BORNEO SCIENCE 9: 19-29, June, 2001

SPATIAL ABILITY TESTS AS A PREDICTOR OF TRAINABILITY IN ARTHROSCOPIC SURGERY

Murnizam Hj. Halik

School of Psychology & Social Work, Universiti Malaysia Sabah,
Locked Bag 2073, 88999 Kota Kinabalu, Sabah.

ABSTRACT. *The purpose of this project was to investigate whether spatial ability tests can act as the predictor of trainability in arthroscopic surgery. The initial part was administering the French Ekstrom Factor Reference Kit (1976) spatial ability tests (n=109). Then some subjects (n=14) were trained in diagnostic arthroscopic surgery procedure. Their performance on selected trials were recorded, and scored by non-medical appraisers (n=8). The outcome showed that subjects' performance in most spatial ability tests correlates with their overall performance in arthroscopic training. Highly significant concordance levels were also found among appraiser. The results highlighted the wisdom of applying spatial ability tests for the selection of future arthroscopic trainee surgeons.*

KEYWORDS. Spatial ability tests, arthroscopic surgery, performance appraisals.

INTRODUCTION

This paper reports on a project to establish predictors of novice trainability in arthroscopic surgery from a range of spatial ability tests. The research investigated whether there is a relationship between individuals' spatial ability and their ability to pick up the skills required to perform arthroscopy. Arthroscopy is a form of *Minimally Invasive Surgery* (MIS). Previously, surgeons had to perform open surgery in order to diagnose the problem and made an instant judgement whether a full-scale operation is needed. The result of surgery at this stage may lead to complications such as infections or damage to the healthy tissues around the affected area. The aim of MIS is to minimise the trauma but still achieve a satisfactory therapeutic result.

The techniques of MIS provide the better option, where an arthroscope and a probe are inserted into the joint through a tiny incision (about 2 cm) called a portal to diagnose the joint. Apart from the probe, other surgical instruments can be inserted into the knee, and these are guided by the surgeon from outside of the knee joints. The video images from the arthroscope are then viewed by the surgeons on a monitor in front of them. This will allow the surgeon to have a unique view of the inside of the joints. The technique of MIS also allows the possibility of performing certain types of surgery without resorting to open surgery. There are also a number of procedures that are difficult to perform in open surgery but can be easily performed in arthroscopy.

For the past two decades, arthroscopic surgery has gained rapid popularity among orthopaedic surgeons and patients alike. Since Professor Takagi of Japan recorded the first arthroscopic procedure in 1918 (using the tools of the time), surgeons have performed thousands of arthroscopic surgery around the world. Spiegler (1995) reported that in 1985, orthopaedic surgeons treated 3.6 million cases and the number jumped to 4.1 million in 1992. McGinty, Johnson, Jackson, McBryde and Goodfellow (1992) estimated around 1.4 million arthroscopic procedures being performed in United States in 1990.

The popular use of arthroscopy has led to several problems. The sophisticated equipment improves the quality of arthroscopy surgery in general but the number of complication is still high. Indeed, studies by Sherman, Fox, Snyder, Pizzo, Friedman, Ferkel, Nuys & Lawley (1986) showed that there was an 8.2 per cent complication rate reported out of 2,640 arthroscopic procedure on the knee. A retrospective survey of the results of 118,590 knee arthroscopies performed by the Arthroscopy Association of North America revealed a total of 930 complications (DeLee, 1985). Bamford, Noble and Davies (1992) pointed out that most, but not all, complications are avoidable.

Arthroscopy differs from most types of conventional surgery due to its unique handling technique. Common problems faced by trainee surgeons are in the area of triangulation of the instruments, navigation inside of the knee cavities and handling of the arthroscope and its instruments. A possible way to achieve a satisfactory standard of arthroscopic surgery is through correct methods of training. Some trainees can learn the skills faster and other takes a longer period to acquire the skills. Clinical observation suggests that some surgeons who have good operative ability have difficulty performing arthroscopy (Barrett, Green & Copeland 1991). Grechenig, Fellinger, Fankhauser & Weiglin (1997) stressed that arthroscopic operations is bound by experience and manual skill and practise is a very essential part of a learning process in surgical fields. Other researchers have also stressed the need for training in arthroscopic surgery (Barrett et al, 1991; Milankov, Jovanovic, Milicic, Savic, Stankovic, Kecojevic & Vukasav, 2000)

The usual practical method of training for arthroscopic surgeon was done using model knees, animals, simple simulators and virtual reality. Chung and Sackier (1998) raised the question whether training sessions are effective in improving surgeons skill in laproscopic surgery (another form of MIS). They believed that an objective evaluation is vital in judging surgeon's performance in training. Murdoch, Bainbridge, Fisher and Webster (1994) believed that objective assessment could be achieved if performance is rated on a very specific set of criteria related to microsurgical tasks. Although there is no doubt that training and objective evaluation is necessary, the best method of doing so has not been established yet.

The method of selection for surgical trainees, as summed up by Gough, Holdsworth, Bell, Keeman, Lagaay, Van De Loo and Droog (1988), depends on the applicant's academic record, the impression given at the interview, the reference submitted and a combination of opportunity and luck.

There is no doubt that with high academic qualification, trainees have little difficulties in acquiring cognitive knowledge about surgery but their surgical skills will be remain untested.

Overall, researchers agree that some skills are important in handling tools and performing a MIS. Investigations by Gibbons R. D, Baker and Skinner (1986) have shown that surgical skills do appear to be correlated significantly with some types of spatial ability, especially flexibility of closure and surface development. Chung and Sackier (1998) mentioned that fine motor skills, depth perception and hand-eye coordination are needed to manipulate tools. Derosis, Bothwell, Sigman & Fried (1998) added ambidexterity as a skill vital in surgery. The same research found a highly significant correlation ($r = 0.72$, $p < 0.001$) between training (a series of motor skills exercise) and surgeons' subsequent performance.

Interest in applying aptitude tests for arthroscopic and endoscopic trainee surgeon had been expressed by Barrett et al. (1991). A computer-based aptitude test had been designed to assess the extra skills required of orthopaedic trainees performing arthroscopy. The results indicated that more than a simple psychomotor skill is needed to perform arthroscopy. This aptitude test, according to them, might be useful to identify junior surgeons who need extra instruction. In some areas that have similar surgical skills as arthroscopy, such as dentistry, aptitude tests had been well researched. According to Murdoch et al. (1994), the majority of those studies suggest that there is a significant relationship between aptitude and dexterity test and performance in dental school. Studies by Murdoch et al. (1994) showed that performance on microsurgical tasks undertaken by trainees to assess their surgical skill have significant correlation with manual dexterity and spatial ability. The research done by Gibbons et. al. (1986), Barrett et al. (1991) and Murdoch et al. (1994) provide support for this project on spatial ability test in the field of arthroscopy.

RESEARCH DESIGN

This project was a partial replication of Gibbons et. al. (1986) study. The researcher developed scoring system guidelines and a demonstration video of a routine diagnostic arthroscopy procedure with the advice and help of an orthopaedic surgeon. The scoring system measured subjects' abilities in performing triangulation, navigation and instrument handling. The project centered on three separate experimental procedures. These were:

1. Paper and pencil tests.
2. Arthroscopy training and recording
3. Performance appraisals.

Each of these procedure is detailed in turn.

Experiment 1: Paper and pencil tests

Subjects

One hundred and nine subjects took part in the paper and pencil test. There were 55 male and 54 female subjects. None of them have any prior experience with arthroscopy or medicine.

The mean age was 22.68 years old, ranging from 19 to 30 years old.

Procedure

The experiment was divided into several sessions of fifteen to seventeen subjects. Seats were arranged in classroom style and booklets were distributed together with two pencils and an eraser. Subjects were briefed by the test administrator about the testing, the confidentiality of their results and the duration of time to complete the testing.

Test Instruments

Six factors from The French Ekstrom Factor Reference Kit (1976) were chosen to measure subjects' spatial ability and collated into a test batteries. The list is shown in Table 1.

Table 1. List of factors, description and types of test used.

	Factor	Description	Type of tests
1	Flexibility of Closure	The ability to hold a given visual perception or configuration in mind so as to disembed it from other well defined perceptual material	A. Hidden Figures Test B. Hidden Patterns Test C. Copying Test
2	Speed of Closure	The ability to unite an apparently disparate perceptual field into a single concept	A. Gestalt Completion Test B. Concealed Words Test C. Snowy Pictures
3	Visual Memory	The ability to remember the configuration, location and orientation of figural material	A. Shape Memory Test B. Building Memory Test C. Map Memory Test
4	Spatial Orientation	The ability to perceive spatial patterns or to maintain orientation with respect to objects in space	A. Card Rotations Test B. Cube Comparisons Test
5	Spatial Scanning	Speed in exploring visually a wide or complicated visual field	A. Maze Tracing Speed Test B. Choosing a Path C. Map Planning Test
6	Visualisation	The ability to manipulate or transform the image of spatial patterns into other arrangements	A. Form Board Test B. Paper Folding Test C. Surface Development Test

Experiment 2: Arthroscopy training and scoring

Subjects

From the list of 109 subjects who participated in the spatial ability test tests, fourteen took part in this experiment, ten males and four female subjects. Their mean age was 26.85 years old. There were no criteria for choosing the subjects except the fact that they were available during the period of training. None of them had any experience in arthroscopy or medicine.

Procedure

Training took place in a hospital room at times to suit the subjects one by one. The estimated duration of each training session was around three hours. The subjects were asked to take a seat in front of a VCR monitor. The administrator explained the experimental conditions to the subjects who were given a chance to ask any questions. After the subjects indicated that they were ready, they were shown a demonstration video of an actual procedure performed by a specialist arthrosopic surgeon. A series of still pictures of the model knee structure extracted from the video were shown and subjects were given five minutes to study them. One of the experimenters then showed the subject a 'live' demonstration of diagnostic arthroscopy and he or she was asked to perform nine trials of the diagnostic arthroscopy. There were no practice trials. The subjects were given instruction orally from the first until the ninth trial. The trainer gave instruction strictly regarding the 'route' of diagnostic arthroscopy and on what action the subjects were supposed to take when they identified the structure. The trainer stopped the sessions at the end of the ninth trial.

Materials and Apparatus

The experiment used a 30 degree arthroscope, a probe and monitor. The model knee used was a *Hillway* plastic knee. There were a set of six still colour pictures of knee structures and a demonstration video. Finally, the experiment used a VCR with monitor and nine blank tapes for each session to record subjects' performance.

Experiment 3: Performance appraisals

Subjects.

There were eight subjects who participated in the scoring of the trainee. None of them had experience in arthroscopy or medicine. Three of them were females. These subjects had not taken part in the spatial ability test tests. They were paid for taking part in this session. The mean age was 27.85 years old, ranging from 24 to 34 years old.

Procedure

The administrator explained to the scorers individually about arthroscopy and the experiments. Each scorer was given a chance to ask any questions regarding the experiment and was then shown the demonstration video. They then viewed a series of still pictures of the model knee structure extracted from the video and are given five minutes to study them. After viewing the stills, the scorer was given guidelines on how to score the recorded procedure and the diagnostic 'route' together with the scoring sheets. Two opportunities to do a practice scoring of randomly selected videos of subjects' performance were provided. The administrator then showed the recorded videos (trial 1, 5 and 9) of all the fourteen subjects. These trials were selected as the researcher sought to evaluate the beginning, middle and final part of subjects' performance in the training sessions.

Materials And Apparatus

The project used a remote contol television monitor and VCR, the recorded video tapes of subjects' performance and set of six colour stills of the inside knee. Scoring sheets, an example of the diagnostic *route*, guidelines for scoring, two pencils and an eraser were provided.

RESULTS

The data gathered from paper and pencil tests were the subjects' mark in each of the tests. In the arthroscopic training, data was gathered from subjects' scoring sheet. The relationship between subjects' performance in paper and pencil tests and in training are shown in Table 2.

Table 2. Correlation between paper and pencil tests and arthroscopic training.

No	Type Of Tests	Trial 1	Trial 5	Trial 9	Overall
1.A	Hidden Figure Test	.79**	.57	.51	.67*
1.B	Hidden Patterns Test	.48	.12	.23	.30
1.C	Copying Test	.02	-.35	-.24	-.19
2.A	Gestalt Completion Test	.05	.19	.02	.09
2.B	Concealed Words Test	.68*	.74*	.78**	.77**
2.C	Snowy Pictures	.23	21	.17	.22
3.A	Shape Memory Test	.30	.12	.40	.30
3.B	Building Memory Test	.57	.66*	.54	.62*
3.C	Map Memory Test	.20	.39	.08	.23
4.A	Card Rotation Test	.62*	.39	.56	.56
4.B	Cube Comparison Test	.78**	.63*	.63*	.72*
5.A	Maze Tracing Speed Test	.70*	.64*	.67*	.71*
5.B	Choosing A Path	.63*	.58	.76**	.70*
5.C	Map Planning Test	.74*	.66*	.69*	.74*
6.A	Form Board Test	.50	.46	.47	.51
6.B	Paper Folding Test	.85**	.58	.66*	.74*
6.C	Surface Development Test	.78**	.70*	.77**	.79**

Significant at $*p<0.05$, $**p<0.01$

The two-tailed Pearson correlation as represented in Table 2 showed that subjects' performance in some of the paper and pencil tests had a significant correlation ($p<0.01$ & $p<0.05$) with their performance in trials one, five and nine. Three factors showed the most significant correlations across three trials. Those factors were spatial orientation, spatial scanning and visualisation. There were no significant correlations in tests 1.B, 1.C, 2.A, 2.C, 3.A and 3.C

The subjects' overall scores obtained in trial one, five and nine of the training are summarised in Table 3.

Table 3. Means, standard deviations and maximum score for subjects' scores in training (n=14).

Trial	Mean	Sd	Max score
1	108.68	28.48	210
5	128.53	27.4	210
9	127.56	28.16	210

The table indicates that there are differences in subjects performance. The mean score for trial one was 108.68 (s.d.=28.48) while in trial five and nine, the mean scores was 128.53 (s.d=27.4) and 127.56 (s.d.=28.16) respectively. As can be seen, the subjects' performance in trial five increased compared to trial one but their performance in trial nine was lower than in trial five.

In order to find out whether there was a significant improvement in subjects' performance from trial one to trials five and nine, a one-way ANOVA was employed (Table 4).

Table 4. Analysis of Variance of subjects' performance in arthroscopic training.

Source	df	Sum of squares	Mean squares	F	p-Value
Between groups	2	28043.4	14021.7	17.85	0.0001
Within groups	333	261491	785.25		
Total	335	289534			

The ANOVA showed that there was a significant difference ($p=0.0001$) in subjects' performance. A *Tukey* test was performed to show where the significant differences lay. The result shows that the performance in trial five and nine were rated significantly higher than trial one indicating that subjects improved their performance in trial five and nine ($p<0.05$). However, there was no significant difference in the rating of subjects performance from trial five to trial nine.

Table 5 displays an inter-correlation of the scores obtained by the subjects. The table shows a significant correlation ($p<0.01$) between the performance on the three trials.

Table 5. Inter-correlation of subjects' scores in training.

	Trial 1	Trial 5	Trial 9
Trial 1	-	.83**	.86**
Trial 5		-	.82**
Trial 9			-

Significant at **p<0.01

The scorers' concordance level is displayed in Table 6. Although it is a non-parametric test, Kendall's coefficient of concordance was used because it is the most straightforward measure of inter-rater consistency, and because it can be converted into an average. For Spearman rank correlation, it does offer a high degree of reliability.

Table 6. Inter-scorers concordance in scoring training performance (n=8).

	Kendall W	Spearman R	Df
Trial 1	0.8355	.81**	13
Trial 5	0.7274	.68*	13
Trial 9	0.8101	.78**	13

Significant at ** p<0.01 *p<0.05

These results demonstrate a significant level of concordance between the scorers ratings for subjects performance in trial one (W=.8355 converted to *rs*=.81, N=14, p<0.01), trial five (W=.7274, *rs*=.68, N=14, p<0.05) and trial nine (W=.8101, *rs*=.78, N=14, p<0.01).

DISCUSSION

This research revealed two important findings arising from the data shown in the results. The first finding answered the question addressed in the first part of the study, that is whether spatial ability tests, as demonstrated in paper and pencil tests and computer tracking task, could be predictors of trainability in arthroscopic surgery.

The results indicated that from the seventeen types of test of the six factors in French-Ekstrom Factor Reference Kit (1976), both the surface development test and the concealed words test correlated highly (*p*<.01) with subjects' performance in overall trials. The only negative correlation (*p*=ns) was found in the Copying Test. All the tests in Spatial Scanning correlated significantly (*p*<.05) with subjects overall performance in the three trials. This result seems to imply that individual performance in nine of the paper and pencil tests (such as surface development test, concealed words test, paper folding test, map planning tests, cube comparison test, maze tracing test, path finding test, hidden figures test and building memory test) may be able to indicate their performance in arthroscopic training.

The second major finding arose from the issue of inter-scorer reliability (Table 7). Using the Kendall's co-efficient of concordance, converted into an average Spearman rank correlation, the results demonstrated a significant level of concordance between the scorers ratings for subjects performance in trial one, trial five and trial nine. The results shows that in general, all scorers were consistent in the scoring the subjects performance.

Besides these two findings, another point of interest is in the trend of correlation between the spatial ability tests and training as shown in Table 2. It appears to have a higher correlation in the first trial compared to the fifth trial. In the ninth trail the correlation is higher than trial five but lower than trial one. One possible explanation for this was that those who performed poorly in spatial ability tests also performed poorly in the first trial of training. These subject might have improve their performance in trial five thus explains the reason for the low correlation. The increase in correlation for trial nine may be the results of overall low performance (as shown in the mean of subjects score in Table 3) .

The results from this study appear to imply that some spatial ability tests can be the predictors of trainability in arthroscopic surgery. The first finding may have answered a few questions raised by some of the researchers as mentioned in the first part of this study. The findings of this research showed a similar pattern with the studies by Gibbons et al. (1986) since this research also showed that flexibility of closure and surface developments have a significant correlation with arthroscopic training. Clearly some tests would be good predictors of future trainee-surgeon performance. Through proper selection, avoidable instances such as fracture of instruments and complications may well be reduced. It follows that an advantage of using these tests is that the time and money invested in the trainee could be minimised. Furthermore, surgeons and psychologists could invent simulators and instruments that would enhance trainee skills and technique before using the real equipment. This could avoid accidental breaking of the costly arthroscopic instruments during training.

The results from the second finding strongly imply that scorers were consistent in giving scores. Questions arise whether expert surgeons should evaluate the performance of trainee surgeons given their different background, technique and personal bias. However, based on this research, scorers who do not have any background in arthroscopy or medicine seem to be able to evaluate subjects' performance consistantly. This research indicated that a well trained non-surgeon could evaluate the performance of potential surgeons. This proposition might be controversial, but at least it would allows senior surgeons to concentrate on their work rather than spending countless hours evaluating their junior partners. The researcher believes that such idea merits a trial run or at least a thorough discussion.

The results in Table 3 showed that subjects' performance improved from trial one to trial five but there was no significant improvement in trial nine compared to trial five. Some human factors, such as complacency, fatigue or over-confidence might contribute to the almost stagnant overall score in trial nine. However, such variables were beyond the experimenter's control.

The author recommends that future researchers in this topic could make some improvements on the short comings of this research. The two areas of this research that really need attention are in the method of training and the design of scoring system. It may be possible that better results can be achieved if researchers made improvements by redesigning the training and scoring method.

CONCLUSION

The implications of this research might benefit people who are involved in selecting and training future arthroscopic surgeons. This project indicated that it might be possible to anticipate individuals future performance in arthroscopic surgery through their performance in spatial ability tests. Efforts should be undertaken to design an objective performance appraisal. This will allow a well-trained non-surgeon to evaluate the surgeons' performance. It would also be valuable to know whether other aspects of human ability can also act as predictors for trainability in arthroscopic surgery.

REFERENCES

Bamford, D., Noble, J., & Davies, D. R. A. 1992. Avoidable complication of arthroscopic surgery. *The Journal of Joint and Bone Surgery (Br)*. **Vol: 74** (supp II): 328.

Barrett, D. S., Green, R. G., & Copeland, S. A. 1991. Arthroscopic and endoscopic skills: A method of assessment. *Annals of the Royal College of Surgeon of England*. **73**: 100-104.

Chung, J. Y., & Sackier, J.M. 1998. A method of objectively evaluating improvements in laproscopic skills. *Surg Endosc*. **12**: 1111-1116.

De Lee, J. C. 1985. Complications of arthroscopy and arthroscopic surgery: A result of a national survey. *Arthroscopy*. **1**: 214-220.

Derosis, A. M., Bothwell, J., Sigman, H. H., & Fried, G. M. 1998. The effect of practice on performance in a laproscopic simulator. *Surg Endosc*. **12**: 1117-1120.

Gibbons, R. D., Baker, R. J., & Skinner, D. B. 1986. Field articulation testing: A predictor of technical skills in surgical residents. *Journal Surgical Residents*. **41**: 53-57.

Gough, M. H., Holdsworth, R., Bell, J. A., Keeman, J. N., Lagaay, M., Loo, V. D., & Droog, A. 1988. Personality assessment techniques and aptitude testing as aids to the selection of surgical trainees: A symposium. *Annals of the Royal College of Surgeon of England*. **70**: 265-279.

Grechenig, M., Fellinger, M., Fankhauser, F. & Weiglin, A. H. 1997. The graz learning and training model for arthroscopic surgery. *Surg Radiol Anat*. **21**: 347-350.

McGinty, J. B., Johnson, L. L., Jackson, R. W., McBryde, A. M., & Goodfellow, J. W. 1992. Uses and abuses of arthroscopy: A symposium. *The Journal of Joint and Bone Surgery.* **74-A**: 1563-1577.

Milankov, M., Jovanovic, A., Milicic, A., Savic, D., Stankovic, M., Kecojevic V., & Vukasav, B. 2000. Arthroscopy of the knee-"surgery without complications". *Med Pregl.* **53(3-4)**: 187-92.

Miller, R. A., & Wickham, J. E. A. 1990. *The endo chair*. Presentation to the european urological association. Amsterdam.

Murdoch, J. R., Bainbridge, L. C., Fisher, S. G., & Webster, H. C. 1994. Can a simple test of visual-motor skill predict the performance of microsurgeons? *Journal of Royal College of Surgeon of Edinburgh.* **39**: 150-152.

Sherman, O. H, Fox, J. M, Snyder S. J, Pizzo, W. D, Friedman, M. J, Ferkel, R. D, Nuys, V., & Lawley, M. J. 1986. Arthroscopy- "No-Problem Surgery." An Analysis of Complications in 2,674 cases. *The Journal of Joint and Bone Surgery*. **68-A**: 256-265.

Spiegler, M. (1995). Knees have no warranties. *American Demographics.* **17**: 20-23.