ORIGINAL REPORTS

Assessment of Competence in Surgical Skills Using Functional Magnetic Resonance Imaging: A Feasibility Study

Marie C. Morris, MSc, BSc Hons,[*] Thomas Frodl, MD,[†] Arun D'Souza, PhD,[†] Andrew J. Fagan, PhD,[‡] and Paul F. Ridgway, MD, MMedSc, FRCS (Gen Surg)[*]

[*]Education Division and Department of Surgery, School of Medicine, Trinity College, Tallaght, Dublin, Ireland; [†]Integrated Neuroimaging, Department of Psychiatry, Trinity College, Dublin, Ireland; and [‡]Centre for Advanced Magnetic Imaging (CAMI), St James's Hospital, School of Medicine, Trinity College, Dublin, Ireland

BACKGROUND: Patient safety is fundamental to modern medical practice; safe surgery saves lives. Ensuring surgical competence is becoming more difficult at a time when surgeons are being trained in fewer hours. Accurate objective assessment of technical skills ability is lacking in standardization. Functional magnetic resonance imaging (fMRI) has a long history in neuroscience, psychiatry, and cognitive studies. Many studies have explored levels of perceived expertise in sports and musical ability. Little has been published on actual rather than perceived motor skills. This study sought to assess the feasibility of utilizing a novel assessment method by measuring blood oxygen level–dependent signal changes (BOLD) in specific brain regions via fMRI during a surgical skills task.

METHODS: Images were acquired using fMRI in a pilot study of 9 subjects (3 experts, 3 intermediates, and 3 novices) when performing and imagining performing a basic surgical procedure: hand tying of surgical knots. Level of expertise was based on years of experience and clinical grade. The quality and quantity of knots were assessed objectively by 2 experts who were independent of the study and blinded to the ability of the candidate. The effect of subject head motion caused by the task itself was assessed. The efficacy of fMRI data analyses in removing artifacts caused by this noise source in the data was explored.

RESULTS: Shifts of less than 1 voxel ($3 \times 3 \times 3.55$ mm^3) were recorded in all participants and were successfully corrected in all cases in the fMRI preprocessing step. Decreased BOLD activity was observed in experts compared to novices when "knot tying" was compared with the control "finger tap." Increased BOLD activity was observed in experts compared with novices when "imagining a task" in the primary visual cortex, an area important in perceptual learning. Experts and intermediates performed consistently with 100% square knots. Novices had an average of 2 slip knots. Regarding knot quantity, the number of knots ranged from 14 to 26 in novices, 38 to 47 in intermediates, and 54 to 58 in experts. A Kruskal-Wallis rank sum test revealed that the difference between the 3 groups was statistically significant in the quantity of square knots tied ($p = 0.147$). Specific regions of interest identified concurred with findings of previous studies and included the left supramarginal, left rolandic operculum, and left postcentral regions.

CONCLUSION: We found that fMRI is a feasible method of exploring actual and perceived motor skill abilities. Head motion during performance of a motor skill does not preclude the attainment of meaningful data. Larger numbers are needed to further investigate these early findings. (J Surg ■:■■■-■■■. © 2014 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: surgical skills, motor skills, competence, assessment, fMRI

COMPETENCIES: Patient Care, Medical Knowledge, Practice Based Learning and Improvement, Systems Based Practice, Professionalism, Interpersonal Skills and Communication

BACKGROUND

The brains of surgical experts may be regarded as the ideal subjects to explore the relationship between cerebral plasticity and learning of complex motor skills. Functional magnetic resonance imaging (fMRI) is a specialized sub

Ethics: Approval for this study was sought and obtained from the local ethics committee.

Correspondence: Inquiries to Marie C. Morris, MSc, BSc Hons, Education Division and Department of Surgery, School of Medicine, Trinity College, Tallaght Hospital Campus, Dublin 24, Ireland; fax: (189) 6-2910; e-mail: morrism4@tcd.ie

type of MRI that has come to dominate the field of brain mapping.[1] It has relatively low invasiveness with the absence of radiation exposure. The availability of reliable, valid objective assessment methods is fundamental in determining the level of surgical skill.[2-5] A variety of direct and indirect measures are currently available for evaluation, but no single method permits high-stakes inferences regarding surgical skills competence. In the postgraduate setting, methods are standardized and systematic and offer some element of comprehensive evaluation.[3] Much work is needed in the undergraduate setting to identify the optimal combination of methods to ensure initial attainment and ongoing maintenance of competence for lifelong practice as a surgeon.

Studies using fMRI represent a novel measure of motor skill ability. It records the patterns of cerebral activation typically following a stimulus presented to the subject; however, it could equally be used to assess brain activation following a motor function, thereby offering the possibility of using fMRI to differentiate between novice, intermediate, and expert skill levels. This hypothesis for pure surgical skill remains untested. It is well known that fMRI is very sensitive to artifacts created by head motion and magnetic field inhomogeneities. This has to date been seen to be prohibitive for actual motor action studies.

Studies with fMRI have been performed in babies to assess basic motor skills such as grasping[6]; however, it is now widely accepted that humans have a motor repertoire that far exceeds object-orientated actions. The best cited example is by Calvo-Merino et al.[7] who performed a study involving fMRI of expert dancers. In this study, expert dancers, on reviewing videos of familiar dance patterns, demonstrated greater bilateral blood oxygen level–dependent (BOLD) signal changes[8] across the whole brain with specific regions of interest (ROIs) identified in the premotor cortex, intraparietal sulcus, right superior parietal lobe, and left posterior superior temporal sulcus (STS). This study suggested that the human brain understands actions by motor simulation. More recently, the area of sports science has utilized fMRI to investigate the effect of expertise on cortical activation during sport anticipation.[9] Further fMRI studies have also shown different patterns of expertise effects in a range of perceptual motor skills in pianists,[10] opera singers,[11] chess players,[12] archery,[13] and tennis players.[14]

This study sought to investigate the use of fMRI to assess motor skill performance, namely hand tying of surgical knots. The previously assumed potential problems in motor skills studies with this approach, i.e., that differences may not show up and essentially image artifacts caused by excessive head motion (caused by the task itself) may obscure any findings, were explored. The aim was to investigate if the presumed prohibitive effects of artifacts could be overcome. Moreover, whether fMRI could potentially differentiate between levels of ability in basic surgical skills, i.e., hand tying of surgical knots, was also investigated in this preliminary research.

METHODOLOGY

Research Setting

This study was undertaken in the Centre for Advanced Magnetic Imaging (CAMI), St James's Hospital, Dublin, Ireland.

Participants

On the advice of fMRI experts, this study involved 9 participants; this sample size was deemed appropriate for a pilot or feasibility study. The 9 participants were divided into 3 groups based on the number of years of surgical experience and expertise as follows:

3 "Novices": Undergraduate medical students in years 3 and 4 of a 5-year program who had 4 hours training in knot tying on jigs.

3 "Intermediates": Senior house officers with at least 2 years of postgraduate surgical experience.

3 "Experts": Consultant/specialist registrars with at least 5 years of postgraduate surgical experience.

Participants were not included in the study if they had a previous head injury with loss of consciousness, psychosis, personality disorders, cortisol medication in the medical history, severe internal or neurological diseases, current alcohol or substance abuse, and dependency. There was no limitation on age or right- or left-handedness.

This criterion for inclusion was supplemental to the usual safety exclusion from fMRI—presence of metal, etc. If participants required glasses to rectify a visual deficit, they were asked to wear contact lenses for the imaging.

PROTOCOL

Written instructions on the study sequence were given to all participants 2 weeks before the study to allow time to address any queries. The researcher's (M.C.M.) phone number was supplied for queries. Ethical approval was obtained for the study from the local institutional review board (Adelaide and Meath Hospital and St James's Hospital Ethics Committee) and written informed consent was obtained from each participant. The study followed the standard fMRI sequencing protocol given in Table 1. Participants were positioned in the scanner in a standardized fashion throughout the study duration, with a commercially available hand-tying jig (Ethicon) that had all metal attachments removed to ensure it was compatible with use in MRI. Participants were equipped with 6 standard-length shoe laces positioned at waist level so as to be easily accessible to the participants while in the scanner to avoid any contribution to head motion during scanning. Angled mirrors were utilized to allow participants to view the jig and all laces without head movement. Ear plugs and head phones were supplied to protect the participants from the

TABLE 1. Basic Workflow of an fMRI Experiment From Image Acquisition, Image Processing, and Task Procedure Through Data Analysis

(1) *Image acquisition*: fMRI images were acquired using a T_2*-weighted SE-EPI–based sequence; slice thickness was set at 4.8 mm.
(2) *Image processing*: Some processing programs are developed as stand-alone systems outside the scanner system. They provide the computational capability to reconstruct the large numbers of images and statistical analyses that identify the active areas.
(3) *Task procedure*: Patients are positioned in the scanner as for a conventional scan. During a functional imaging series, multiple repeated images of the brain are acquired over a given period. To identify brain tissue involved in language, sensory, visual, auditory, hand movement, and other targeted functions, the patient acts accordingly during the activity epoch. The beginning and end of this activity period are cued by a visual or auditory signal.
(4) *Data analysis*: Statistical analyses are used to identify areas of the brain activated by specific tasks and are based on a multistage comparison of stimulation and resting intensity levels and of multiple replications.

SE-EPI, spin-echo–echo-planar imaging.

high decibel of noise generated by the scanner. A standardized set of audio instructions was relayed to each participant while in the scanner. An emergency bell was supplied and tested at the start of each scan; this was secured in close proximity to the participant's dominant hand.

STUDY MODEL

The tasks presented to each participant were event related, with each trial in the task lasting 10 seconds and consisting of alternative finger tapping (control), rest period (null event), " pick" up shoe lace (null event), knot tying (action), rest period, and imagining tying a knot (perception); details of the protocol are presented in Table 2. "Finger tapping" was utilized as a control, acting as a very basic motor function during which all participants' BOLD activity would be similar. "Imagine" was utilized as an abstract thought process to differentiate between thoughts and actions. This sequence was repeated 10 times, resulting in a total of 380 dynamics per participant as per previously cited studies.

All MRI data were acquired using a 3T Achieva scanner (Philips Medical Systems, the Netherlands) in CAMI, St James's Hospital, Dublin. The fMRI protocol consisted of the acquisition of a high-resolution 3D T_1-weighted structural dataset (spoiled gradient echo sequence with repetition time/echo time = 8.5/3.9 ms and $1 \times 1 \times 1$ mm^3 spatial resolution), followed by an fMRI experiment (spin-echo–echo-planar imaging sequence with repetition time/echo time = 2000/35 ms, in-place resolution = 3×3 mm^2, 4.8-mm slice thickness, and 380 dynamic scans each with 2-s duration). A short T_2-weighted image was recorded over 2 minutes to elicit incidental findings (as per the CAMI's procedures), and all scans were reviewed by a consultant radiologist.

DATA ANALYSIS

Data were analyzed using SPM8 (Statistic Parametric Mapping, Wellcome Trust Centre for Neuroimaging, UCL, UK) running under MATLAB (The MathWorks Inc, USA). The threshold was set at 15 voxels to reduce the likelihood of false recovery rates. Using SPM 8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8), fMRI data were preprocessed using the following steps: first, compensation of systematic, slice-dependent time shifts; second, the elimination of systematic odd-even slice intensity differences due to interleaved acquisition; and third, rigid body correction for interframe head motion within and across runs. Next, coregistration of the structural T_1 image to the functional scans was carried out. Spatial normalization to standard $3 \times 3 \times 3$ mm^3 Montreal Neurological Institute space was then applied to the functional images and to the structural image to allow for intersubject analysis. The

TABLE 2. fMRI Study Sequence

Scan	Detail	Action	Duration	
Diagnostic scan	3D volumetric scan	To co-register with fMRI data	8 min	
Diagnostic scan	Short T_2-weighted image	To elicit incidental findings	2 min	
fMRI study—run 1	Baseline—control	Participants required to tap index fingers and thumbs together intermittently	10 s	
	Rest period	Null event—no movement	10 s	Repeat 10 times
	Perform task 1	Hand tie knots on jig	10 s	
	Rest period	Null event—no movement	10 s	
	Perform task 2	Lie still and imagine tying a knot on the jig	10 s	
	Rest period	Null event—no movement	10 s	

functional images were then spatially smoothed (smoothing full width at half maximum = 8 mm).

Statistical significance was accepted at $p < 0.001$ for the whole brain and $p < 0.05$ for ROIs in keeping with neurosciences research practice. To guarantee objective and consistent labeling of brain regions, the Automated Anatomic Labeling Toolbox was used to identify in which activities the differences were observed.

Competence in knot tying was objectively assessed with regard to the quality and quantity of knots tied using the commercial jig. The tool utilized to assess the knots was devised in-house and was an item-specific checklist. Face and content validities were ensured through expert review by consultants who were not involved in the study. Each participant's knots jig was independently assessed by 2 experts in surgical skills; both were blinded to the categorization of participants (novice, intermediate, or expert) for validation.

RESULTS

The voxel shift or "artifact" during data acquisition was recorded at less than the length of 1 voxel (3 mm^3) movement throughout the entire study sequence for all participants. This is at least as low as artifacts reported in the previous studies cited. The "tapping" activity appears to be an accurate control, as the BOLD activity in the motor cortex was the same in the novice, intermediate, and expert groups as would be anticipated with such a basic motor action. There was less "BOLD" activity in experts than in novices during knot tying compared with during finger tapping (Fig. 1). With regard to the abstract thought process "imagine," an increase in BOLD activity was found in experts compared with novices at the temporal parietal junction and posterior STS, as represented in Figure 2.

Most of the group comparisons (experts vs intermediates vs novices) revealed no significant differences ($p < 0.001$) after correcting for the whole brain. However, there were clusters of activity differences for some comparisons. The activity difference for contrast "tie:tap" was smaller in experts than in novices (Fig. 1 and Table 3). This was not statistically significant but may become significant in a follow-up study using larger numbers and a priori data. Regarding the "imagine" task, increased activity was seen in the primary visual cortex (V1) in experts, which is an area important in perceptual learning. This difference is significant at cluster level $p < 0.05$ (Fig. 2 and Table 4). ROIs at peak voxel level p (familywise error corrected) included the left supramarginal, left rolandic operculum, and left postcentral regions, which is where " imagining activity" would be expected.

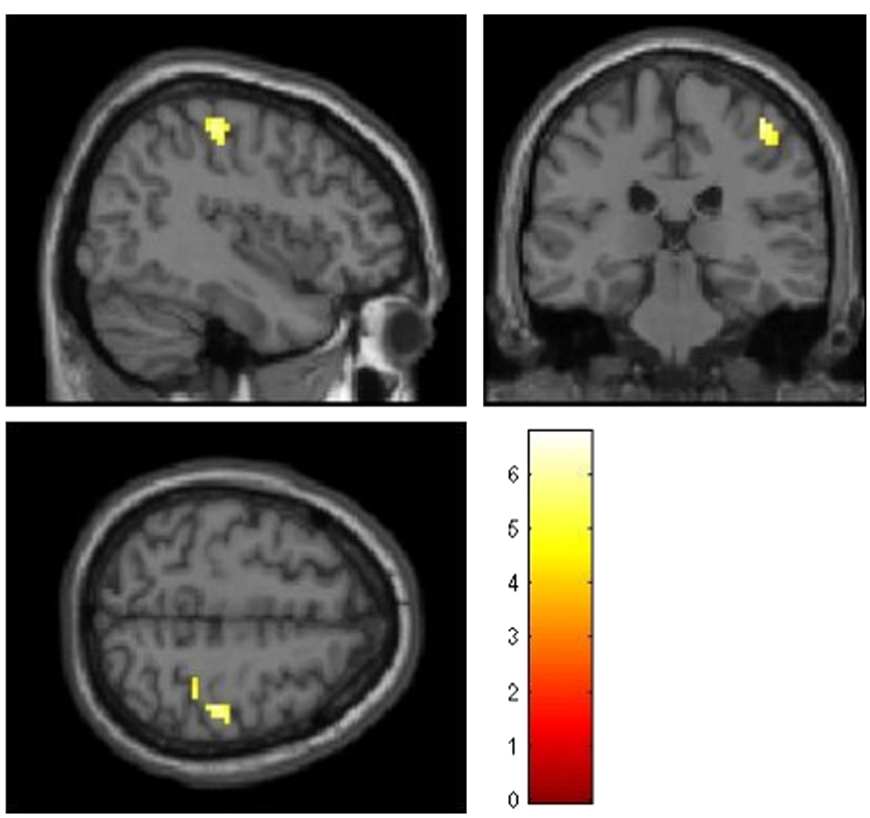

FIGURE 1. The right postcentral cortex is shown, in which experts show less BOLD responses compared with novices when they tie knots in contrast to finger tap. All differences have a threshold at $p < 0.001$.

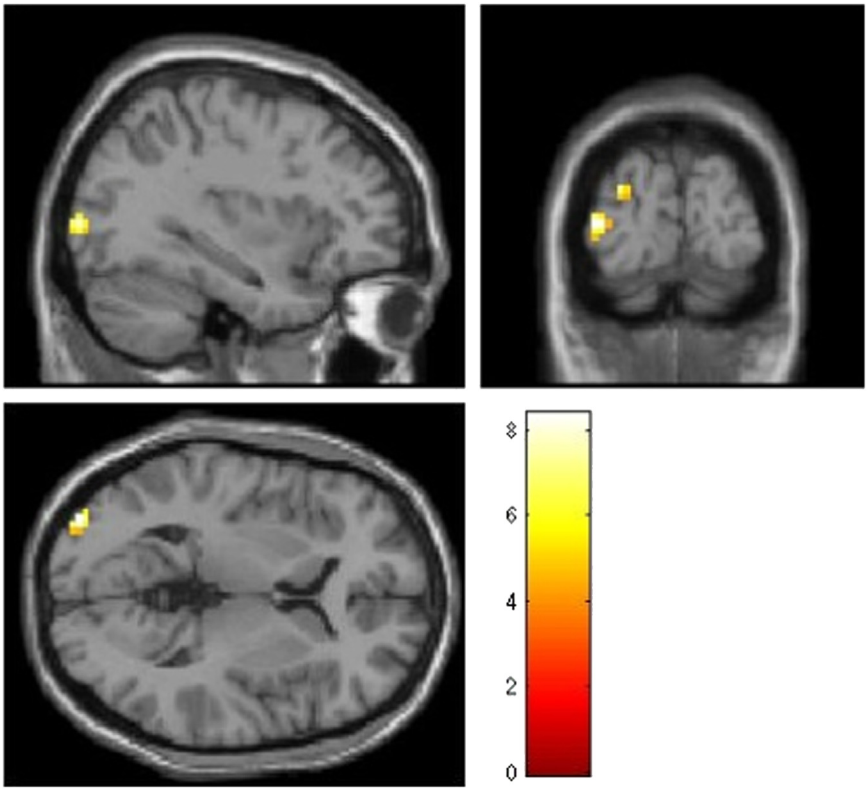

FIGURE 2. The middle occipital cortex is shown, in which experts show more BOLD responses compared with novices when they imagine knot tying in contrast to actually tying knots. All differences have a threshold at $p < 0.001$.

Regarding surgical skills competence, experts and intermediates performed consistently with 100% square knots. Novices had an average of 2 slip knots. Regarding knot quantity, the number of knots ranged from 14 to 26 in novices, 38 to 47 in intermediates, and 54 to 58 in experts (Table 5). A Kruskal-Wallis rank sum test revealed that the difference between the 3 groups was statistically significant in the quantity of square knots tied, with $p = 0.147$. However, further analysis to identify whether the difference was between the novices and intermediates or between intermediates and experts was limited by the small sample size.

TABLE 3. Differences in fMRI BOLD Responses Between Experts and Novices During Knot Tying

				Coordinates (mm)		
K	*FWE*$_{cluster}$	Region	*t*	*x*	*y*	*z*
16	0.671	Cerebellum R	6.77	6	−82	−28
37	0.353	Postcentral R	6.14	42	−28	56
		Postcentral R	4.91	25	−37	56

FWE, familywise error.

Regions are indicated where experts show less BOLD responses compared with novices when they tie knots in contrast to finger tap. *K* = number of voxels (> 15). All differences have a threshold at $p < 0.001$. FWE cluster represents the p value after correction on the whole brain.

DISCUSSION

Our study shows that assessing actual motor skill performance is feasible using fMRI techniques. In particular, it demonstrates that skills requiring considerable hand movement do not create prohibitive levels of motion-induced artifacts (and consequently, it is possible to obtain meaningful data using fMRI). The BOLD activity during "finger tap" was equal in all categories, indicating that this was an accurate control event. The activity difference for {contrast

TABLE 4. Differences in fMRI BOLD Responses Between Experts and Novices During "Imagine"

				Coordinates (mm)		
K	*FWE*$_{cluster}$	Region	*t*	*x*	*y*	*z*
22	0.542	Middle occipital L	8.39	−33	−91	4
18	0.626	Middle occipital L	8.09	−24	−66	20
19	0.604	Cerebellum R	7.56	21	−79	−24

FWE, familywise error.

Regions are indicated where experts show more BOLD responses compared with Novices when they imagine knot tying in contrast to really tying knots. *K* = number of voxels (> 15). All differences have a threshold at $p < 0.001$. FWE cluster represents the p value after correction on the whole brain.

TABLE 5. Quantity of Knots by Category

Category	Age (y)	Designation	No. of Square Knots
Novice	23	Med student year 4	26
Novice	22	Med student year 4	22
Novice	22	Med student year 4	14
Intermediate	29	SHO-year 1	38
Intermediate	27	SHO-year 1	42
Intermediate	27	SHO-year 1	47
Expert	38	SpR	54
Expert	50	Consultant	58
Expert	39	Consultant	55

Med, medical; SHO, senior house officer; SpR, specialist registrar.

tie > tap} was smaller in experts than in novices (Fig. 1 and Table 3), and thus it would appear that experts demonstrate less BOLD activity when they are proficient in the performance of tying a knot, indicating that this motor action had become "automatic" through repeated practice, requiring little conscious thought processes. This concurs with the fMRI findings by Wright et al.[9] in sports-related anticipation research.

Regarding {imagine > tie: Experts > Novices}, this comparison shows that for the contrast {*imagine* > *tie*} the differences were greater for experts than they were for novices (Fig. 2 and Table 4). Thus, although both experts and novices may likewise show more activity during imagining particular action vs executing this action, it is the experts for whom this difference is more prominent. Moreover, this difference is located in the V1, an area that is very important in perceptual learning. Activity in this area relates to "seeing with the mental eye" in this context. Given that neurons in V1 are monocular with highly specialized input and given that it is experts who show enlarged activity ("imagining") this is as expected. This is a very interesting find and warrants further inquiry, as it implies that V1 being active without any actual retinal input could constantly be interfering (experience dependent) with what we think we see (similar to memory in the case of witnesses). This abstract thought process "imagine" resulted in increased BOLD activity at the temporal parietal junction and posterior STS in experts when compared with novices. This agrees with the data from the study by Calvo-Merinos et al.,[7] wherein similar ROIs were identified with expert dancers reviewing familiar and unfamiliar dance videos. Thus, it would appear that action observation and imaging the performance of a motor skill involve BOLD activity in the same brain regions. This finding needs to be explored with a larger sample size.

An interesting observation with the 9 participants was that of what appeared to be "performance anxiety." Novices and intermediate participants undertook the fMRI sequence and appeared unperturbed with regard to the quality and quantity of knots actually tied. However, all in the expert group sought feedback on their performance and that of their peer group. Perhaps this was owing to the fact that their performance would be evaluated by 2 independent reviewers. This is being investigated further in a current follow-up study.

Our study demonstrates that despite the limitation of a small sample size, there is evidence of differences in BOLD activity between actual skill performance (tie) and in perceived skill ability (imagine) in novices and experts. Larger numbers are needed to explore these findings. Correlations between BOLD activity data and formal training and years of practice can be explored to attempt to identify at what point in intern training the novice becomes proficient and at what point in residency training the intermediates become refined. This has potential clinical and educational implications. Subsequently, a further larger-cohort study will investigate possible differences in activation patterns across the whole brain during an actual suturing task and imagining a complex surgical procedure.

LIMITATIONS

This study has a few limitations, which include a small sample number, an artificial study setting in an fMRI scanner, and the cost implications of using this technology.

CONCLUSION

This study demonstrates that fMRI appears to be a feasible method for differentiating levels of surgical skills ability. The potential for utilizing fMRI objectively to identify actual rather than perceived surgical skills ability is tangible. New ROIs have been identified, namely left supramarginal, left rolandic operculum, and left postcentral regions, which warrants further investigation. fMRI is a promising method for identifying the cerebral regions involved in the process of motor learning and providing insight into the mechanisms underlying the generalization of actions. Our study suggests that, although limited data exist at present, with ongoing research, fMRI may guide training and assessment in surgical skill acquisition.

ACKNOWLEDGMENTS

The authors acknowledge assistance from Aliz Takas in the Centre for Advanced Magnetic Imaging (CAMI) and also institutional support to CAMI from the Health Research Board (HRB), Ireland.

REFERENCES

1. Belliveau JW. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*. 1991;254(5032):716-719.

2. Harden R. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ*. 1979;13(1):41-54.
3. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative “bench-station” examination. *Am J Surg*. 1997;173(3):226-230.
4. Kogan J. Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *J Am Med Assoc*. 2009;302(12):1316-1326.
5. Van der Vleuten CPM, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ*. 2005;39(3):309-317.
6. Grafton ST, Arbib MA, Fadiga L, Rizzolatti G. Localizations of grasp representations in humans by positron emission tomography. *Exp Brain Res*. 1996;112(1):103-111.
7. Calvo-Merino B, Glaser DE, Grezes J, Passingham RE, Haggard P. Action observation and acquired motor skills—an fMRI study with Expert Dancers. *Cereb Cortex*. 2005;15(8):1243-1249.
8. Ogawa S. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A*. 1990;87(24):9868-9872.
9. Wright MJ, Bishop DT, Jackson RC, Abernethy B. Functional MRI reveals expert-novice differences during sport-related anticipation. *Neuroreport*. 2010;21(2):94-98.
10. Landau S, D`Esposito M. Sequence learning in both pianiats and nonpianists: an fMRI study of motor expertise. *Cogn Affect Behav Neurosci*. 2006;6(3):246-259.
11. Kleber B, Veit R, Birbaumer N, Gruzelier J, Lotze M. The brain of opera singers: experience-dependant changes in functional activation. *Cereb Cortex*. 2010;20(5):1144-1152.
12. Bilalic M. Many faces of expertise: fusiform face area in chess experts and novices. *J Neurosci*. 2011;31(28):10206-10214.
13. Kim J, Lee HM, Kim WJ, Park HJ, Kim SW, Moon DH. Neural correlates of pre-performance routines in expert and novice archers. *Neurosci Lett*. 2008;445(3):236-241.
14. Wright MJ, Jackson RC. Brain regions concerneed with perceptual skills in tennis: a fMRI study. *Int J Psychophysiol*. 2007;63(2):214-220.