



Cross-Over Validation of an Endoscopic Training Box Task

Sanjay M. Salgado MD EdM*, Blake A. Niccum MD, Michael L. Kochman MD

Abstract

Background & Aims: Gastroenterology training programs have increasingly incorporated endoscopic simulation into their curricula to augment competency assessment and clinical training. One mechanical simulator model that has demonstrated significant promise is the Thompson Endoscopic Skills Trainer Box (TEST Box). When administering the TEST Box, we have noted that trainees differ in their approach to the tip deflection module, raising concerns pertaining to construct validity.

Methods: Five attending gastroenterologists with at least three years of independent practice participated in a randomized crossover study designed to provide validity data on the endoscopic tip-deflection task of the TEST Box, using two alternative approaches: the “grasping” method, in which a participant grasps the target object with a forceps in order to move the object into the intended location, and the “pass-through” method, in which a participant passes the closed forceps through the central hole in the object and subsequently opens the forceps in order to pick up and move the object into the intended location.

Results: The average scores of “grasping” and “pass-through” methods were 36.0 and 88.2, respectively. The use of the “grasping” method resulted in significantly lower scores compared to the “pass-through” method ($p = 0.027$; CI -94.5 to -9.9) when performing the tip deflection module of the TEST Box.

Conclusions: As we have historically observed that a cohort of trainee participants routinely utilize the “pass-through” method to complete the tip deflection module of the TEST Box during competency assessments, these observations reveal limitations in the validity of the tip deflection module of the TEST Box.

Keywords: Endoscopic simulation; TEST Box; Tip deflection; Forceps technique; Grasping; Pass-through

Introduction

Digestive diseases affect 60-70 million Americans resulting in more than 15 million endoscopic procedures annually in the United States (US) [1,2]. Trainees have been historically deemed competent at performing these procedures after crossing a minimum threshold of procedural volume [3]. Over time, however, studies revealed that the number of procedures required to achieve endoscopic competence varies widely by individual, spurring the development of novel measurement and assessment tools [4-6].

Modern gastroenterology (GI) training programs utilize a wide variety of instruments to assess endoscopic competency. A 2015 study found that most GI programs rely on procedural volume and subjective attending evaluations

Affiliation:

Division of Gastroenterology, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA

*Corresponding author:

Sanjay M. Salgado, Division of Gastroenterology, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA

Citation: Sanjay M. Salgado, Blake A. Niccum, Michael L. Kochman. Cross-Over Validation of an Endoscopic Training Box Task. *Journal of Surgery and Research*. 7 (2024): 191-195.

Received: March 01, 2024

Accepted: March 11, 2024

Published: May 08, 2024

for competency assessment [7]. A minority of programs track objective procedural metrics (i.e. cecal intubation rate) or utilize objective skills assessment tools, such as the American Society of Gastrointestinal Endoscopy (ASGE) Assessment of Competency in Endoscopy (ACE) Tool [7,8].

Over the past 10-15 years, GI training programs have also started to incorporate endoscopic simulation into their curricula to augment competency assessment and clinical training [7,9,10]. Simulator devices span a wide spectrum, including live animal models, *ex vivo* and *in vivo* animal models, mechanical models, and virtual reality computer simulators, each of which have inherent advantages and limitations [10]. Live animal models, for instance, tend to be realistic but limited by cost, infrastructure requirements, and ethical concerns [10]. On the other hand, mechanical models are typically more affordable, convenient, and standardized, but often fall short in realism and applicability [10].

One mechanical model that has demonstrated significant promise in the assessment of endoscopic competency is the Thompson Endoscopic Skills Trainer Box (TEST Box; licensed by EndoSim, Bolton, MA) [11]. Introduced in 2014, the TEST Box is a compact, low-cost mechanical endoscopic simulator that consists of a training box with five modules, each of which focuses on a particular endoscopic skill, including retroflexion, tip deflection, torque, polypectomy, and navigation/loop reduction [11]. In the original study describing this device, the authors evaluate the validity of the TEST Box as an assessment tool using the contemporary framework for validity [11,12]. This framework defines validity across several disciplines: test content (i.e. consistency between the content of a test and the construct it is intended to measure), internal structure (i.e. degree of inter-relationships between test components), response process (i.e. extent by which participants' thought processes and actions match those intended by the test administrators), relationship to other variables external to the test, and consequences of testing [12]. The authors of the study provide evidence that supports the validity of the TEST Box with respect to test content, internal structure, and response process. Follow up studies have revealed correlation of TEST Box scores with level of endoscopic experience and various endoscopic metrics [13-15], suggesting validity with respect to variables external to the test.

When administering the TEST Box at our institution, we have noted that participants routinely vary in their approach to the tip deflection module, raising concerns pertaining to construct validity. Despite receiving identical instructions preceding the module as per the Thompson Endoscopic Skills Trainer Instructions [16], some participants opt to move objects by grasping them with forceps (which we will term the "grasping" method), while others accomplish the same task by passing the forceps through the central hole in the object and then opening the forceps to prevent the object

from sliding off ("pass-through" method). Our preliminary analysis has suggested that the "pass-through" technique may lead to higher scores than the "grasping" strategy. In this randomized crossover study, we sought to evaluate whether participants using the "pass-through" method to complete the tip deflection module of TEST Box scored higher than participants using the "grasping" approach.

Methods

A randomized crossover study was designed to provide validity data on the endoscopic tip-deflection task of the TEST Box. University of Pennsylvania IRB approval was obtained in June of 2021. Inclusion criteria included attending general gastroenterologists with at least three years of independent practice and without prior exposure to the endoscopic TEST Box. We excluded trainees, and any attendings with advanced endoscopic training.

Participants were randomized based on random numbers generated using computer software (Excel 2016; Microsoft Corp., Redmond, WA) to perform the tip deflection TEST Box task via a "grasping" or "pass-through" method. All participants were read the instructions for the Tip-Deflection task as described in the "Score Sheet and Instructions" guide from the ENDOSIM website [16]. For the "grasping" method, participants were additionally instructed to grasp the object with an FG-47L-1: Olympus Rat Tooth Grasping Forceps in order to move the object into the intended location (see Figure 1). For the "pass-through" method, participants were instructed to pass the closed forceps through the central hole in the object and subsequently open the forceps in order to pick up and move the object into the intended location. After completion of this task, participants were given a 30 second break before performing the same task using the alternative method. For all trials, an FG-47L-1: Olympus Rat Tooth Grasping Forceps was used.

Scoring was based on the "Score Sheet and Instructions" guide from the ENDOSIM website. Each object transferred into the correct compartment was awarded 10 points. If the participant completed the task under 5 minutes, they were awarded 1 additional point for each second remaining on the 5-minute timer. Instructions were provided by the same physician in all trials. Scoring was conducted by the same physician in all trials, in addition to one of three research assistants. Total scores were compared after each trial to ensure that there was not a scoring discrepancy.

We hypothesized that the "pass-through" method would be superior to the "grasping" method for obtaining a higher score on the tip deflection task of the TEST Box. Based on analysis of the performance of senior fellows, we estimated that the mean score for the "grasping" and "pass-through" methods would be 35 and 80, respectively, with a standard deviation of 25. We calculated a necessary sample size of 5 participants to achieve 80% power at a significance level of

.05. A post-hoc power calculation was conducted after results were obtained, which confirmed that the initial sample size was sufficient.

Results

Five attending general gastroenterologists (three women, two men) participated during April 2022. No scoring discrepancies were identified between the scorers in the recorded trials. The mean task scores were analyzed using a paired t-test. All statistics are reported as mean ± standard error of the mean. Correlations are reported as Pearson’s correlation coefficients. All statistics were performed by using Stata 12.0 (StataCorp, College Station, Texas).

The average scores of “grasping” and “pass-through” methods were 36.0 and 88.2, respectively (see Table 1). A paired t-test indicated that use of the “grasping” method resulted in lower scores compared to the “pass-through” method ($p = 0.027$; CI -94.5 to -9.9). A 95% confidence interval for the true difference in population means resulted in the interval of (-3.466, -.034). Two participants were able to fully complete the task using the “pass-through” method within the allotted five minutes. No participant was able to fully complete the task using the “grasping” method in the allotted time.

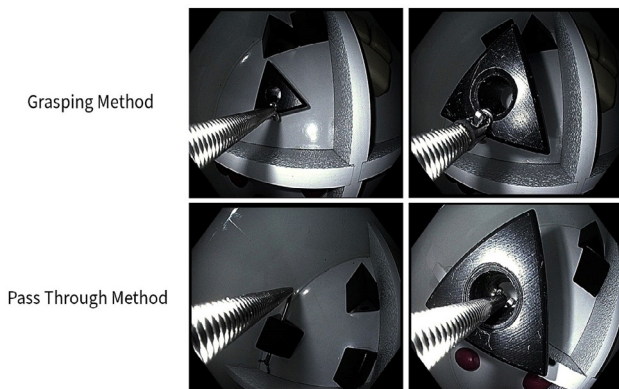


Figure 1: For the “grasping” method (above), participants were instructed to grasp the object with an FG-47L-1: Olympus Rat Tooth Grasping Forceps in order to move the object into the intended location. For the “pass-through” method (below), participants were instructed to pass the closed forceps through the central hole in the object and subsequently open the forceps in order to pick up and move the object into the intended location

Discussion

In an effort to fortify assessment of endoscopic competency at the University of Pennsylvania, all general GI fellows began completing the TEST Box on an annual basis starting in 2020. We have had experience with its use and application in our trainees since 2017. When their performances were observed, it was noted that participants routinely varied in their approach to the tip deflection module despite receiving identical instructions matching those displayed on the ENDOSIM webpage [16], with some utilizing the “grasping” method and others employing the “pass-through” technique. The observed differences in the performance of this task based on the chosen completion method raised concerns of the construct validity of the module. To formally evaluate the relationship between these two strategies and module score, we performed a randomized crossover study in which general attending gastroenterologists without prior exposure to TEST Box were required to complete the tip deflection module using specifically the “grasping” or “pass-through” method, followed shortly thereafter by completion of the module using the alternative approach in a randomized cross-over design. In this study, we found that participants scored significantly higher on the tip deflection module when using the “pass-through” method compared to the “grasping” technique. To place the score difference observed in this study in a greater context, the mean difference in the scores by using the “pass-through” method in the tip deflection task in this singular task (52 points), would account for two-thirds of the observed overall score difference between first and second year gastroenterology fellows in all five tasks, as reported by the prior validation studies [13].

Prior studies have provided evidence broadly supporting the validity of the TEST Box as an assessment tool based on the contemporary framework for validity. In the 2014 study introducing the TEST Box, the authors support validity of test content by demonstrating that eight surveyed experts collectively felt that the modules comprising the TEST Box were realistic (content validity index (CVI) 0.88), relevant (CVI 1.00), and representative (CVI 0.88), and by showcasing that the majority of 54 surveyed participants with variable endoscopic experience felt that the TEST Box could differentiate between levels of endoscopic experience

Table 1: Participant Scores by Testing Method

Participant	Pass-Through Method	Grasping Method	Score Difference	P Value
1	80	30	50	
2	60	40	20	
3	113	60	53	
4	128	20	108	
5	60	30	30	
Median Score	80	30	50	
Mean Score + SEM	88.2 ± 13.9	36.0 ± 6.8	52.2	0.027

(82%) and should be used as a practice tool prior to clinical cases (93%) [11]. Likewise, they support validity of internal structure by revealing that inter-module correlation ranged from 0.67 to 0.93, that each module contributed between 16.0% and 26.1% of total score, and that completion of the TEST Box two consecutive times by the same participant with the same proctor resulted in similar scores [11]. In their original study, the creators of the TEST Box also argue that the TEST Box is valid with respect to response process by citing that the same printed instructions were used by proctors in all cases and that TEST Box score did not vary by proctor (297.6 using Proctor 1 vs. 308.1 using Proctor 2, $p=0.94$) [11]. Follow up studies have since provided data supporting the validity of the TEST Box in relationship to other variables by demonstrating correlation between TEST Box score and both level of endoscopic experience and endoscopic measures of performance as evaluated based on the ASGE ACE Tool [13-15].

Despite this constellation of findings supportive of the validity of the TEST Box, our results bring to light apparent weaknesses in validity that should be addressed. The current scoring rubric does not specify how the participants can manipulate and move the target objects. Some participants complete the tip deflection module using the “grasping” technique and others employ the “pass-through” strategy. We argue that the content and construct validity is decreased due to a reduction in both relevance (as the “pass-through” method is not a traditional endoscopic technique) and representativeness (as variation in scores between participants could reflect forceps technique rather than proficiency at tip deflection). Additionally, we propose that variation in forceps strategy reduces response process validity, as the actions of participants using the “pass-through” method very likely differ from those intended by the test creators. Moreover, as utilizing the “pass-through” technique may lead to a higher score on the tip deflection module than would be expected based on a user’s endoscopic skill, employing this non-traditional forceps approach could limit validity with respect to internal structure (by altering the inter-relationships between an individual’s module scores) and relationship of TEST Box score to other variables external to the test (i.e., assessment based on the ASGE ACE Tool). As such, we believe that the instructions for the tip deflection module should be modified to explicitly state that participants must move objects using the “grasping” method, and that the “pass-through” technique is not permitted.

Of note, these validity limitations may not have surfaced in the 2014 study introducing the TEST Box as all participants in the original study may have completed the tip deflection module using the “grasping” method. In support of this view, surveyed experts in that study rated the tip deflection module equivalent to the other four modules with respect to realism

(CVI 0.88), relevance (CVI 1.00), and representativeness (CVI 0.88) [11], which we believe would likely not have occurred if some participants had utilized the “pass-through” technique, as this method of using forceps is not traditionally applicable to clinical cases. Additionally, the standard error (18.1) of the tip deflection scores in that study was similar in magnitude to that of the other four modules (ranging 13.3-19.4). In light of our findings, we believe that the standard error of this module would likely have been higher than that of the other modules if some participants had used the “grasping” method while others had employed the “pass-through” approach.

Our study has a number of strengths. All participants had a similar level of clinical experience, and none had previously been exposed to the TEST Box. Additionally, the crossover study design reduced risk of confounding and accommodated a small sample size. Moreover, scores for the tip deflection module were tabulated using the scoring system from the ENDOSIM website [16], which match those used in the prior studies involving the TEST Box [11,13-15], maximizing the external validity of our findings.

We also acknowledge several limitations. The crossover format of our study introduces risk that participant performance on the second attempt was influenced by the first experience navigating the module. However, we are hopeful that the impact of this potential confounding factor is low as prior study has shown that two immediately consecutive completions of the TEST Box result in similar scores [11], as participants were only allowed to take a very short break (30 seconds) between attempts (mitigating risk of interval change in endoscopic skill), and as randomization was used to determine which forceps technique each participant utilized first. Additionally, the equivalent level of all participants in our study prevented analysis of how our findings translate to endoscopists with different degrees of experience. Our study is also small, albeit sufficiently powered. Finally, as participants solely completed the tip deflection module, we were unable to analyze the impact of using different forceps techniques during the tip deflection module on the internal structure validity of the TEST Box as a whole.

Conclusion

In summary, we conducted a randomized crossover study in which we found that completion of the tip deflection module of the TEST Box using a “pass-through” forceps technique is associated with better performance than using the traditional “grasping” approach. As we have historically observed that a cohort of participants routinely utilize the “pass-through” method to complete the tip deflection module of the TEST Box during competency assessments, we believe that our observations reveal limitations in the validity of the tip deflection module, in particular pertaining to test content and

response process, as well as potentially internal structure and relationship to other variables. Going forward, we believe that the instructions for the tip deflection module of the TEST Box should be modified to ensure that all participants are required to use a specified forceps technique in order to increase the validity of this promising modality of assessing endoscopic competence. We would recommend the “grasping” method as the specified technique, as we believe this more greatly translates as a skill utilized in intraluminal endoscopy. To better understand the importance of our findings, further study is required to determine how frequently participants elect to utilize the “pass-through” technique as opposed to the traditional “grasping” approach when navigating the tip deflection module using the current instructions.

Conflicts of Interest

Sanjay M. Salgado has no conflicts of interest or financial ties to disclose.

Blake A. Niccum has no conflicts of interest or financial ties to disclose.

Michael L. Kochman has received consulting fees from Olympus, Virgo systems, Boston Scientific, Medtronic, Castle Pharmaceuticals, Dark Canyon Labs, and has an equity interest in Dark Canyon Labs, Virgo Systems, EndoSound.

References

1. U.S. Department of Health and Human Services NIOH. Opportunities and Challenges in Digestive Diseases Research: Recommendations of the National Commission on Digestive Diseases (2009).
2. Peery AF, Crockett SD, Murphy CC, et al. Burden and Cost of Gastrointestinal, Liver, and Pancreatic Diseases in the United States: Update 2018. *Gastroenterology* 156 (2019): 254-272.
3. Eisen GM, Baron TH, Dominitz JA, et al. Methods of granting hospital privileges to perform gastrointestinal endoscopy. *Gastrointest Endosc* 55 (2002): 780-783.
4. Spier BJ, Benson M, Pfau PR, et al. Colonoscopy training in gastroenterology fellowships: determining competence. *Gastrointest Endosc* 71 (2010): 319-324.
5. Sedlack RE. Training to competency in colonoscopy: assessing and defining competency standards. *Gastrointest Endosc* 74 (2011): 355-366.
6. Sedlack RE. The Mayo Colonoscopy Skills Assessment Tool: validation of a unique instrument to assess colonoscopy skills in trainees. *Gastrointest Endosc* 72 (2010): 1125-1133.
7. Patel SG, Keswani R, Elta G, et al. Status of Competency-Based Medical Education in Endoscopy Training: A Nationwide Survey of US ACGME-Accredited Gastroenterology Training Programs. *Am J Gastroenterol* 110 (2015): 956-962.
8. Sedlack RE, Coyle WJ, Obstein KL, et al. ASGE's assessment of competency in endoscopy evaluation tools for colonoscopy and EGD. *Gastrointest Endosc* 79 (2014): 1-7.
9. Jirapinyo P, Thompson CC. Current status of endoscopic simulation in gastroenterology fellowship training programs. *Surg Endosc* 29 (2015): 1913-1919.
10. Goodman AJ, Melson J, Aslanian HR, et al. Endoscopic simulators. *Gastrointest Endosc* 90 (2019): 1-12.
11. Thompson CC, Jirapinyo P, Kumar N, et al. Development and initial validation of an endoscopic part-task training box. *Endoscopy* 46 (2014): 735-744.
12. Education AERAAPANCoMi. Standards for Educational and Psychological Testing. In: Washington, DC: American Educational Research Association (2014).
13. Jirapinyo P, Kumar N, Thompson CC. Validation of an endoscopic part-task training box as a skill assessment tool. *Gastrointest Endosc* 81 (2015): 967-973.
14. Ou A, Shin CM, Goodman AJ, Poles MA, Popov VB. Endoscopic part-task training box scores correlate with endoscopic outcomes. *Surg Endosc* 35 (2021): 3592-3599.
15. Tamai N, Aihara H, Kato M, et al. Competency assessment for gastric endoscopic submucosal dissection using an endoscopic part-task training box. *Surg Endosc* 33 (2019): 2548-2552.
16. ENDOSIM Evidence-Based Simulation. Thompson Endoscopic Skills Trainer Instructions Web site (2022).