Validation of the Imperial College Surgical Assessment Device (ICSAD) for labour epidural placement

Validation du dispositif d'évaluation chirurgicale du Imperial College (ICSAD) pour le positionnement de la péridurale pour le travail obstétrical

Megan A. Hayter, MD · Zeev Friedman, MD · M. Dylan Bould, MB ·
John G. Hanlon, MD · Rita Katznelson, MD · Bruno Borges, MD ·
Viren N. Naik, MD

Received: 19 December 2008 / Revised: 9 March 2009 / Accepted: 12 March 2009 / Published online: 2 April 2009
© Canadian Anesthesiologists' Society 2009

Abstract

Purpose Technical proficiency in anesthesia has historically been determined subjectively. The purpose of this study was to establish the construct validity for the Imperial College Surgical Assessment Device (ICSAD), a measure of hand motion efficiency, as an objective assessment tool for technical skill performance, by examining its ability to distinguish between operators of different levels of experience performing a labour epidural. Concurrent validity for the ICSAD was investigated by comparison to a validated task specific checklist (CL) and global rating scale (GRS). **Methods** A single blinded, prospective, controlled study design compared three groups of subjects: novice residents (<30 epidurals), experienced residents (>100

epidurals), and staff anesthesiologists (>500 epidurals). Performance was measured using the ICSAD (number of movements, path length, time) and scores from a CL and GRS graded by examiners blinded to the level of training. Data were analyzed by multivariate analysis of variance (MANOVA).

Results Twenty-nine subjects were recruited. Novice residents had longer path lengths compared to experienced residents ($P = 0.031$) and staff anesthesiologists ($P = 0.0004$), made more movements ($P = 0.012$) and took more time than staff ($P = 0.009$). Novice residents scored significantly worse on the GRS compared to experienced residents ($P = 0.029$) and staff ($P = 0.01$) and had significantly lower CL scores compared to staff ($P = 0.003$).

Conclusions Construct and concurrent validity for the ICSAD was established for a regional anesthesia technique by demonstrating that it can distinguish between operators of different levels of experience and by comparing it to the current standards of technical skill assessment.

Attributable to: SMART (St. Michael's Anesthesia Research into Teaching) Group, Department of Anesthesiology, St. Michael's Hospital, University of Toronto, 30 Bond St., Toronto, ON M5B 1W8, Canada and the Department of Anesthesia and Pain Medicine, Mount Sinai Hospital, University of Toronto, Canada.

M. A. Hayter, MD (✉) · M. D. Bould, MB ·
J. G. Hanlon, MD · V. N. Naik, MD
Department of Anesthesiology, St. Michael's Hospital,
University of Toronto, 30 Bond St., Toronto,
ON M5B 1W8, Canada
e-mail: megan.hayter@utoronto.ca

Z. Friedman, MD · B. Borges, MD
Department of Anesthesiology, Mount Sinai Hospital,
University of Toronto, Toronto, ON, Canada

R. Katznelson, MD
Department of Anesthesia, Toronto General Hospital,
University of Toronto, Toronto, ON, Canada

Résumé

Objectif La compétence technique en anesthésie est traditionnellement évaluée de façon subjective. L'objectif de cette étude était d'établir la validité de la méthode sous-tendant le dispositif d'évaluation chirurgicale de l'Imperial College (ICSAD - Imperial College Surgical Assessment Device), une mesure de l'efficacité des mouvements de la main, en tant qu'outil d'évaluation objectif de la performance des compétences techniques. Pour ce faire, nous avons examiné la capacité du dispositif à faire la distinction entre les opérateurs de différents niveaux d'expérience réalisant une péridurale pour le travail

obstétrical. La validité concomitante de l'ICSAD a été évaluée en comparant le dispositif à une liste de contrôle spécifique à la tâche validée (CL) et une échelle d'évaluation générale (GRS).

Méthode Un concept d'étude contrôlée, prospective et en simple aveugle a comparé trois groupes : résidents novices (<30 péricuriales), résidents expérimentés (>100 péricuriales) et anesthésiologistes traitants (>500 péricuriales). La performance était mesurée avec l'ICSAD (nombre de mouvements, longueur du parcours, temps) et des scores d'une CL et d'une GRS donnés par des évaluateurs ne connaissant pas le niveau de formation. Les données ont été analysées par analyse multivariée de la variance (MANOVA).

Résultats Vingt-neuf sujets ont été recrutés. Les résidents novices avaient des parcours plus longs comparativement aux résidents expérimentés ($P = 0,031$) et aux anesthésiologistes traitants ($P = 0,0004$), faisaient plus de mouvements ($P = 0,012$) et ont pris plus de temps que les anesthésiologistes traitants ($P = 0,009$). Les résidents novices ont eu des scores considérablement moins bons sur l'échelle GRS comparativement aux résidents expérimentés ($P = 0,029$) et aux anesthésiologistes traitants ($P = 0,01$) et des scores sur la LC considérablement plus bas par rapport aux anesthésiologistes traitants ($P = 0,003$).

Conclusion La validité de la méthode et la validité concomitante de l'ICSAD a été établie pour les techniques d'anesthésie régionale en démontrant que ce dispositif peut faire la distinction entre des opérateurs de différents niveaux d'expérience et en comparant le dispositif aux normes actuelles d'évaluation des compétences techniques.

Changes to medical training programs and a reduction in residents' working hours have prompted a movement towards competency based training programs requiring reliable and objective measures to assess performance. Traditionally in anesthesia, medical knowledge has been evaluated by written examinations, and decision-making skills have been examined in oral exam format. Historically, technical skill proficiency has been assessed by direct observation and by a variety of self-reported procedure lists or logbooks.¹ There has been a recent trend towards establishing more objective measures of technical skill in medicine.² Direct observation with criteria, such as checklists and global rating scales, may improve the assessment of technical skill by reducing observer bias.³

Within anesthesiology, construct validity has been established for a variety of task-specific checklists and global rating scales.^{3,4} Checklists measure expertise by separating the technical skill into its smallest component tasks and recording whether the task is completed.

Task-specific checklists have been recently validated for various technical skills within anesthesia, including the interscalene nerve block, the labour epidural, and fiberoptic intubation.¹⁻³ Unlike checklists, global rating scales (GRS) allow for consideration of aspects of manual dexterity, such as respect for tissue, flow of procedure, movement efficiency, and instrument handling. The GRS was recently demonstrated within anesthesia as being capable of distinguishing between operators with various levels of experience in performing labour epidurals and interscalene nerve blocks.^{1,2} Despite the addition of objective criteria for evaluation, both the task-specific checklist and the global rating scale retain a degree of subjectivity. Although checklists and GRSs are widely used for technical skill assessment in various specialties, they remain in their infancy for use in technical skill assessment and have not been incorporated into formative and summative evaluations.⁴⁻⁶

Surgical specialties have initiated a trend towards a more objective and quantifiable measure of technical skill proficiency. The Imperial College Surgical Assessment Device (ICSAD) is a hand motion analysis device designed to evaluate hand motion efficiency in surgeons and also to measure their manual dexterity. Construct validity has been established for different measurements produced by the ICSAD in open, laparoscopic, and micro-surgery.⁵⁻¹⁰ It uses an electromagnetic tracking system (Isotrak II; Polhemus Inc., Colchester, VT, USA) consisting of an electromagnetic field generator and two 10 mm sensors that are attached to the dorsum of each hand.⁷ By processing the Cartesian coordinate information, the ICSAD produces three dexterity scores: total distance moved by each hand, number of movements, and total time.

In anesthesiology, the ICSAD has potential for use as an objective assessment tool of manual dexterity for technical skills. The purpose of this study was to establish the construct and concurrent validity of the ICSAD in the field of anesthesiology by comparing the hand motion efficiency of three groups of anesthesiologists with different levels of experience performing an epidural for labour analgesia. We hypothesized that the ICSAD would be able to discriminate between operators of different levels of experience in performing a labour epidural in all three ICSAD dexterity scores: path length, number of movements, and time, thus establishing construct validity for assessment of this skill. Furthermore, we postulated that there would be a correlation between the three ICSAD dexterity scores and a previously validated task-specific checklist² and GRS,^{1,2} therefore establishing concurrent validity for the ICSAD measurements.⁶ The epidural for labour analgesia was examined because of its importance in the field of anesthesiology and its particular complexity as a technical skill for anesthesiologists.⁴

Materials and methods

After institutional review board approval (Mount Sinai Hospital and St. Michael's Hospital) and informed consent, residents and attending anesthesiologists were recruited to participate in the study. Three groups of physician subjects were recruited over a 4-month period: novice residents, experienced residents and attending anesthesiologists. The timeframe was based on the availability of novice anesthesia residents completing a 4-month rotation at the beginning of their second post-graduate training year. The recruitment pool included all attending anesthesiologists and experienced residents at Mount Sinai Hospital, and all novice residents at both Mount Sinai and St. Michael's Hospitals. All subjects who were approached agreed to participate in the study, no subjects declined to participate.

The novice group included anesthesia residents who had performed less than 30 unsupervised labour epidurals at one of the participating institutions. The following is the current standard of care that takes place at our institutions: first, novice residents receive didactic lectures on labour epidural insertion; then, they observe five labour epidural insertions; and after that, they perform five labour epidurals with attending anesthesiologists observing. The residents then perform the task independently. Based on previous studies that describe the learning curve for this procedure, the experienced resident group included anesthesia residents who had performed over 100 epidurals.^{8–10} The attending anesthesiologists were participating obstetrical anesthesiologists at Mount Sinai Hospital, and they had each performed in excess of 500 epidurals. There were no exclusion criteria for physician subjects.

Informed consent was obtained from all parturients prior to participation. In order to control for technical difficulty, parturients were excluded from the study if they were morbidly obese (a body mass index >35), possessed spinal deformities, or were unable to sit still for the procedure due to excessive labour pain.

During the 4-month period, all subjects were observed once. The labour epidurals were performed with parturients in the sitting position. A midline approach was used between the L2 and L4 interspaces. An Arrow[®] 17-gauge epidural needle and catheter set was used at both hospital sites (Arrow[®], Reading, PA, USA). In order to control for fatigue and workload factors, all epidurals were performed between 08:00 AM and 12:00 PM.

Two separate recordings were made during each observation: a videotape recording and an ICSAD recording. Both recordings started simultaneously with local skin anesthetic infiltration, and both were stopped when the epidural catheter was secured in place. Two of the authors, M.H. or B.B., performed the videotaping using a Canon ZR400 digital camcorder mounted on a tripod. Videotaping was completed

in a blinded manner without audio recording to ensure proper masking of the identity and level of training of the subject. Blinding was achieved by only videotaping the subjects' hands throughout the procedure. The authors involved in videotaping the procedures were not involved in the grading process. During each observation, the subjects had the ICSAD probes applied to the dorsum of each of their hands. The ICSAD probes were applied following the preparation and sterile draping of the parturient. Following a second hand wash with alcohol, the probes were secured to the dorsum of each hand with an op-site. A sterile gown was then donned, followed by sterile gloves over the ICSAD probes.

At the end of the data collection period, the videotaped sessions were copied from the camcorder to DVDs in random order before being assessed and graded. Subjects were only identified in the videos by a unique randomly allocated study number. The two independent examiners were anesthesiologists from different centres with expertise in labour epidural placement. They were blinded to the level of experience of the subjects. The examiners evaluated each performance using a task-specific checklist and a GRS, both of which had been validated in previous studies.^{2,6} Both examiners were trained in the use of the checklist and GRS for the grading of epidural anesthesia procedures. One of the examiners had previous experience grading performances using the checklist and GRS for a prior study.

Assessment of labour epidural placement

Three assessment tools were used to evaluate the performance of each subject in the insertion of a labour epidural catheter: the ICSAD, the task-specific labour epidural checklist (Table 3 in Appendix), and a seven domain GRS (Table 4 in Appendix). The ICSAD was used to provide a purely objective measure of hand motion efficiency. We postulated that novice residents would have longer path lengths, would make more movements, and would take more time than experienced residents and attendings, as measured by the ICSAD.

The ICSAD uses an electromagnetic tracking system (Isotrak II; Polhemus Inc, Colchester, VT, USA) consisting of an electromagnetic field generator and two 10 mm sensors that are attached to the dorsum of each hand at the mid-shaft point of the third metacarpal.⁷ Surgical latex gloves secure the trackers and provide an aseptic condition for the placement of the epidural. Robotic Video and Motion Analysis Software (ROVIMAS) retrieves the time-stamped Cartesian (X, Y, and Z) coordinates at a resolution of 1 mm and a frequency of 20 Hz. Hand movements are defined by changes in hand velocity. The ICSAD tracking device processes the Cartesian coordinate information and produces values for total distance moved by each hand, number of movements, total time, and average hand velocity.

ROVIMAS then collates the raw data and extrapolates it into three discrete dexterity scores which are: (i) *number of movements* made by each hand, (ii) the *distance travelled* by each hand (path length in millimetres), and (iii) the *total time* (in sec) taken to complete the procedure. A Gaussian filter of 16 and a velocity tolerance level of $7.5 \text{ mm} \cdot \text{sec}^{-1}$ were used to eliminate background noise so that only meaningful actions were registered as movements.¹¹

The task-specific checklist consisting of 27 items rated on a three point scale was recently validated for obstetric epidural placement.² A score of 0, 1, or 2 was given when a stage was either not performed, poorly performed, or performed well, respectively.

The GRS was previously validated for obstetric epidural procedures as well as for interscalene nerve blocks.^{1,2} It consists of seven domains. Each of the domains is evaluated on a 5-point Likert scale with behavioural anchors and focuses on broad categories, such as the flow of the procedure and the use of instruments, rather than on the specifics of the manual task.

Statistical analysis

Statistical analysis was performed with SPSS 13.0® (SPSS Inc, Chicago, IL, USA). The primary endpoints included the three ICSAD dexterity scores (path length, number of movements, and total time), the score on the task-specific checklist, and the score on the GRS.

Although the individual Likert scales that make up the GRS are clearly ordinal in nature, the overall scores from the GRS behave empirically as a parametric variable¹² and, consequently, we chose to use parametrical statistical analysis, as have other researchers.⁶ Likewise, checklists with a large number of items are suitable for parametric analysis. Furthermore, the data from the ICSAD dexterity scores is tightly inter-related and skewed in nature. As a result, all of our primary endpoints, the ICSAD, the checklist, and the GRS were analyzed parametrically using a multivariate analysis of variance (MANOVA).⁶ Significant differences were analyzed using a Tukey's post hoc test.

Intraclass correlation coefficients were calculated to assess the inter-rater reliability of assessments provided by the two examiners. Following guidelines suggested by

Landis and Koch, a value >0.80 = excellent agreement, $0.61\text{--}0.80$ = substantial agreement, $0.41\text{--}0.60$ = moderate agreement, $0.21\text{--}0.40$ = fair agreement, $0.00\text{--}0.20$ = slight agreement, and less than 0.00 = poor agreement.¹³

Pearson correlation coefficients were used to establish the concurrent validity of the ICSAD with the previously validated task-specific checklist and GRS. Pearson correlation coefficients were calculated to demonstrate the relationship between all three outcome measures, namely, the three ICSAD dexterity scores, the task-specific checklist, and the GRS. Following the guidelines of Cohen, a positive or a negative value between 0.5 and 1.0 indicates a large effect, $0.3\text{--}0.5$ indicates a medium effect, and $0.1\text{--}0.3$ indicates of small effect.¹⁴ A negative Pearson correlation coefficient reflects an inverse relationship, which can be either a large, medium, or small effect between the two outcome measures being assessed.

Results

A total of 29 anesthesiologists participated as subjects (Table 1). There were nine novice residents, eight experienced residents and 12 staff anesthesiologists. A MANOVA for the ICSAD dexterity scores of path length, number of movements, and time demonstrated a significant relationship between all three dexterity scores and experience groups, $F(6, 46) = 2.41$, $P = 0.042$. Univariate testing found a significant association between groups and each of the dexterity scores: path length: $F(2, 27) = 7.00$, $P = 0.004$; movement: $F(2, 27) = 4.99$, $P = 0.015$; time: $F(2, 27) = 5.64$, $P = 0.010$. A Tukey's post hoc analysis was conducted to evaluate the pair-wise differences among the means of each group for path length, movement, and time. The ICSAD data demonstrated that novice residents made significantly more movements ($P = 0.012$) than staff anesthesiologists; they also had a significantly longer path length than both experienced residents ($P = 0.031$) and staff anesthesiologists ($P = 0.0004$), and they took a significantly longer time than staff anesthesiologists ($P = 0.009$) to complete the labour epidurals (Fig. 1).

The interrater reliability of the task-specific checklist was excellent, with an intraclass correlation coefficient (ICC)

Table 1 Demographic characteristics

	Novice residents (<i>n</i> = 9)	Experienced residents (<i>n</i> = 8)	Staff anesthesiologists (<i>n</i> = 12)
Number of postgraduate months; median (IQR)	13.0 (13.0–14.0)	23.0 (23.0–28.5)	120.0 (60.0–240.0)
Male/female	5/4	3/5	10/2
Number of epidurals; median (IQR)	12.00 (5.0–17.5)	>100	>500

IQR interquartile range

Fig. 1 Performance of Imperial College Surgical Assessment Device (ICSAD) dexterity scores by experience level: number of movements, path length (mm), and time (sec)

Table 2 Pearson correlation coefficients between the three ICSAD dexterity scores and the checklist and global rating scale

ICSAD endpoint	Checklist	<i>P</i>	Global rating scale	<i>P</i>
Path length	−0.28	0.16	−0.52	0.006**
Number of movements	−0.53	0.005**	−0.41	0.032*
Time	−0.29	0.15	−0.51**	0.006**

** $P < 0.01$; * $P < 0.05$

and 95% confidence interval (CI95) of 0.63 (0.35, 0.80), $P < 0.001$. Novice residents scored significantly lower on the checklist than staff anesthesiologists ($P = 0.003$).

The interrater reliability of the GRS was excellent, with an ICC and CI95 of 0.66 (0.40, 0.82), $P < 0.001$. Novice residents had significantly lower global rating scores than experienced residents ($P = 0.029$) and staff anesthesiologists ($P = 0.01$).

Concurrent validity of the ICSAD was evaluated by comparing it to the task-specific checklist and the GRS. Pearson correlation coefficients between the three ICSAD dexterity scores (path length, number of movements, and time) and the task-specific checklist or the GRS demonstrated a fair to strong correlation (Table 2). As expected, a negative correlation was demonstrated between the individual ICSAD dexterity scores and the task-specific checklist and the GRS.

Discussion

The results of this study indicate that the Imperial College Surgical Assessment Device (ICSAD) is a reliable measure of technical performance of a labour epidural catheter

insertion, and it can discriminate between operators of different levels of training. Specifically, the three dexterity scores of the ICSAD demonstrate a significant difference between operators of various levels of experience. In addition, the data from the current gold standards of technical skill assessment, i.e., the checklist and the GRS, support the concurrent validity of the ICSAD. The weak correlation between time and the task-specific checklist is supported by theories of expertise which distinguish experts from experienced non-experts by their ability to remain adaptive to the specifics of the situation rather than to speed-of-task completion.¹⁵

Interestingly, the outcome measures did not make an overall distinction between experienced residents and staff anesthesiologists. A possible explanation for this finding is that experienced residents, having completed over 100 labour epidurals, have progressed to becoming experienced non-experts. Experienced non-experts are individuals who perform well on routine problems by unreflectively and automatically applying the standard technique.¹⁶ This is supported by previous literature on learning curves for labour epidurals.⁹ After completing over 100 epidurals, anesthesia residents achieve competency and may approach failure rates similar to staff anesthesiologists.^{9,10} An interesting future study would be to include patients with a variety of complicating factors. Perhaps this added complexity would allow finer distinction of skill level and a better understanding of the cognitive and technical processes that experts utilize.

An important limitation of this study was the fact that we were not sufficiently powered to detect a difference between novice and experienced residents on most ICSAD dexterity scores, and a difference between experienced residents and attending anesthesiologists on all ICSAD dexterity scores. However, our findings regarding data from the ICSAD were supported by the scores on the checklist and the GRS, which detected similar differences between the three subject groups. As such, the sensitivity of the ICSAD as an evaluation has not yet been established. Future studies are indicated that are powered sufficiently to determine the sensitivity of the ICSAD.

Traditionally, anesthesiology and surgical specialties have relied primarily on informal subjective assessments as a means of both formative and summative assessment of technical skill.¹ The current climate in medical education favours objective and reproducible systems of technical skill assessment. Anesthesia lags behind other specialties in the application of more rigorous assessment tools for technical skill performance. Previous studies examining the success rates of regional anesthesia have attempted to quantify a minimum number of procedures required to attain consistency in regional anesthesia.^{8–10} However, this methodology only documents number of procedures and

not proficiency or success rate. Moreover, attaining a prescribed number of procedures does not guarantee competence, as trainees could continue to perform the skill incorrectly.¹⁷

Regional anesthesia techniques are more difficult to learn than the manual skills required for administering general anesthesia. Recently, various checklists for specific regional anesthetic techniques were developed.^{1,2} Although checklists provide objective parameters by which the observer can evaluate the trainee, there remains a degree of subjectivity in the assessment of the performance. Global rating scales provide a gestalt impression of the performance, irrespective of whether checklist criteria were met; however, this is based on the observer's opinion. The ICSAD provides a means of reliably and objectively assessing manual dexterity while performing procedures. The surgical literature has demonstrated that economy of movement and path length and reduced total time taken to complete procedures is associated with increasing expertise.^{18–21} This study begins to demonstrate the construct and concurrent validity of the ICSAD for a regional anesthesia procedure.

While regional anesthesia procedures are ever-increasing in popularity,²² there remain a limited number of evaluation tools available to formally assess competency with regional techniques. Training programs are becoming

more accountable to regulatory bodies in evaluating physicians in training against accepted standards which continue to be developed and validated.²³ As such, the ICSAD may provide an objective assessment tool to provide a quantified result that has been compared to an acceptable standard. Future research that is sufficiently powered to discriminate operators of various degrees of experience is needed in order to establish the ICSAD as an effective assessment tool for anesthesiology trainees. In the future, these assessment tools may form the basis of a national evaluation standard for documenting acquisition of technical skills.

Acknowledgments The authors sincerely thank the anesthesiology residents and staff at the University of Toronto for their participation in the study.

Financial support Supported in part by the Royal College of Physicians and Surgeons of Canada through the 2008 Medical Education Research Grant.

Conflicts of interest None declared.

Appendix

See Tables 3 and 4.

Table 3 Examiners task-specific checklist for epidural anesthesia

Stages performed	Not performed	Performed poorly	Performed well
1 Ensures patient is positioned comfortably and safely in the middle of the bed			
2 Adjusts height of bed appropriately			
3 Carefully prepares a sterile work surface			
4 Pours antiseptic solution (or has nurse pour it) without contaminating the epidural set			
5 Washes hands and puts on gloves in a sterile fashion			
6 Optimally positions him/herself for the procedure			
7 Prepares the skin at the back widely and aseptically (skin prep × 3)			
8 Allows solution to dry			
9 Neatly lays out and prepares all necessary equipment (needles, syringes, local anesthetic)			
10 Asks patient to arch her back			
11 Places drape over patient's back in a sterile fashion			
12 Landmarks site of injection after palpating iliac crests			
13 Warns patient of needle insertion			
14 Infiltrates subcutaneous layers with local anesthetic			
15 Places epidural needle with correct positioning of bevel			
16 Inserts epidural needle through skin, subcutaneous tissue and into ligament before attaching the syringe			
17 Attaches air/saline filled syringe to the needle hub with needle well controlled			

LOR loss of resistance

Table 3 Examiners task-specific checklist for epidural anesthesia

Stages performed	Not performed	Performed poorly	Performed well
18 Braces hand/s holding the needle against patient's back in complete control of the needle			
19 Slowly advances needle through supra spinous and inter spinous ligaments and into ligamentum flavum while applying pressure on the plunger (continuous or intermittent)			
20 Identifies LOR and immediately releases pressure on the plunger			
21 Notes depth of needle insertion before threading catheter			
22 Warns patient about possible paresthesia during catheter threading			
23 Detaches the syringe and threads the catheter to a depth of 4–5 cm			
24 Pulls the needle out while maintaining correct catheter placement			
25 Carefully aspirates from catheter			
26 Injects test dose through flushed filter			
27 Fixes the epidural catheter securely			

LOR loss of resistance

Table 4 Global rating scale for epidural anesthesia

	1	2	3	4	5
Preparation for procedure	Did not organize equipment well. Has to stop procedure frequently to prepare equipment		Equipment generally organized. Occasionally has to stop and prepare items		All equipment neatly organized prepared and ready for use
Respect for tissue	Frequently used unnecessary force on tissue or caused damage		Careful handling of tissue but occasionally caused inadvertent damage		Consistently handled tissues appropriately with minimal damage
Time and motion	Many unnecessary moves		Efficient time/motion but some unnecessary moves		Clear economy of movement and maximum efficiency
Instrument handling	Repeatedly makes tentative or awkward moves with instruments		Competent use of instruments but occasionally appeared stiff or awkward		Fluid moves with instruments and no awkwardness
Flow of procedure	Frequently stopped procedure and seemed unsure of next move		Demonstrated some forward planning with reasonable progression of procedure		Obviously planned course of procedure with effortless flow from one move to the next
Knowledge of procedure	Deficient knowledge		Knew all important steps of procedure		Demonstrated familiarity with all aspects of procedure
Overall performance	Very poor		Competent		Clearly superior
Overall, should the candidate:		Pass	Fail		

References

1. Naik VN, Perlas A, Chandra DB, Chung DY, Chan VW. An assessment tool for brachial plexus regional anesthesia performance: establishing construct validity and reliability. *Reg Anesth Pain Med* 2007; 32: 41–5.
2. Friedman Z, Katznelson R, Devito I, Siddiqui M, Chan V. Objective assessment of manual skills and proficiency in performing epidural anesthesia—video-assisted validation. *Reg Anesth Pain Med* 2006; 31: 304–10.
3. Naik V, Matsumoto E, Houston P, et al. Learning fiberoptic intubation on a simple model transfers to the O.R. *Can J Anesth* 2001; 48: A2. abstract.
4. Neequaye SK, Aggarwal R, Van Herzele I, Darzi A, Cheshire NJ. Endovascular skills training and assessment. *J Vasc Surg* 2007; 46: 1055–64.
5. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative “bench station” examination. *Am J Surg* 1997; 173: 226–30.
6. Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998; 73: 993–7.
7. Ridgway PF, Ziprin P, Datta VK, et al. Laboratory-based validation of a novel suture technique for wound closure. *Ann Plast Surg* 2002; 49: 291–6.
8. Konrad C, Schupfer G, Wietlisbach M, Gerber H. Learning manual skills in anesthesiology: is there a recommended number of cases for anesthetic procedures? *Anesth Analg* 1998; 86: 635–9.
9. Naik VN, Devito I, Halpern SH. Cusum analysis is a useful tool to assess resident proficiency at insertion of labour epidurals. *Can J Anesth* 2003; 50: 694–8.
10. Kopacz DJ, Neal JM, Pollock JE. The regional anesthesia “learning curve”. What is the minimum number of epidural and spinal blocks to reach consistency? *Reg Anesth* 1996; 21: 182–90.
11. Brydges R, Classen R, Larmer J, Xeroulis G, Dubrowski A. Computer-assisted assessment of one-handed knot tying skills performed within various contexts: a construct validity study. *Am J Surg* 2006; 192: 109–13.
12. Carifio J, Perla R. Resolving the 50-year debate around using and misusing Likert scales. *Med Educ* 2008; 42: 1150–2.
13. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977; 33: 363–74.
14. Cohen J. Statistical power analysis. *Curr Dir Psychol Sci* 1992; 1: 98–101.
15. Mylopoulos M, Regehr G. Cognitive metaphors of expertise and knowledge: prospects and limitations for medical education. *Med Educ* 2007; 41: 1159–65.
16. Moulton CA, Regehr G, Mylopoulos M, MacRae HM. Slowing down when you should: a new model of expert judgment. *Acad Med* 2007; 82: S109–16.
17. Watts J, Feldman WB. Assessment of technical skills. In: Neufeld VR, Norman GR, editors. *Assessing clinical competence*. New York: Springer; 1985. p. 259–74.
18. Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surg* 2001; 193: 479–85.
19. Moorthy K, Munz Y, Dosis A, Bello F, Chang A, Darzi A. Bimodal assessment of laparoscopic suturing skills: construct and concurrent validity. *Surg Endosc* 2004; 18: 1608–12.
20. Saleh GM, Voyatzis Y, Hance J, Ratnasothy J, Darzi A. Evaluating surgical dexterity during corneal suturing. *Arch Ophthalmol* 2006; 124: 1263–6.
21. Smith SG, Torkington J, Brown TJ, Taffinder NJ, Darzi A. Motion analysis. *Surg Endosc* 2002; 16: 640–5.
22. Borgeat A, Ekatodramis G, Schenker CA. Postoperative nausea and vomiting in regional anesthesia: a review. *Anesthesiology* 2003; 98: 530–47.
23. Steinbrook R. The debate over residents’ work hours. *N Engl J Med* 2002; 347: 1296–302.