1** Patient and operative characteristics versus surgeons' rating of operative difficulty

Variables	Easy (%) <i>n</i> , (%)	Moderate (%)	Difficult (%)	<i>p</i>
Gender				
Male	9 (23.7)	18 (36.7)	12 (54.5)	
Female	29 (76.3)	31 (63.3)	10 (45.5)	0.055
Age (range)	63.6 (25.2–91.5)	70.8 (24.7–94.5)	65.9 (24.8–86.1)	0.742
Hospital				
Public	14 (36.8)	11 (22.4)	10 (45.5)	
Private	24 (63.2)	38 (77.6)	12 (54.5)	0.117
Operation				
Right hemicolectomy	30 (78.9)	42 (85.7)	16 (72.7)	
Redo-bariatric surgery	8 (21.1)	7 (14.3)	6 (27.3)	0.414
Minimally invasive technique				
Robotic	3 (7.9)	10 (20.4)	3 (13.6)	
Laparoscopic	35 (92.1)	39 (79.6)	19 (86.4)	0.259
Surgeon seniority				
Consultant	32 (84.2)	44 (89.8)	18 (81.8)	
Fellow	6 (15.8)	5 (10.2)	4 (18.2)	0.602

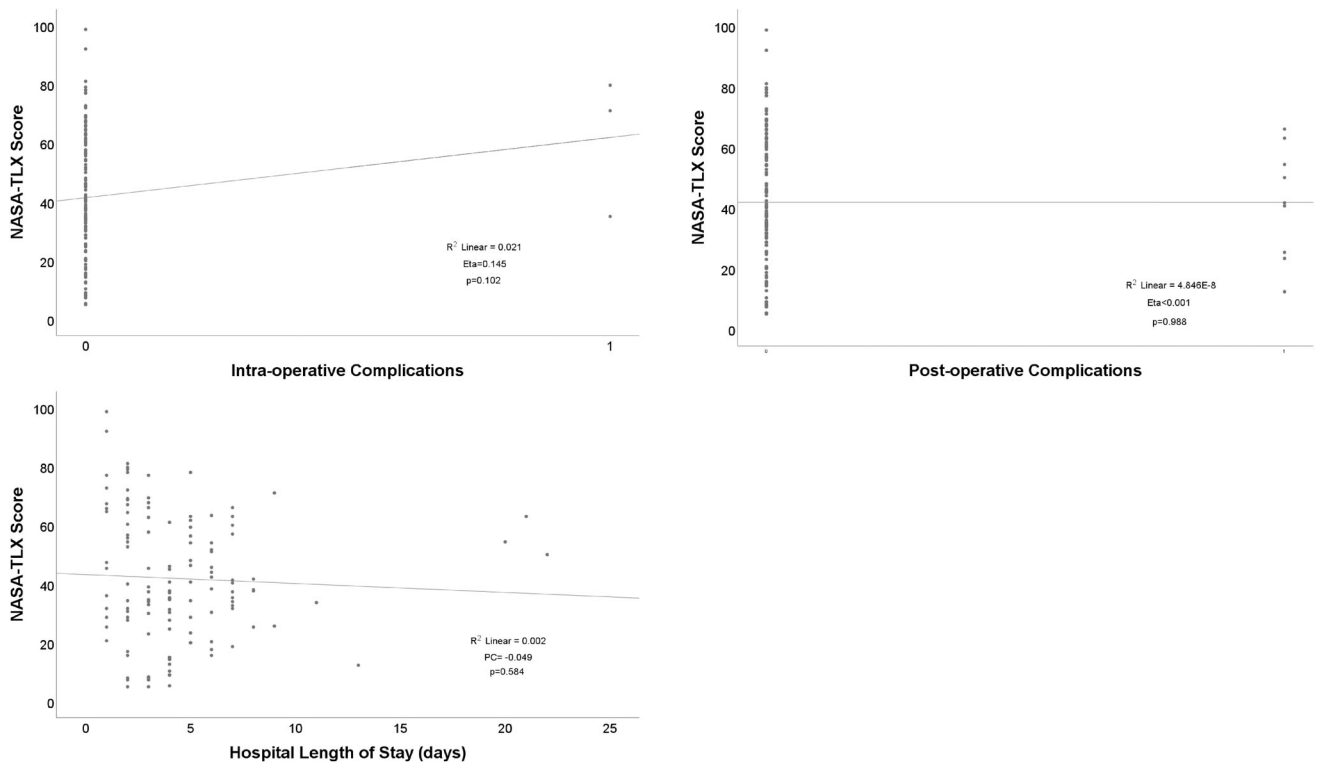
**Table 2** Clinical parameters versus surgeons' rating of operative difficulty

Variables	Easy (%) <i>n</i> , (%)	Moderate (%)	Difficult (%)	<i>p</i>
Previous surgery				
No	16 (42.1)	20 (40.8)	6 (27.3)	
Yes	22 (57.9)	29 (59.2)	16 (72.7)	0.475
Transfusion requirement				
No	37 (97.4)	47 (97.9)	22 (100)	
Yes	1 (2.6)	1 (2.1)	0	0.757
Estimated blood loss (range)	10 (0–100) mls	20 (0–300) mls	82.5 (15–800) mls	<0.001
Operative time (range)	90 (35–180) mins	105 (75–240) mins	160 (90–320) mins	<0.001

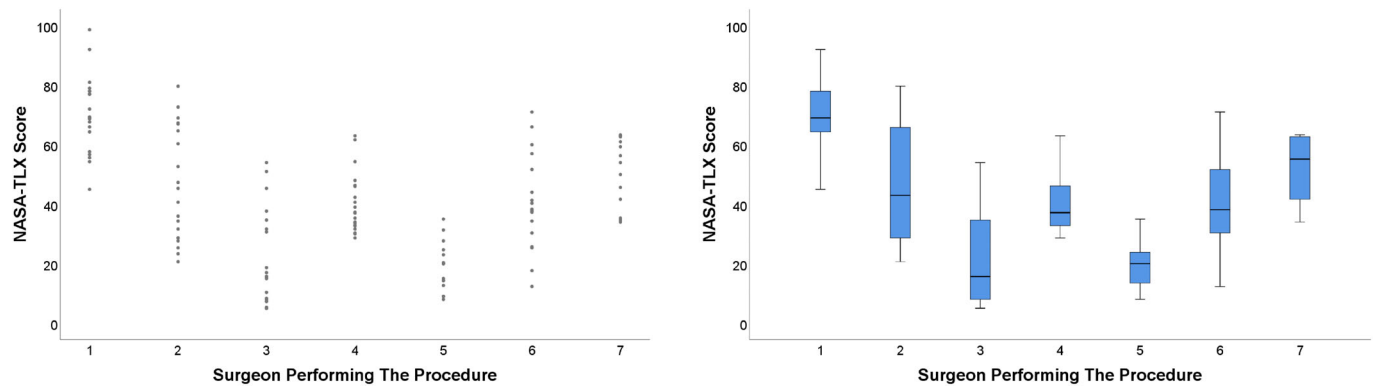
Abbreviation: mls, millilitres.



**Fig. 1.** NASA-TLX score versus surgeon’s assessment, operative time, blood loss & previous surgery. NASA-TLX: National Aeronautics and Space Administration Task Load Index. Surgeon’s assessment: 1 = Easy, 2 = Moderate, 3 = Difficult; Previous surgery: 0 = no previous surgery, 1 = prior surgery.



**Fig. 2.** NASA TLX score vs. length of stay, intra- and post-operative complications. NASA-TLX: National Aeronautics and Space Administration Task Load Index. Complications: 0 = no; 1 = yes.



**Fig. 3.** Individual NASA TLX scores generated by each surgeon median and IQR (whisker plot) and 95% confidence interval (box plot). NASA-TLX: National Aeronautics and Space Administration Task Load Index.

## Results

Seven surgeons enrolled 127 patients, with four surgeons enrolling the planned 20 patients and three having enrolment impaired by restrictions on surgery due to the COVID 19 pandemic, enrolling 18, 15 and 14 patients. These were six consultant surgeons (four colorectal and two bariatric), all with >10 years of consultant experience, and one supervised post fellowship colorectal trainee (who performed 14 laparoscopic right hemicolectomies). The participating hospitals were both public (32% of episodes) and private (68% of episodes), and located in Melbourne, Sydney and Brisbane, Australia. There were 72 laparoscopic right hemicolectomies, 15 robotic right hemicolectomies, and 40 re-do bariatric operations. The patients were 64% female, with one surgeon being female.

Table 1 shows the patient characteristics against the surgeons' assessment of the degree of difficulty ('easy', 'moderate' or 'difficult'), while Table 2 correlates the variables against this "traffic light" assessment by the surgeons. One surgeon did not complete the traffic light assessment, leaving 109 operations for assessment. Two patients received a blood transfusion, and there were only two conversions to open surgery and three intra-operative complications. The surgeon's assessment correlated with estimated blood loss ( $p < 0.001$ ) and operation duration ( $p < 0.001$ ), but not with transfusion requirement ( $p = 0.757$ ) nor previous surgery ( $p = 0.475$ ).

There was a statistically significant correlation when the NASA TLX score was compared to the surgeons' assessment of surgical difficulty ( $R^2 = 0.375$ ,  $p < 0.001$ ), a history of previous abdominal surgery ( $R^2 = 0.046$ ,  $p = 0.015$ ), intra-operative blood loss ( $R^2 = 0.056$ ,  $p = 0.007$ ) and operation duration ( $R^2 = 0.111$ ,  $p < 0.001$ ) (Fig. 1). Post-operative complications and in-hospital length of stay did not show a statistical correlation (Fig. 2).

Figure 3 graphically demonstrates the spread of individual scores for each surgeon. This is important to show as it suggests the NASA TLX can identify differences between individual operations performed by the same surgeon.

## Discussion

The ability to accurately quantify the technical difficulty of an operation is a valuable research and teaching tool. The current

parameters used to assess operative difficulty are imprecise and have only been recorded as uncorrelated observations rather than robust measurements that can be incorporated in detailed analysis. The NASA TLX offers a single numerical score for a task that appears to accurately reflect the technical difficulty experienced by the surgeon.

The NASA TLX is not a tool for predicting how difficult an individual operation may be, but is designed to measure how difficult it was to perform. It could be useful in determining the perceived difficulty of a specific surgical technique, changes in difficulty relating to training or the best methods of training to accelerate improvements through the learning curve. The benefits of validating this tool in surgery include its potential use in studies measuring the impact of interventions that reduce surgical difficulty, and the impact of training techniques.

The correlation with the surgeon's categorical assessment of whether the operation was easy, moderate or difficult is a strong reflection of this. The numerical score offers more granularity to the simple three option categorical description, and presents a more robust tool for analysis. The NASA TLX correlated with the surgeon's assessment as well as blood loss, operation duration and previous surgery; however the surgeon's assessment only correlated with blood loss and operation duration. This suggests that the NASA TLX could be more sensitive to changes in technical difficulty.

The duration of an operation may reflect the degree of difficulty, but does not take in to account other factors, such as the time of day or whether the surgeon was feeling tired or stressed. Division of adhesions is time consuming but may not be difficult to perform. Other factors that can make an operation longer without necessarily adding to the degree of difficulty include the experience or seniority of the surgeon and any personnel or equipment issues during surgery. There will have been some overlap in assessment between the duration of operation and the history of previous abdominal surgery, as the presence of adhesions will relate to both. The NASA TLX score correlated significantly with both of these parameters. Interestingly the NASA TLX did correlate with a history of previous abdominal surgery, however the surgeon's assessment did not, suggesting that the NASA TLX may have been more sensitive at detecting a difference.

The lack of correlation with intra- and post-operative complications does not detract from the utility of the NASA TLX. Indeed the occurrence of complications may be more dependant on patient related factors rather than operative factors. In this study, intra-operative (three) and post-operative (nine) complications were infrequent events. A technically difficult operation performed well may still not see post-operative complications. Intra-operative complications are generally rare, and are likely attributed to a technically more difficult operation. In a larger data set or with a more difficult operation being assessed, this may be born out more clearly against the NASA TLX, and could be an area for further study. The lack of correlation with in hospital length of stay is also not a concern for the utility of the NASA TLX as there are many factors that can result in a longer hospital stay that do not relate to the technical difficulty of an operation. These include coexistent medical illness, social and post-discharge arrangements, and the hospital setting (public or private admission).

The individual scores produced by each surgeon were moderately spread, suggesting that the NASA TLX can adequately differentiate between degrees of difficulty (see Fig. 3). This is possibly one of the more important aspects of this study as the weaknesses of the other clinical variables, along with the bluntness of the categorical three-tiered assessment are overcome by a more detailed scoring system. The NASA TLX provides a graded ordinal numerical score that lends itself well to individual or cohort study and statistical analysis.

The experience from the ADIPOSE study, which randomized obese patients into two groups for pre-operative weight loss, highlighted this potential benefit. There was a clear observable difference in the scores between the treatment arm and the control arm (NASA TLX scores of 129 versus 216). The ADIPOSE study did not accrue an adequate number of patients to produce a statistically significant result, however the NASA TLX appeared to be beneficial in a “real world” research and clinical setting.<sup>10</sup>

At the time of our study design there were no publications of the use of the NASA TLX in surgery, and we felt that it was reasonable to perform an initial assessment of this tool to assess an association with technical difficulty. While this study was being performed there have been three publications that have also reported their initial assessments. Lowndes *et al.* used a modified NASA TLX in 662 operations, retrospectively collecting patient and procedural factors. They demonstrated that the NASA TLX score varied across specialties, and also noted that when workload was higher the operation duration was longer.<sup>11</sup> Law *et al.* (from the same institution as Lowndes *et al.*) also published a study of 238 operations showing that procedure type and surgical approach (including robotic surgery) effected workload.<sup>12</sup> Zheng *et al.* compared the NASA TLX score with the surgeon’s blink frequency in 42 operations, reporting that higher blink frequency was associated with higher mental workload.<sup>13</sup> These studies have not prospectively correlated the NASA TLX with other possible measures of surgical difficulty. The strengths in this current study were the prospective nature of the data collection, along with the

direct correlation with other potential markers of technical difficulties. There has not been a definitive validation of the NASA TLX in the surgical sphere as yet, and we believe that there is now scope to expand on this initial work in larger and more broad reaching studies.

A weaknesses of this study was that only two moderately difficult operations were assessed, and that the majority of surgeons were experienced consultant surgeons. This may limit the generalisability to trainees and will need to be separately evaluated in a trainee cohort. It would be beneficial to study the NASA TLX further in a broader range of complicated operations. This study included one post-fellowship trainee, however there was not enough statistical power to make a conclusion about the utility of the NASA TLX across different levels of experience. There was no “gold standard” measure of operative difficulty to assess the NASA TLX against. The evaluated outcome parameters are in common use in trials reporting clinical difficulty measures, and there is potential that these NASA-TLX data will assist investigators in future trial design.

## Conclusions

The NASA TLX score provides a graded ordinal numerical score that correlated significantly with the surgeon’s assessment of the technical difficulty of an operation, along with a history of previous abdominal surgery, operative time and intra-operative blood loss. This provides a potential tool that can be used in both research and teaching settings to measure changes in technical difficulty with the introduction of a new technique or technology, along with a trainee’s progress over time.

## Acknowledgement

Open access publishing facilitated by Monash University, as part of the Wiley - Monash University agreement via the Council of Australian University Librarians.

## Conflict of interest

None declared.

## Author contributions

**Stephen W. Bell:** Conceptualization; formal analysis; investigation; methodology; project administration; writing – original draft; writing – review and editing. **Joseph C. H. Kong:** Formal analysis; writing – original draft; writing – review and editing. **David A. Clark:** Data curation; writing – original draft; writing – review and editing. **Peter Carne:** Data curation; writing – original draft; writing – review and editing. **Stewart Skinner:** Data curation; writing – original draft; writing – review and editing. **Stephen Pillinger:** Data curation; writing – original draft; writing – review and editing. **Paul Burton:** Data curation; writing – original draft; writing – review and editing. **Wendy Brown:** Conceptualization; methodology; supervision; writing – original draft; writing – review and editing.

## References

1. Kawada K, Sakai Y. Can we predict surgical difficulty of rectal surgery? *Ann. Laparosc. Endosc. Surg.* 2018; **3**: 44.
2. NASA. Nasa Task Load Index (TLX) v. 1.0 Manual, 1986.
3. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research (PDF). In: Hancock PA, Meshkati N (eds). *Human Mental Workload, Advances in Psychology* 52. Amsterdam: North Holland, 1988; 139–83. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9).
4. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 Years Later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. *Sage J.* 2006; **50**: 904–8.
5. Young G, Zavelina L, Hooper V. Assessment of workload using NASA task load index in Perianesthesia nursing. *J. Perianesth. Nurs.* 2008; **23**: 102–10.
6. Davis DHJ, Oliver M, Byrne AJ. A novel method of measuring the mental workload of anaesthetists during simulated practice. *Br. J. Anaesth.* 2009; **103**: 665–9.
7. Hoonakker P, Carayon P, Gurses A *et al.* Measuring workload of ICU nurses with a questionnaire survey: the NASA task load index (TLX). *IIE Trans. Healthc. Syst. Eng.* 2011; **1**(2): 131–43.
8. NASA TLX. Available from URL: <https://humansystems.arc.nasa.gov/groups/tlx/index.php>
9. Dindo D, Demartines N, Clavien P-A. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann. Surg.* 2004; **240**: 205–13.
10. Bell S, Venchiarutti R, Warrier S, Stevenson A, Solomon M. A perspective on surgical randomised controlled trials. *ANZ J. Surg.* 2019; **89**: 998–9.
11. Lowndes BR, Forsyth KL, Blocker RC *et al.* NASA-TLX assessment of surgeon workload variation across specialties. *Ann Surg.* 2020 Apr; **271**: 686–92.
12. Law KE, Lowndes BR, Kelley SR *et al.* NASA-task load index differentiates surgical approach: opportunities for improvement in colon and Rectal surgery. *Ann. Surg.* 2020; **271**: 906–12.
13. Zheng B, Jiang X, Tien G, Meneghetti A, Panton N, Atkins S. Workload assessment of surgeons: Correlation between NASA TLX and blinks. Sages.Session Number: SS02 – Instrumentation/Ergonomics. Program Number: S008. Available from URL: <https://www.sages.org/meetings/annual-meeting/abstracts-archive/workload-assessment-of-surgeons-correlation-between-nasa-tlx-and-blinks/>

## Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Appendix S1.** Supporting Information.