Validation of the National Aeronautics and Space Administration-Task Load Index as a tool to evaluate the learning curve for endoscopy training

Rachid Mohamed MD FRCPc1, Maitreyi Raman MD FRCPc MSc1, John Anderson BSc FRCP(Edin) MD2, Kevin McLaughlin MD MRCP PhD3, Alaa Rostom MD FRCPc MSc1, Sylvain Coderre MD FRCPc MSc1

BACKGROUND: Although workplace workload assessments exist in different fields, an endoscopy-specific workload assessment tool is lacking.

OBJECTIVE: To validate such a workload tool and use it to map the progression of novice trainees in gastroenterology in performing their first endoscopies.

METHODS: The National Aeronautics and Space Administration Task Load Index (NASA-TLX) workload assessment tool was completed by eight novice trainees in gastroenterology and 10 practicing gastroenterologists/surgeons. An exploratory factor analysis was performed to construct a streamlined endoscopy-specific task load index, which was subsequently validated. The 'Endoscopy Task Load Index' was used to monitor progression of trainee exertion and self-assessed performance over their first 40 procedures.

RESULTS: From the factor analysis of the NASA-TLX, two principal components emerged: a measure of exertion and a measure of self-efficacy. These items became the components of the newly validated Endoscopy Task Load Index. There was a steady decline in self-perceived exertion over the training period, which was more rapid for gastroscopy than colonoscopy. The self-efficacy scores for gastroscopy rapidly increased over the first few procedures, reaching a plateau after this period of time. For colonoscopy, there was a progressive increase in reported self-efficacy over the training period.

DISCUSSION: The present study validated an Endoscopy Task Load Index that can be completed in<1 min. Practical implications of such a tool in endoscopy education include identifying periods of higher perceived exertion among novice endoscopists, facilitating appropriate levels of guidance from trainers.

Key Words: Cognitive load; Endoscopy training; Procedural skills; Workload assessment

The road to acquiring competence, and at times excellence, in the performance of any skill requires a combination of innate biological capacities, dedicated teachers and many hours of training. The process of skills acquisition has been described as a sequential process involving three major phases (1). The first phase, or novice phase, involves intense concentration to fully understand the activity and avoid making mistakes. The second phase is an evolution to a more fluid, less cognitively arduous step in which trainees begin to perform at an acceptable level, with fewer major mistakes. The final phase involves a process of automation, in which the skill is precisely and smoothly performed with little or no conscious cognitive involvement. Different terminology has been used to describe a similar sequence of events. The terms 'unconscious' and 'conscious' incompetence have been used to describe the early training stage, evolving to conscious competence (akin to Ericsson's second phase) and, finally, unconscious competence for the more automated phase of skill acquisition (2).

First described in 1971, endoscopy of the entire colon, or colonoscopy, is a common diagnostic procedure performed worldwide (3). During a colonoscopy training session, a novice endoscopist must attend to myriad sensory stimuli: visual stimuli from the endoscopic image of the colon on the monitor; verbal stimuli from the patient, nurse and trainer; as well as tactile/proprioceptive stimuli from the nurse and trainer; as well as tactile/proprioceptive stimulus from the nurse and trainer; as well as tactile/proprioceptive stimulus from the...
In addition to these sources of cognitive load, colonoscopy also places physical demand on the trainee, such as the process of straightening the colonoscope or loop resolution, a technique critical to the successful, safe and comfortable advancement of the instrument through the colon. Examination of the upper gastrointestinal tract (ie, esophagogastroduodenoscopy [EGD]) also provides many of these workload demands, although likely to a lesser extent.

The above-mentioned sources of workload, in addition to other potential sources such as time demand and frustration/annxiety, are captured in a workload assessment tool known as the National Aeronautics and Space Administration Task Load Index (NASA-TLX). This subjective assessment of workload has been used in >500 published articles worldwide in several domains including medicine. Subjective ratings are able to tap the essence of workload, and provide a valid, sensitive and practically useful indicator.

The objectives of the present study were threefold. The first was to identify the principal components of the NASA-TLX tool when applied to endoscopy training and to then use these data to create an endoscopy-specific rating tool. Our second objective was to assess the construct validity of the endoscopy task load rating tool. Our final study objective was to use the endoscopy task load rating tool to map our trainees’ workload and perceived performance during the early phases of their training. Such attempts to measure mental effort during deliberate practice and relating these measures to the level of performance has been shown to adequately represent the efficiency of the ongoing learning processes. We hypothesized that as trainees progress through their initial endoscopies, a steady, gradual decrease in workload would occur, correlating with an equally steady improvement in performance.

METHODS

Participants

Participants were eight first-year gastroenterology residents at the University of Calgary (Calgary, Alberta). This gastroenterology residency training program lasts two years, during which residents perform, on average, 200 colonoscopies and 200 EGDs. Proficiency in both procedures is a requirement for graduating from the program. The study was conducted over two time periods, July 1 to September 30, 2009, and July 1 to September 30, 2010. There were four participants during each time period. This enabled data collection for the residents’ procedural learning curve during the first three months of their training. Ten practicing gastroenterologists and colorectal surgeons with >5 years of experience in performing colonoscopy were asked to complete the NASA-TLX rating after two consecutive colonoscopies. Before beginning the study, ethics approval from the Conjoint Health Research Ethics Board at the University of Calgary was obtained, in addition to written informed consent from each of the participants.

Materials

A slightly modified version of the NASA-TLX rating tool was used for the present study. The unmodified NASA-TLX tool was initially piloted on three experienced gastroenterologists performing colonoscopy on two occasions. After their feedback, six items on the rating scales were kept, but some of the descriptors were modified for clarity and/or to make them more relevant to endoscopy. The final rating tool had six items: mental demand, physical demand, time demand, effort, performance, and frustration and anxiety. Each of these items was rated on a visual analogue scale that was interpreted as the participant’s subjective rating of each variable.

Procedure

The present analysis was a prospective observational cohort study. The participants were asked to complete the NASA-TLX rating for each of their colonoscopies and EGD procedures during the study time period. Cognizant of the effect that patient variability may have on the workload of endoscopy, the participants were asked not to rate procedures for patients with known previously difficult colonoscopies (failed or successful), one or more pelvic surgeries, two or more abdominal surgeries, previous colonic resection, as well as patients deemed to experience excessive anxiety over the procedure. These exclusion criteria were not used for the colonoscopies rated by the practicing gastroenterologists and colorectal surgeons.

Statistical analyses

The reliability of the NASA-TLX survey was assessed using Cronbach’s α coefficient. Before performing factor analysis, the Kaiser-Meyer-Olkin (KMO) test was used to assess the appropriateness of performing this analysis on this dataset (ie, to ensure KMO statistic >0.5). Using the NASA-TLX survey as the unit of analysis, an exploratory factor analysis on the individual items of this tool was performed. This technique reduces a set of items to a smaller number of underlying principal components and, in doing so, uncovers the latent structure of the set of items. Factor analysis can evaluate discrimination by statistically testing whether two or more items differ. Items are considered to be measuring different constructs if they load most heavily on different principal components. Items that load most heavily, or converge, on the same principal component are considered to be measuring the same construct.

In the analysis, a Pearson product moment correlation matrix for the NASA-TLX items was initially constructed and then used principal component analysis to extract factors. A cut-off threshold was used for factor extraction of eigenvalue ≥1 (Kaiser rule). Factor loading was then performed on the extracted factors, followed by factor rotation using the Varimax method with Kaiser normalization. A cut-off threshold of 0.5 was used for factor loading.

Having identified the principal components of the NASA-TLX tool, simplification this tool was sought using a data-reduction technique. For this, a weighted composite score for each component was created, on which more than one item loaded. Weighting for each item corresponded to its factor loading score. Linear regression was then used to identify the minimum number of items that would allow explanation of ≥80% of the variance (R²) of each weighted composite score.

To compare NASA-TLX scores for colonoscopies performed by practicing gastroenterologists/surgeons and residents, a two-sample t-test was used and Cohen’s d as a measure of effect size. A repeated-measures ANOVA was used to evaluate whether there was a change in NASA-TLX ratings over time. For this analysis, the between-subject variable was participant and the within-subject variable was procedure number. STATA version 11.0 (StataCorp, USA) was used for the analyses.

RESULTS

Principal components of the NASA-TLX for endoscopy

In the factor analysis, all 276 surveys for colonoscopy and 128 surveys for EGD were included. The participants completed a mean of 34.5 (range 12 to 54) surveys for colonoscopy and 32 (range 13 to 45) for EGD. The alpha coefficients and KMO statistics for the colonoscopy surveys were 0.78 and 0.78, respectively, while the corresponding results for EGD were 0.90 and 0.82.

For the colonoscopy surveys, two principal components (eigenvalues of 3.0 and 1.0) were identified that explained 66% of the overall variance. Five of the six individual items loaded on the first factor, which were interpreted as ‘exertion’, while a single item loaded on the second factor (‘self-efficacy’). Factor loading scores are shown in Table 1. For the weighted exertion score, no single variable could explain ≥80% of the variance for this score, but the combination of effort and physical demand provided an R² of 0.89. Therefore, the tool was simplified to include three items – a weighted combination of effort and physical demand (hereby referred to as ‘exertion’) and performance as an indicator of ‘self-efficacy’. These two measures, exertion and self-efficacy, are based solely on endoscopists’ self-assessment (via NASA-TLX) and not on any specific objective measures of achievement (such as cecal intubation, detection rates, etc).
For analysis of the EGD surveys, the same two principal components explained 80% of the overall variance. Loading of individual items paralleled that of the colonoscopy data (Table 1), as did data reduction to simplify the rating of the exertion score. Once again, the combination of effort and physical demand provided the optimal R² reduction to simplify the rating of the exertion score. Once again, the items paralleled that of the colonoscopy data (Table 1), as did data ents explained 80% of the overall variance. Loading of individual

**TABLE 1**

Principal components of the National Aeronautics and Space Administration Task Load Index (NASA-TLX) for colonoscopy and esophagogastroduodenoscopy (EGD)

<table>
<thead>
<tr>
<th>NASA-TLX item</th>
<th>Colonoscopy Factor 1 (Exertion)</th>
<th>Colonoscopy Factor 2 (Self-efficacy)</th>
<th>EGD Factor 1 (Exertion)</th>
<th>EGD Factor 2 (Self-efficacy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental demand</td>
<td>0.77</td>
<td>0.87</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>Physical demand</td>
<td>0.63</td>
<td>0.91</td>
<td>0.62</td>
<td>0.76</td>
</tr>
<tr>
<td>Time demand</td>
<td>0.62</td>
<td>0.76</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>Effort</td>
<td>0.89</td>
<td>0.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>0.78</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frustration and anxiety</td>
<td>0.79</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**DISCUSSION**

Our final study objective was to map the evolution of exertion and self-efficacy during initial training procedures of our novice endoscopists. As shown in Figure 1, for both gastroscopy and colonoscopy there was a steady decline in self-perceived exertion over this training time period. It is notable, however, that the mean exertion score of our novices at the end of their first three months (50.2) was still substantially higher than the mean exertion score of our experts (27.8). The self-efficacy scores for gastroscopy rapidly increase over the first few procedures, reaching a plateau after this period of time. For colonoscopy, there is a progressive increase in reported self-efficacy over the first three quarters of procedures, which is followed by a drop in self-efficacy scores over the next quartile. The final mean self-efficacy score for the novices is much lower than that for their expert counterparts.

From these findings, it appears that for gastroscopy, there is a rapid acquisition of basic skills resulting in a corresponding decrease in exertion over the first training procedures, with a rapidly increasing perception of self-efficacy that quickly reaches a plateau to a relatively stable level. This finding may reflect the ease of intubation during gastroscopy as well as the relatively consistent anatomy encountered. For the most part, once intubation of the esophagus is mastered, the manoeuvres to navigate through the esophagus into the stomach and duodenum are consistent. The higher exertion and lower self-efficacy reported in the first few procedures may represent the intubation process, which can be challenging and create a high amount of cognitive and physical demand. Once competence and experience is attained with esophageal intubation, the trainee can relax and dedicate more time and is believed to fine tune the skills necessary for the remainder of the procedure.

For colonoscopy, a similar pattern of decreasing exertion over the first training procedures emerges, although a more gradual decline in exertion scores exists compared with gastroscopy. This is perhaps not surprising because colonoscopy is generally believed to be more intrinsically challenging and demanding than gastroscopy. As well, the decline may be more gradual because novice trainees over their first few colonoscopies become more comfortable with several of the 'basic' colonoscopy manoeuvres (eg, movement of dials, torque steering), yet become aware of more advanced and challenging techniques, such as loop resolution and even possibly polypectomy.
This awareness of colonoscopy looping, and the need to resolve it, may explain the perceived drop in self-efficacy apparent in the last quartile of study colonoscopies. Learning curve theories demonstrate that learning curves, in general, do not proceed in a smooth, linear fashion, but are characterized by constant fluctuations (14). For colonoscopy training, it is quite possible that trainees spend their first training procedures acquiring the ‘basics’ of colonoscopy: movement of the dial, tip control and torque steering. During this time, they become increasingly adept and comfortable with these basic techniques, with a concomitant decrease in exertion and increase in self-efficacy. However, at some point (approximately 30 procedures in our study), trainees (with the advice of their trainers) become aware that more advanced techniques are required to navigate consistently and safely to the cecum.

Specifically, loop formation and the art of loop reduction became important. The majority of difficulties encountered during a colonoscopy result from lack of progression of the instrument on insertion, commonly as a consequence of looping of the colonoscope within the colon. To progress safely, effectively and with minimal patient discomfort, normally requires straightening of the instrument and resolution of the loop (6). The decreased self-efficacy scores achieved in this period may, in fact, represent loop reduction that will be interpreted by the inexperienced trainee as a shorter distance reached as less of the colonoscope is inserted into the patient when, in reality, they are in a better position to proceed.

There are several practical implications for endoscopy education that can be derived from the present study. First, the study validated a brief Endoscopy Task Load Index that can be completed in <1 min. The Endoscopy Task Load Index is easily applied and can be tracked over an entire training period, as well as applied to therapeutic interventions such as polypectomy, managing gastrointestinal bleeds and endoscopic retrograde cholangiopancreatography. It has the potential to identify periods of higher perceived exertion and facilitate appropriate levels of guidance from the trainers. Early in their training, novices will not be able to handle the additional workload imposed by questioning or verbose directions from the trainer. Verbal instruction of scope position and movement at this stage can be simplified by using a set of 12 direct, simple terms such as: tip up, tip down, tip left, tip right, clockwise torque, anticlockwise torque, insufflate, aspirate, advance/ push forward, withdraw/pull back, stop and slowly (15). As the trainee becomes more comfortable and exertion decreases, the amount of direct, hands-on supervision can be gradually decreased, thereby fostering more independence.

Work has been performed on developing formal curricula encompassing both physical and cognitive aspects of procedure teaching for colonoscopy (16,17). This is a concept that is at much further stages of evolution in the surgical realm (18-20). Much of the existing work in endoscopy has largely focused on the technical skills through simulators (21-25) and with the aid of other learning tools (4). The Endoscopy Task Load Index could be a valuable guide to the potential benefit (or harm) of new educational interventions.

While it is clear that skills acquisition improves with experience, debate remains as to the necessary number of procedures required for independent competence (26-28). A tool such as the Endoscopy Task Load Index, followed longitudinally, could provide a better definition of the transition point between the intermediate and the fully automated phases of expertise development. While our study clearly showed a difference between the trainees and the experts, we did not extend it sufficiently to help determine that exact point of transition. Such knowledge may also be helpful in assessing and following expert clinicians in need of more advanced skills training.

There were several important limitations to the present study. First, the sample size was small and the study did not extend beyond the first 50 procedures; hence, other important transition points on the road to colonoscopy competence (200 procedures by some groups [29]), such as polypectomy, cannot be ascertained. In addition, we did not evaluate the trainers’ hands-on involvement and degree of supervision, which naturally varies among educators and could have influenced perceived cognitive workload of the trainee. The study used group learning curves, which can yield a misleading picture of what is occurring in individual subjects (30). Furthermore, the study did not include objective parameters, such as cecal intubation, time and detection rates, nor was repeat assessment of practicing gastroenterologists/surgeons performed. Future studies will address the impact of magnetic endoscopic imaging (6) on both novice workload and colonoscopy performance.

Colonoscopy skills training and acquisition is in an exciting state of flux and development. With increasing emphasis of quality assurance measures in endoscopy, formalizing skills training to ensure competence of those providing the service are of paramount importance. The inclusion of objective tools that encompass both the physical and cognitive components of learning will be best suited to identify the optimal methods of teaching.

DISCLOSURES: This work originated from the University of Calgary, Canada, in affiliation with Gloucestershire Hospitals NHS Foundation (United Kingdom).

APPENDIX 1: MODIFIED NASA-TLX WORKLOAD ASSESSMENT TOOL

1. Mental Demand:
How much mental and perceptual activity was required to perform the procedure (eg, thinking, decision-making, looking, listening, remembering, reasoning, etc)? Was the procedure easy or demanding, simple or complex, exacting or forgiving?
0 Low 100 High

2. Physical Demand:
How much physical activity was required to perform the procedure (eg, pushing, pulling, turning, controlling, etc.)? Was the procedure trivial or demanding, easy or strenuous, restful or laborious?
0 Low 100 High

3. Time Demand:
How much time pressure did you feel due to the rate or pace of the procedure? Was the pace slow and leisurely, or rapid and frantic? Were there external time pressures (eg, personal, other duties) not directly related to the procedure?
0 Low 100 High

4. Effort:
How hard did you have to work mentally and physically to perform and learn during this procedure?
0 Low 100 High

5. Performance:
During this procedure, how successful do you think you were in accomplishing the goals set out by the trainer or yourself? How satisfied were you with your performance in accomplishing these goals? (NOTE: Good is on left hand side of scale, poor is on right)
0 Good 100 Poor

6. Frustration/Anxiety:
How discouraged, irritated, stressed and anxious (versus gratified, satisfied were you with your performance in accomplishing these goals? (NOTE: Good is on left hand side of scale, poor is on right)
0 Low 100 High
REFERENCES


• Reducing the signs and symptoms, inducing major clinical response and clinical remission, inhibiting the progression of structural damage and improving physical function in adult patients with moderately to severely active rheumatoid arthritis (RA). Can be used alone or in combination with methotrexate (MTX) or other disease-modifying antirheumatic drugs (DMARDs).

When used as first-line treatment in recently diagnosed patients who have not been previously treated with MTX, HUMIRA should be given in combination with MTX. Can be given as monotherapy in case of intolerance to MTX or when treatment with MTX is contraindicated.

• In combination with MTX, reducing the signs and symptoms of moderately to severely active polyarticular juvenile idiopathic arthritis (JIA) in patients 4 to 17 years of age who have had an inadequate response to one or more DMARDs. Can be used as monotherapy in case of intolerance to MTX or when continued treatment with MTX is not appropriate. HUMIRA has not been studied in children aged less than 4 years.

• Reducing the signs and symptoms in patients with active ankylosing spondylitis (AS) who have had an inadequate response to conventional therapy.

• Reducing the signs and symptoms of active arthritis and inhibiting the progression of structural damage and improving the physical function in adult psoriatic arthritis (PsA) patients. Can be used in combination with MTX in patients who do not respond adequately to MTX alone.

• Reducing the signs and symptoms and inducing clinical remission in adult patients with moderately to severely active Crohn’s disease (CD) who have had an inadequate response to conventional therapy, including corticosteroids and/or immunosuppressants. HUMIRA is indicated for reducing the signs and symptoms and inducing clinical remission in these patients if they have also lost response to or are intolerant to infliximab.

• Reducing the signs and symptoms and inducing and maintaining clinical remission in pediatric patients 13 to 17 years of age weighing ≥40 kg with severely active Crohn’s disease and/or who have had an inadequate response or were intolerant to conventional therapy (a corticosteroid and/or aminosalicylate and/or an immunosuppressant) and/or a tumour necrosis factor alpha antagonist.

• Treatment of adult patients with moderately to severely active ulcerative colitis (UC) who have had an inadequate response to conventional therapy including corticosteroids, azathioprine and/or 6-mercaptopurine (6-MP) or who are intolerant to such therapies. The efficacy of HUMIRA in patients who have lost response to, or were intolerant to, TNF blockers has not been established.

• Treatment of adult patients with chronic moderate to severe psoriasis (Ps) who are candidates for systemic therapy. For patients with chronic moderate plaque psoriasis, HUMIRA should be used after phototherapy has been shown to be ineffective or inappropriate.

Limited data are available for treatment with HUMIRA in children weighing <15 kg. Safety and effectiveness in pediatric patients in indications other than polyarticular JIA have not been established. Clinical trial data for patients aged 4 to 6 years are limited.

The safety and efficacy of HUMIRA were authorised in pediatric patients 13 to 17 years of age weighing ≥40 kg with severely active Crohn’s disease and/or who have had an inadequate response or were intolerant to conventional therapy.

Contraindications

• Severe infections such as sepsis, tuberculosis and opportunistic infections
• Moderate to severe heart failure (NYHA class III/IV)

Most serious warnings and precautions

Hepatosplenic T-Cell Lymphoma (HSTCL): Very rare post-marketing reports of HSTCL, a rare aggressive lymphoma that is often fatal, have been reported. Most of the patients had prior infliximab therapy as well as concomitant azathioprine or 6-mercaptopurine use for Crohn’s disease. The potential risk with the combination of azathioprine or 6-mercaptopurine and HUMIRA should be carefully considered.

Infections: Serious infections have been reported. Hospitalization or fatal outcomes associated with infections have been reported. Many of the serious infections have occurred in patients on concomitant immunosuppressive therapy that, in addition to their underlying disease, could predispose them to infections. Treatment with HUMIRA should not be initiated in patients with active infections. In patients who have been exposed to tuberculosis, and patients who have traveled in areas of high risk of tuberculosis or endemic mycoses, the risks and benefits of treatment with HUMIRA should be considered prior to initiating therapy.

As with other TNF blockers, patients should be monitored closely for infections (including tuberculosis) before, during and after treatment with HUMIRA. Administration of HUMIRA should be discontinued if a patient develops a serious infection or sepis, and appropriate therapy should be initiated. Physicians should exercise caution when considering the use of HUMIRA in patients with a history of recurrent infection or with underlying conditions which may predispose them to infections, or patients who have resided in regions where tuberculosis and histoplasmosis are endemic.

Pediatric Malignancy: Lymphoma and other malignancies, some fatal, have been reported in children and adolescent patients treated with TNF blockers, including HUMIRA.

Other relevant warnings and precautions

• Concurrent administration with other biologic DMARDs or other TNF antagonists not recommended
• Surgery: Close monitoring for infection required
• Patients with congestive heart failure: Cases of worsening congestive heart failure (CHF) and new onset CHF
• Hematologic events: Pancytopenia, including aplastic anemia, and medically significant cytopenia
• Hypersensitivity reactions, including anaphylaxis and latex allergic reactions
• Autoimmunity
• Immunizations: Live vaccines must be avoided. It is recommended that pediatric patients, if possible, be brought up to date with all immunizations in agreement with current immunization guidelines prior to initiating HUMIRA therapy
• Infections: Tuberculosis (TB) (including reactivation and new onset of TB), opportunistic infections (including invasive fungal infections), and hepatitis B virus reactivation
• Malignancies including malignancies in pediatric patients and young adults, lymphoma and non-lymphoma malignancy
• Neurological events: New onset or exacerbation of demyelinating disease
• Pregnant women: HUMIRA may cross the placenta; infants born to women treated with HUMIRA during pregnancy may be at increased risk for infection
• Nursing women: Breastfeeding is not recommended for at least five months after the last HUMIRA treatment
• Geriatrics: Higher incidence of infections and malignancies

For more information

Please consult the Product Monograph at http://webprod5.hc-sc.gc.ca/dpd-bdpp/index-eng.jsp for important information relating to adverse reactions, drug interactions, and dosing information which have not been discussed in this piece.

The Product Monograph is also available by calling at 1-888-704-8271.

References
