

Construction and validation of a surgical skills assessment tool for general surgery residency program

Construção e validação de um instrumento de avaliação de habilidades técnicas para programas de residência em cirurgia geral

ELIZABETH GOMES DOS SANTOS, TCBC-RJ¹; GIL FERNANDO DA COSTA MENDES DE SALLES¹

A B S T R A C T

Objective: To develop and validate an instrument for measuring the acquisition of technical skills in conducting operations of increasing difficulty for use in General Surgery Residency (GSR) programs. **Methods:** we built a surgical skills assessment tool containing 11 operations in increasing levels of difficulty. For instrument validation we used the face validity method. Through an electronic survey tool (Survey MonKey®) we sent a questionnaire to Full and Emeritus members of the Brazilian College of Surgeons – CBC – all bearers of the CBC Specialist Title. **Results:** Of the 307 questionnaires sent we received 100 responses. For the analysis of the data collected we used the Cronbach's alpha test. We observed that, in general, the overall alpha presented with values near or greater than 0.70, meaning good consistency to assess their points of interest. **Conclusion:** The evaluation instrument built was validated and can be used as a method of assessment of technical skill acquisition in the General Surgery Residency programs in Brazil.

Key words: Medical Residency. General surgery. Medical Education. Training programs.

INTRODUCTION

A good medical General Surgery Residency (GSR) program is one that forms skilled surgeons capable of providing the best possible medical care to the population, wherever they operate, in big cities or outside them. For that training to be appropriate and the results satisfactory, there needs to be a structured and balanced curriculum, with objectives defined at different stages of specialized training, which should include an extensive breakdown of theoretical knowledge and an equal opportunity for all residents to develop the technical skills essential to the surgeon¹, since "failure is not an option"².

Assessments should occur at the end of each period or year of training so that one knows if the resident can proceed to a more advanced training stage. At the end of the residency, a judicious assessment of technical skills should be mandatory so that certification can be granted to a competent professional, able to offer patients excellence in care and to independently exercise his/her duties³, pursuant to Specialist Title⁴ that will be granted by law at the end of the program in Brazil.

In 1872, in Germany, Theodore Billroth had already recognized the need of some form of more complete assessment to the general surgeon beyond the cognitive dimension, since even a great theoretical

knowledge is not warranty of technical skill⁵. Much has been published about evaluation. In the view of Satava et al., there is a deficiency in the field of evaluation of surgeons' performance and technical skill⁶. To Bhatti, there is no reliable and objective single test able to assess all the details involved in the competence of a surgeon³.

One can interpret the evaluation in various forms, and all its methods have strengths and weaknesses. The word comes from Latin *a + valere* and means to assign a value to the studied object. It implies to judge, to appraise. When it comes to students, residents in the case, assessment is almost synonymous with verification or measurement⁷. In one of his many ideas, to evaluate is to compare and to classify according to criteria defined by an evaluation instrument. When evaluating technical skills it is necessary to have a set standard⁸ (criterion-referenced assessment), whose outcome may be favorable or not to the performance of who is being evaluated⁹. The importance of the assessment certainly does not end in the result obtained; but when it comes to obtaining the Specialist Title, to meet the criteria for "certification" is the ultimate goal.

Most of the evaluation tests used today purports to the assessment of knowledge and for that, among others, multiple-choice tests are very good. When done carefully, with elaborate options, they are able to assess theoretical knowledge, but they do not evaluate the physician's ability

1. Faculdade de Medicina, Universidade Federal do Rio de Janeiro, RJ, Brasil.

to examine a patient (clinical competence) or perform an operation (surgical skills). Alone, they do not allow attesting the performance of a surgeon¹⁰.

So far there is no single test that includes the evaluation of all the requirements needed to test technical skills, allowing to conclude whether the surgeon is competent or not. Surgical performance can be assessed through direct observation in the operating room by means of a simulator, through videos, by OSATS (Objective Structured Assessment of Technical Skill)⁸, and even oral tests, by asking the resident to describe an operation¹¹. As positive points we can say that they are safe, have credibility and validity, but few are directed to specific aspects of the procedures, some are unaffordable for some hospitals, others require a long time to apply, and require a very large apparatus, such as video recordings, multiple testing stations or simulators.

Technical skill is the psychomotor domain of knowledge, is the "set of acquired practices, especially by demonstration, repetition and critical reworking, which provide the surgeons with expertise know-how along with the ability to make decisions and solve issues in their field", and includes accuracy and error prevention. These capabilities are very difficult to be measured¹². Ericsson et al. support the hypothesis that the specialist is developed by an intense and efficient training program¹³. The psychomotor domain development is achieved through movement repetition. This means that the greater the number of times the resident performs a given operation, the more he/she will develop manual dexterity, until automation is reached, when movements are performed without thinking about them.

Only in the last decade there has been a greater emphasis on evaluation of manual dexterity, motivated perhaps by the perception of educators that "innate talent", the quality that makes the resident need a lot fewer repetitions to achieve excellence, is a rare genetic peculiarity, and for those who lack it, rigorous training is key, with movements repetition in a greater number of times until the future surgeon can be considered competent, which is achieved with a large volume of operations¹⁴.

Bhatti and Cummings³, citing Miller, propose that the acquisition of surgical competence is based on four stages. For this demonstration, they used a pyramid with a wider base. At the bottom is the "knowing" that corresponds to knowledge (cognition), which is necessary to reach the next stage. In the fourth and final stage is what the individual "actually does", and is where the surgeon effectively demonstrates his/her performance. In surgery, it means practical performance, knowing how to perform the operation³.

In the regulation of GSR programs by the National Commission of Medical Residency (Brazil) there is no standard tool for technical skill assessment, either during or at the end of the programs⁴. In a thus formatted system, it is assumed that at the end of the program the resident

will be able to receive the Specialist Title, even if one is not aware of how many and what kind of operations the resident performed during the training period. There is no performance assessment that enables the identification of strengths and weaknesses of the training process so that the program can be improved, or which ratifies the decision not to certify the resident if he/she does not meet the predetermined standards.

Therefore, the aims of this study were to build a tool to assess the technical ability to GSR programs residents in conducting operations of increasing difficulty, and validate the instrument with General Surgery specialists, Full and Emeritus Members of the Brazilian College of Surgeons (CBC).

METHODS

Based on the most prevalent diseases in the context of daily General Surgery exercise, present in most hospitals that house GSR programs across the country, whether in universities or not, we developed an instrument including 11 operations at increasing levels of difficulty (Table 1), considered by the CBC General Surgery specialists as indispensable for those who will receive the General Surgery Specialist Title, and a questionnaire with four questions for each operation to assess the technical ability when performing them (Table 2). Among the eleven procedures proposed, three were considered of little difficulty, six of medium difficulty and two, great difficulty.

Validation consisted of two parts, the first related to the operations' degree of difficulty (Table 1), and the second, specific, considering particularities of each operation (Table 2). The questionnaire was constructed with closed answer questions, the options being: agree, when the experts were in agreement with the scale of the operations difficulty degrees and with the ability of the criteria to demonstrate skill when performing the operation; and disagree, when they thought otherwise.

We used face validation method, one in which a group of people connected to the area of interest of what is being validated judges whether that instrument is able to evaluate what it proposes. Through an electronic research tool (Survey MonKey®), we sent a questionnaire to Full and Emeritus members of CBC of all Brazilian states, all bearers of the CBC Specialist Title. The number of experts was based on a simple random sample, with sampling error not exceeding 5% according to proportional stratified sampling, where N = 1329 represented the total of Surgeons registered with the CBC as general surgeons and n = 307 showing the minimum number of surgeons to be consulted. We sent the emails in three distinct stages. Between the first and second delivery there was an interval of 30 days, and between the second and the third, the interval was 15 days.

For the analysis of the data collected, we used Cronbach's alpha to determine the internal consistency of responses or the average correlation of items in a survey instrument, to assess its reliability. We performed the analysis with the complete number of observations for each operation. For some procedures, we could not calculate the coefficient due to the almost absolute occurrence of concordant answers for the relevant questions (almost 100% agreement).

This work was presented to the Ethics in Research Committee at the Hospital Universitário Clementino Fraga Filho - UFRJ under the number 115,032 and was approved on October 4th, 2012.

RESULTS

Of the 307 questionnaires sent, we received 100 responses. We considered usable the responses that followed the alternatives: "agree" or "disagree". In some cases, it was necessary to eliminate the answer to a question because it did not meet the established criteria.

We observed that for almost all operations the global alpha coefficients presented with values near or greater than 0.70, a good expression of internal consistency to assess the matters of interest (Table 3). Only the procedure "colostomy" had an alpha below the 0.70 average, 0.62. We then withdrew the item considered inconsistent (Table 2), this operation retaining three assessment items, and the alpha coefficient increased to 0.69. Therefore, the test showed that all the operations-related questions were consistent and can be applied with reliability.

DISCUSSION

The changes taking place in general surgery practice, new technologies that require greater skills for

the domain, the social demands on patients safety and the legal prohibitions on animal training have required new behaviors when training and evaluating General Surgery residents.

In GSR the lack of an assessment tool of technical skill development can sometimes compromise the outcome of residents performance evaluations, since in most cases these refer only to knowledge (cognitive domain) and behavior (affective domain)³. The use of indicators such as hospital stay, postoperative complications, mortality, presence of foreign bodies left in patients, iatrogenic injuries and dehiscence are part of some competency assessment protocols, but are not suitable for technical skill assessment *per se*^{12,13}.

Building an instrument to assess gain in technical skills is not an easy task, and passes through several stages. Everyone involved in the education of future general surgeons may question what operations are essential to know how to perform, and in what number, for the resident to be considered a specialist.

In this work, the operations were arbitrarily chosen to be part of everyday life for all general surgeons practicing in urban centers or outside of them, and were judged by other specialists (members of the Brazilian College of Surgeons) as representing increasing levels of execution difficulty, and indispensable for those who receive a Specialist Title. The evaluation criteria for each of the procedures were also considered important, meaning the resident who performs them properly can be considered competent in carrying out that specific operation.

The most common form of data collection is the use of questionnaires. With the advent of computers, the electronic scientific research has increased enough to be a facilitating factor¹⁵. It offers many advantages, low cost, eliminates the embarrassment of a personal interview and favors a large sample. As disadvantages, there are the lack of understanding of those who read the questionnaire, the difficulties of internet access, and the main, the low response rate, as stated by several authors¹⁶⁻¹⁸. The literature states that these rates can vary between 13 and 35%¹⁵⁻¹⁸. In this work, the index was 32.6%, which is within the acceptable for this type of research.

To validate is to deem correct, true. There are different validation levels: face validity, content validity and criterion-related validity¹⁹⁻²¹. The face one was chosen, which, being simpler, is usually the first validation method to be used in a study²¹. Validation is a general measure, and although it is considered by many as weak evidence, does not make it incorrect. It is the kind in which a test is evaluated by a group of selected people involved in that study field, who state that that test is able to assess what it has set to assess. In other words, a test has face validity when it "seems" to measure what it is intended to measure. It can be used alone or as a first step to validation of an instrument, and researchers assume that the results are representative of reality. In our case, it means that the list

Table 1 - Operations proposed for evaluation, according to the degree of difficulty.

Degree Difficult	Operation
Small	Inguinal Herniorrhaphy Gastrostomy
Medium	Incisional Herniorrhaphy Colostomy Enterectomy Laparoscopic Cholecystectomy Biliodigestive derivation Partial gastrectomy Partial colectomy
Great	Simple Hepatectomy Distal Pancreatectomy

Table 2 - Specific content of the technical skills assessment tool.**Validation of specific criteria of performance evaluation for each procedure****Inguinal Herniorrhaphy**

- a. Correctly identifies the structures corresponding to the inguinal ligament, conjoint tendon, inguinal triangle?
- b. Properly identifies the anatomical relations with the femoral vessels?
- c. Properly handles the hernia sac?
- d. Properly Fixes the mesh?

Gastrostomy

- a. Properly incises the abdominal wall for Gastrostomy?
- b. Chooses a suitable gastric segment for positioning of the catheter?
- c. Properly fixes the catheter to the stomach?
- d. Properly fixes the catheter to the abdominal wall?

Incisional Herniorrhaphy

- a. Properly prepares the aponeurosis for fixing the mesh?
- b. Properly fixes the mesh? (adequate distance from the aponeurotic edges, adequate distance between stitches)
- c. Properly handles the contents of the hernia sac?
- d. Properly handles the hernia sac?(resection of excess, base ligation, possibility of use of the excess for abdominal cavity insulation)

Colostomy

- a. Chooses the correct abdominal wall spot for colostomy positioning?
- b. Exercises caution not to let the colon in tension?
- c. Incises the colon in proper extension?
- d. Adequately performs the lateral colostomy? (early maturation)*

Enterectomy

- a. Properly uses linear staplers for suture of small bowel resection?
- b. Properly uses intestinal clamps?
- c. Properly performs manual entero-enteric anastomosis?
- d. Observes the care required for the success of the anastomosis. (angles' stitches, vascularization, tension on suture line)

Laparoscopic Cholecystectomy

- a. Performs the first abdominal puncture with the due care?
- b. Recognizes the anatomical repairs for surgical safety? (dissection along the gallbladder, Calot's Triangle exposure, exposure of junction of cystic duct and choledocus)
- c. Properly dissects the gallbladder pedicle (cystic duct, choledocus, cystic artery)
- d. Uses cautery with the required caution? (low current, away from the main biliary tree area)

Biliodigestive derivation with jejunum

- a. Appropriately chooses the bowel segment for anastomosis? (proper distance to excluded loop)
- b. Properly incises biliary tree and bowel loop?
- c. Properly performs the anastomosis? (no tension)
- d. Appropriately positions the cavity drain?

Partial Gastrectomy

- a. Recognizes stomach's vascular pedicles? (left gastric and celiac trunk; right gastric, left and right gastroepiploic, short vessels)
- b. Recognize stomach's anatomical divisions?
- c. Adequately performs vascular ligatures? (type of suture, type of ligature)
- d. Adequately performs the reconstruction of transit? (Roux-en-Y, B I or B II, submucosal hemostasis, equal distance from the edges).

Partial Colectomy

- a. Recognizes anatomical landmarks and vascular pedicle of segment to be resected.
- b. Adequately performs vascular ligatures?
- c. Properly uses linear and circular staplers?
- d. Adequately performs anastomosis for the reconstruction of the transit? (repairs in the angles, extramucosal suture, mesenterium position, absence of tension)

Simple Hepatectomy

- Correctly recognizes the liver vascularization landmarks?
- Adequately performs digitoclasis, ligatures and resection?
- Recognizes the intrahepatic ducts and ligates them separately?
- Adequately uses the different methods for hemostasis? (monopolar current, ultrasonic, argon scalpel, biological glues, etc.)

Distal Pancreatectomy

- Properly recognizes and dissects planes for access to the pancreas?
- Properly identifies splenic artery and vein?
- Adequately performs parenchymal suture?
- Adequately treats the pancreatic duct?

* Item withdrawn from the questionnaire after validation by experts.

Table 3 - Cronbach's alpha coefficients for each of the 11 operations evaluated.

Operation	Number of responses	Cronbach's alpha coefficient
Inguinal Herniorrhaphy	99	0.69
Gastrostomy	98	NP
Incisional Herniorrhaphy	97	0.85
Colostomy	97	* 0.62
Enterectomy	98	0.89
Laparoscopic Cholecystectomy	97	0.76
Biliodigestive derivation	94	NP
Partial gastrectomy	96	0.67
Partial colectomy	97	NP
Simple Hepatectomy	96	0.93
Distal Pancreatectomy	98	0.98

* After exclusion of the fourth evaluation item of this operation, the alpha coefficient went on to 0.69.

NP: not processed due to absolute agreement of answers.

of chosen operations is representative of the increasing levels of complexity, and that the questions on them may indicate the technical skill of the evaluated resident when performing them¹⁹⁻²¹.

The use of this questionnaire was considered practical and objective by addressing specific aspects of the procedures. Another advantage is that there was no high cost involved in its application.

In conclusion, to certify a Specialist in General Surgery is a big responsibility. We believe that improving evaluation methods translates commitment to improving

the care provided and to provide the population with the quality care it deserves. The built assessment tool was validated and can be used as a method of evaluation of technical skill gain in the Medical Residency Program in General Surgery in Brazil.

Acknowledgments

Our thanks to the Brazilian College of Surgeons for providing the National list of Full and Emeritus Members, and to the Specialists who responded to the questionnaire.

R E S U M O

Objetivo: construir e validar um instrumento para aferir a aquisição de habilidades técnicas na realização de operações de graus crescentes de dificuldade para ser utilizado na Residência Médica em Cirurgia Geral (RMCG). **Métodos:** foi construído um instrumento de avaliação de habilidades cirúrgicas contendo 11 operações em níveis crescentes de dificuldade. Para a validação do instrumento foi usado o método de validação de face. Por meio de uma ferramenta de pesquisa eletrônica (Survey MonKey®) um questionário foi enviado para membros Titulares e Eméritos do CBC de todos os estados brasileiros, portadores de Título de Especialista pelo CBC. **Resultados:** Dos 307 questionários enviados foram recebidas 100 respostas. Para a análise dos dados coletados foi utilizado o teste alfa de Cronbach. Observou-se, de uma forma geral, que os alfas globais se apresentaram com valores próximos ou superiores a 0,70, expressando uma boa consistência interna das perguntas para avaliar os respectivos aspectos de interesse. **Conclusão:** O instrumento de avaliação construído foi validado e pode ser usado como um método de avaliação da aquisição de habilidade técnica na Residência Médica em Cirurgia Geral no Brasil.

Descritores: Residência Médica. Cirurgia Geral. Educação Médica. Programas de Treinamento.

REFERENCES

- Grantcharov TP, Bardram L, Funch-Jensen P, Rosemberg J. Assessment of technical surgical skills. *Eur J Surg.* 2002;168(3):139-44.
- Panait L, Larios JM, Brenes RA, Fancher TT, Ajemian MS, Dudrick SJ, et al. Surgical skills assessment of applicants to general surgery residency. *J Surg Res.* 2011;170(2):189-94.
- Bhatti NI, Cummings CW. Competency in surgical residency training: defining and raising the bar. *Acad Med.* 2007;82(6):569-73.
- Brasil. Ministério da Educação [homepage na Internet]. Regulação da Residência Médica [acesso em jul 2014]. Disponível em <http://www.mec.gov.br>.
- O'Shea JS. Becoming a surgeon in the early 20th century: parallels to the present. *J Surg Edu.* 2008;65(3):236-41.
- Satava RM, Gallagher AG, Pellegrini CA. Surgical competence and surgical proficiency: definitions, taxonomy and metrics. *J Am Coll Surg.* 2003;196(6):933-7.
- Luckesi CC. Avaliação da aprendizagem escolar: estudos e proposições. 19ª ed. São Paulo: Cortez; 2008.
- Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg.* 1997;84(2):273-8.
- Fried GM, Feldman LS. Objective assessment of technical performance. *World J Surg.* 2008;32(2):156-60.
- Cheung MT, Yau KK. Objective assessment of a surgical trainee. *ANZ J Surg.* 2002;72(5):325-30.
- Faurie C, Khadra M. Technical competence in surgeons. *ANZ J Surg.* 2012;82(10):682-90.
- Mendes OM. Avaliação formativa no ensino superior: reflexões e alternativas possíveis [acesso em jul 2013]. Disponível em: http://www.proiac.ufr.br/sites/default/files/avaliacao_formativa_olenir.pdf
- Ericsson KA, Nandagopal K, Roring RW. Toward a science of exceptional achievement: attaining superior performance through deliberate practice. *Ann N Y Acad Sci.* 2009;1172:199-217.
- Sooriakumaran P. Is UK surgical training in crisis? A trainee's perspective. *Int J Surg.* 2004;2(2):127; discussion 128-9.
- Coleman JJ, Esposito TJ, Rozycki GS, Feliciano DV. Early subspecialization and perceived competence in surgical training: are residents ready? *J Am Coll Surg.* 2013;216(4):764-71; discussion 771-3.
- Barbosa EF. Instrumentos de coleta de dados em pesquisas educacionais [acesso em jul 2014]. Disponível em: http://www.tecnologiadeprojetos.com.br/banco_objetos/%7B363E5BFD-17F5-433A-91A0-2F91727168E3%7D_instrumentos%20de%20coleta.pdf
- Vieira HC, Castro AE, Schuch Júnior VF. O uso de questionários via e-mails em pesquisas acadêmicas sob a ótica dos respondentes. XIII SEMEAS Seminários em administração. 2010. [acesso em mai 2012]. Disponível em: <http://www.ead.fea.usp.br/semead/13semead/resultado/trabalhosPDF/612.pdf>.
- Vasconcelos L, Guedes LFA. E-surveys: vantagens e limitações dos questionários eletrônicos via internet no contexto da pesquisa científica. X SEMEAS, FEA-USP. 2007. [Acesso em jul 2014]. Disponível em: http://www.ead.fea.usp.br/Semead/10semead/sistema/resultado/an_resumo.asp?cod_trabalho=420.
- Bright E, Vine S, Wilson MR, Masters RS, McGrath JS. Face validity, construct validity and training benefits of a virtual reality TURP simulator. *Int J Surg.* 2012;10(3):163-6.
- Shuttleworth M. Face validity. [acesso em set 2014]. Disponível em: <https://www.explorable.com/face-validity>.
- Trochim WMK. Measurement validity types. [acesso em jul 2014]. Disponível em: <http://www.socialresearchmethods.net/kb/measval.php>.

Received at: 30/05/2015

Accepted for publication: 25/07/2015

Conflict of interest: none.

Source of funding: none.

Mailing address:

Elizabeth Gomes dos Santos

E-mail: eligsant54@gmail.com