The predictive value of aptitude assessment in laparoscopic surgery: a meta-analysis

Kelvin H Kramp,[1] Marc J van Det,[2] Christiaan Hoff,[1] Nic J G M Veeger,[3,4] Henk O ten Cate Hoedemaker[5] & Jean-Pierre E N Pierie[1,6]

CONTEXT Current methods of assessing candidates for medical specialties that involve laparoscopic skills suffer from a lack of instruments to assess the ability to work in a minimally invasive surgery environment.

OBJECTIVES A meta-analysis was conducted to investigate whether aptitude assessment can be used to predict variability in the acquisition and performance of laparoscopic skills.

METHODS PubMed, PsycINFO and Google Scholar were searched to November 2014 for published and unpublished studies reporting the measurement of a form of aptitude for laparoscopic skills. The quality of studies was assessed with QUADAS-2. Summary correlations were calculated using a random-effects model.

RESULTS Thirty-four studies were found to be eligible for inclusion; six of these studies used an operating room performance measurement. Laparoscopic skills correlated significantly with visual–spatial ability ($r = 0.32$, 95% confidence interval [CI] 0.25–0.39; $p < 0.001$), perceptual ability ($r = 0.31$, 95% CI 0.22–0.39; $p < 0.001$), psychomotor ability ($r = 0.26$, 95% CI 0.10–0.40; $p = 0.003$) and simulator-based assessment of aptitude ($r = 0.64$, 95% CI 0.52–0.73; $p < 0.001$). Three-dimensional dynamic visual–spatial ability showed a significantly higher correlation than intrinsic static visual–spatial ability ($p = 0.024$).

CONCLUSIONS In general, aptitude assessments are associated with laparoscopic skill level. Simulator-based assessment of aptitude appears to have the potential to represent a job sample and to enable the assessment of all forms of aptitude for laparoscopic surgery at once. A laparoscopy aptitude test can be a valuable additional tool in the assessment of candidates for medical specialties that require laparoscopic skills.

Medical Education 2016: 50: 409–427
doi:10.1111/medu.12945

Discuss ideas arising from the article at www.mededuc.com discuss.

---

[1] Department of Surgery, Medical Centre Leeuwarden, Leeuwarden, The Netherlands
[2] Department of Surgery, Hospital Group Twente, Almelo, The Netherlands
[3] Department of Epidemiology, University Medical Centre Groningen, Groningen, The Netherlands
[4] Department of Epidemiology, Leeuwarden Medical Center, Leeuwarden, the Netherlands
[5] Department of Surgery, University Medical Center Groningen, Groningen, the Netherlands
[6] Post Graduate School of Medicine, University Medical Center Groningen and University of Groningen, Groningen, the Netherlands

Correspondence: Kelvin H Kramp, Department of Surgery, Medical Centre Leeuwarden, Leeuwarden Henri Dunantweg 2, 8934 AD Leeuwarden, The Netherlands. Tel: (058) 286 60 70; E-mail: k.h.kramp@gmail.com

INTRODUCTION

The assessment of candidates for training in a medical discipline is a critical component of selection procedures in medical education. However, although surgical techniques are becoming increasingly difficult to master, no scientific methods are currently used to evaluate the potential to acquire these surgical skills. This is especially worrisome in the field of laparoscopic surgery, which is coming to represent the mainstay method in an increasing list of procedures in abdominal surgery, gynaecology and urology. During laparoscopic surgery, there is no direct visualisation of the operative field or direct contact with intra-abdominal organs. Consequently, laparoscopic surgery requires a set of skills that differs from those used in conventional surgery. The majority of trainees overcome the ergonomic difficulties associated with laparoscopic surgery during laparoscopic skills training, but research has raised concerns about large individual differences during and after training that may be dependent on aptitude. For instance, the studies of both Schijven *et al.*[1] and Grantcharov and Funch-Jensen[2] were able to distinguish four groups that showed different patterns of improvement during laparoscopic skills training: (i) those who were proficient at the beginning; (ii) those who achieved proficiency through training; (iii) those who showed improvement without reaching proficiency, and (iv) those who showed no improvement. A third study by Bosker *et al.*[3] showed that one of 13 participants (7.7%) seemed to have difficulties in learning to perform a laparoscopic sigmoid resection, although there were no factors that could have caused an increase in the difficulty of procedures performed.

There is growing evidence in the literature that some of these differences in ability to learn and perform laparoscopic surgery can be explained in part by aptitude. Aptitude for minimally invasive surgery (MIS) is often divided into three abilities that are generally accepted to be of innate nature: visual–spatial ability; perceptual ability, and psychomotor ability.[4] Visual–spatial ability refers to the ability to mentally visualise or manipulate objects; perceptual ability refers to the ability to interpret two-dimensional representations of three-dimensional objects, and psychomotor ability refers to the ability to perform motor movements that involve eye–hand coordination, bimanual dexterity and short reaction times. The question of whether testing of these abilities can be used in the selection of trainees is currently a topic of vehement debate.[5–9] Although there is evidence that the ability to learn and perform laparoscopic surgery can be assessed with measurements of these aptitudes, there are also reports that contradict such a correlation.[10–14] To date, reviews that have aimed to reach an univocal conclusion on this topic have been mainly descriptive in nature, have lacked a quantitative analysis or have investigated a broad spectrum of surgical skills.[4,8,14] Therefore, a meta-analysis was conducted to evaluate whether aptitude assessments can be used to predict the ability to acquire and perform laparoscopic skills and to quantify how much of difference between individuals can be predicted by aptitude assessment.

METHODS

Search strategy

A systematic literature search in PubMed, PsycINFO and Google Scholar was conducted in November 2014 to find studies that measured laparoscopic skills and aptitude (Fig. 1). The query used to identify the available literature in PubMed was '((Space Perception[MeSH]) OR (Visual Perception[MeSH]) OR (Psychomotor Performance[MeSH]) OR (Aptitude[Mesh])) AND (Laparoscopy[MeSH])'. The query used in PsycINFO was 'Laparoscopy OR Laparoscopic' and was limited to studies in human subjects aged > 18 years. The queries used in Google Scholar were the word 'laparoscopy' combined with the phrase 'visual–spatial ability', and the word 'laparoscopy' combined with the phrase 'psychomotor ability'. To identify unpublished literature, dissertation and thesis databases and conference abstract books were hand-searched using the keywords 'visual–spatial', 'spatial ability', 'visual perception', 'spatial perception', 'psychomotor' and 'aptitude' for additional relevant titles. Finally, the reference lists of the studies retrieved were scanned for additional relevant studies and the names of key authors were used as search terms in PubMed and Google Scholar. The abstracts of studies with relevant titles were reviewed; only studies that assessed aptitude among subjects and measured laparoscopic skill level were considered eligible for inclusion.

Data extraction and quality assessment

From the final list of studies, information was extracted on: the specific aptitude tests used; the reported correlations between the results of aptitude tests and the performance level of participants; numbers of participants; characteristics of partici-

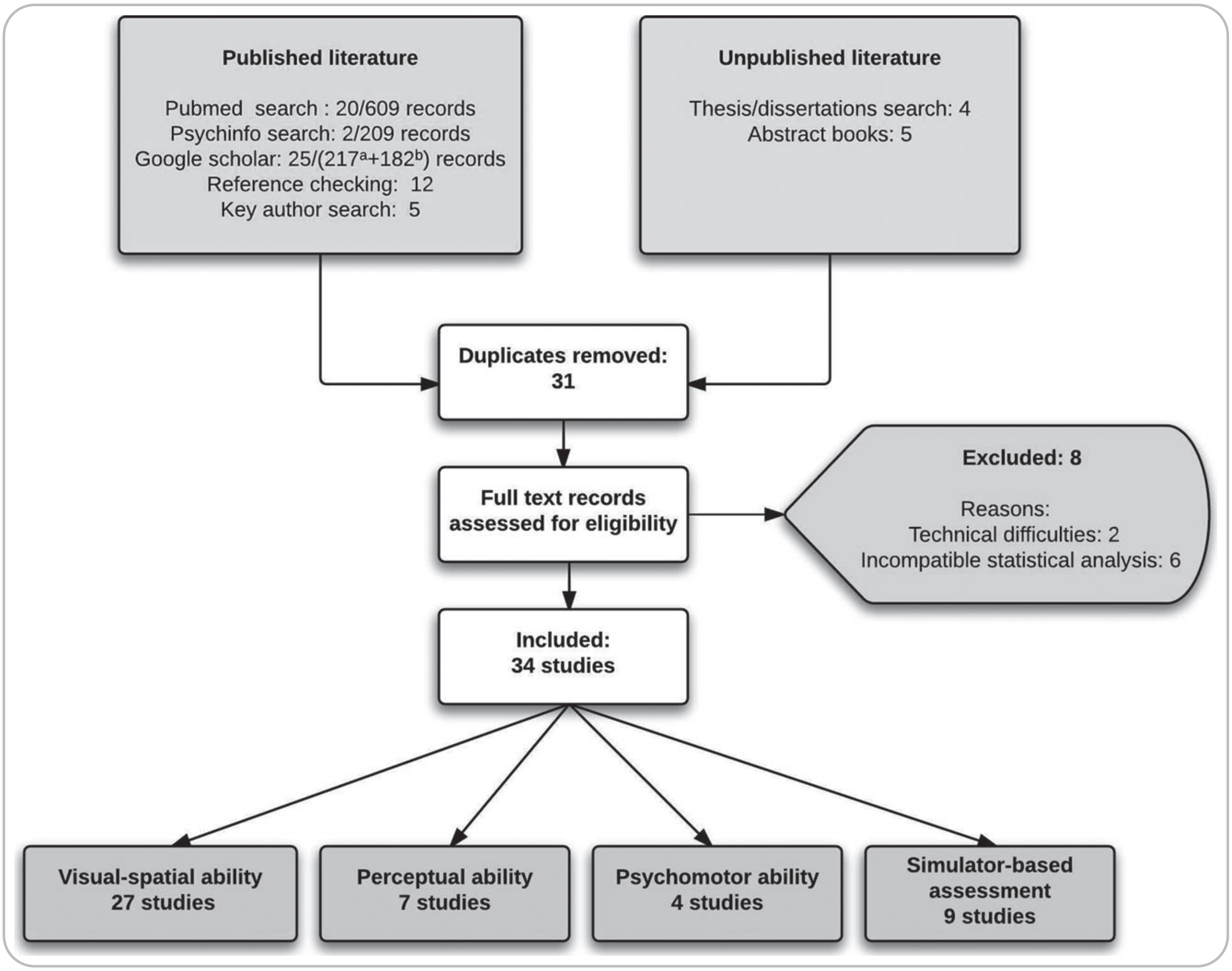

Figure 1 Flow diagram of search strategy. The Google Scholar search was conducted with the word 'laparoscopy' combined with the phrase 'visual spatial ability'[a] and 'laparoscopy' combined with the phrase 'psychomotor ability'[b]. Abstract books were sourced from the European Association for Endoscopic Surgery, the Society of American Gastrointestinal and Endoscopic Surgeons, and the Association of Laparoscopic Surgeons of Great Britain and Ireland. Theses and dissertations were found by using key author names as search terms in Google Scholar

pants; methods used to measure laparoscopic skills; parts of the learning curve in which laparoscopic skills were measured; publication status, and country of origin.

The quality of the included studies was assessed using the modified form of the QUADAS-2 (QUality Assessment of Diagnostic Accuracy Studies).[16] QUADAS-2 is a quality assessment system for diagnostic studies. It involves four domains of assessment that should be tailored to fit the study in question: subject selection; index test; reference test, and flow and timing. The first three of these domains are evaluated based on aspects related to the risk for bias and concerns related to the applicability of the study results, whereas the domain of flow and timing is only assessed based on aspects related to risk for bias. Because the goal of this meta-analysis was to investigate the predictive validity of aptitude tests, which can be seen as index tests, the QUADAS-2 was considered to be a suitable quality assessment tool.

Meta-analysis

The collected correlations were coded such that positive correlations indicate a proportional relationship and negative correlations an inverse proportional relationship between aptitude test scores and laparoscopic performance metrics. For studies that did not report the relationship between aptitude scores and laparoscopic skills with Pearson or Spearman correlation coefficients, the reported results were converted into correlation coefficients using the formulae shown in Appendix S1. Fisher *Z–r* transformation was used to translate the Pearson, Spearman and converted correlations into effect sizes.[17]

If the actual value of a non-significant correlation was not reported, the lead authors were contacted for additional data. If data could not be obtained from the authors, we used either of two different strategies to address the missing correlation: (i) there is no relationship between the two variables, or (ii) the sample size is too small to achieve a level of significance. The first option can mathematically be considered as a correlation of 0. Non-significant correlations with an unreported value were therefore coded as $r = 0$ in dataset DS_0. The second option was evaluated by substituting the maximum achievable correlation coefficient (the critical value of the Pearson correlation coefficient based on the number of participants) for the unreported non-significant correlations in dataset, DataSet with r = Critical Value (DS_{cv}).

Some studies used multiple groups of participants with different characteristics (e.g. medical students, trainees, consultants, etc.). In these cases, the correlations for each group of participants were calculated into mean 'participant group' effect sizes. This was done by computing the mean value of the correlations between aptitude test outcomes and laparoscopic skill level reported in the study for a specific group of participants.

A correction was applied to the variance to compensate for the partial independence between correlations because of commonalities in the study setting in which the different correlations were measured within a participant group.[17] As no correlations that could be used to correct the partial interdependence between the reported correlations were identified in the literature, $r_x = 0.5$ was used as a compromise between the two extremes (Formula 6, Appendix S1).

After estimating mean participant group effect sizes and their variance, summary correlations were calculated for the different forms of aptitude. Because a high degree of variety (heterogeneity) in methodology among the included studies was observed, the random-effects model was used to calculate the mean correlations. Heterogeneity tests (Cochrane *Q*) were performed to assess the variety among studies and a p-value of < 0.10 was considered to indicate differences of statistical significance. The I^2 statistic was calculated to estimate the percentage of variance that can be attributed to variation among studies. Heterogeneity was defined by percentage as: low (25%); medium (50%), and high (75%).[18]

Visual–spatial ability moderator analysis

When statistically significant heterogeneity is observed, a moderator analysis can be performed to investigate whether variations in results among studies are caused by differences in study methodology. In this meta-analysis, a moderator analysis was conducted to evaluate the different factors that might have influenced the relationship between visual–spatial ability and laparoscopic skills. Moderators were set as: (i) a recently published 2 × 2 classification of visual–spatial ability; (ii) method of measuring laparoscopic skills; (iii) participant characteristics, and (iv) components of the learning curve. The random-effects model was used to calculate the summary correlation for each subgroup within the moderator analysis. A pooled τ^2 was used to estimate the imprecision of subgroup summary correlations and subgroups were compared with heterogeneity *Q* according to the procedure described by Borenstein *et al.*[17] Heterogeneity *Q*-values with p-values of < 0.05 (two-tailed) were considered statistically significant in the moderator analysis.

2 × 2 classification of visual–spatial ability

It is currently accepted that visual–spatial ability is not a uniform ability.[19,20] Uttal *et al.*[19] recently proposed a classification of visual–spatial ability on the basis of two fundamental properties of visual–spatial ability tests. The first distinction is whether a visual–spatial ability test utilises intrinsic or extrinsic information. Intrinsic information contains the characteristics that define an object. Extrinsic information is information that comes from relationships between groups of objects or the relationship of an object to a framework. The second distinction refers to whether a visual–spatial ability test requires a static or dynamic mental process to complete. Static visual–spatial ability tests contain fixed objects, whereas dynamic visual–spatial ability tests require the mental visualisation of a spatial change in an object or perspective. These two properties of visual–spatial ability tests can be used to classify them in four categories: intrinsic static; intrinsic dynamic; extrinsic static, and extrinsic dynamic. Uttal *et al.*[19] suggested that a visual–spatial ability test score is an indication of visual–spatial ability in one of the four categories that cannot plainly be generalised to other categories. For example, artists seem to use their visual object ability (intrinsic static) in their profession, whereas engineers appear to depend more on visual–spatial translational ability (intrinsic dynamic and extrinsic dynamic).[21,22]

Other moderators

A further three moderators were used to evaluate heterogeneity among studies in visual–spatial ability: measurement method; participant characteristics, and part of the learning curve in which the performance level was measured.

The association between visual–spatial ability and laparoscopic skills may differ on the basis of the method used to measure skill level. To investigate whether the method of measuring performance influences the correlation with visual–spatial ability, correlations were categorised according to whether they were derived from assessments using: (i) a video trainer; (ii) a virtual reality simulator; (iii) laparoscopic camera navigation, or (iv) laparoscopic surgery on an animal or human.

Whether the study sample represents the characteristics of the population of interest can be of key importance in the validation of prognostic tools. In the moderator *participant characteristics*, the correlations were therefore categorised according to whether they referred to: (i) non-medical students; (ii) medical students; (iii) novice trainees, or (iv) trainees who had received training in laparoscopy and consultants.

A typical learning curve has been documented in laparoscopic skills training on video trainers, virtual reality simulators and in the operating room (OR).[23,24] The learning curve starts with a baseline performance. After the first performance there are improvements with each repetition. The improvements decrease in size as experience accrues according to the learning rate. As the improvements become increasingly smaller with each repetition, a learning plateau is reached.[25] The strength of the association between visual–spatial ability test scores and laparoscopic skills may differ between learning phases. In the moderator *learning curve*, correlations were therefore categorised according to whether they referred to: (i) baseline performance; (ii) learning rate, or (iii) learning plateau.

Publication bias

To visually assess the sample of studies for publication bias, a funnel plot was created with 95% pseudo confidence intervals (CIs). Quantitative assessment of publication bias was performed with the Egger and Begg tests with p-values of < 0.05 (two-tailed) considered as indicative of significance.[26,27]

RESULTS

Search results

The results of the literature search are shown in Fig. 1. Eight studies were excluded because they used a methodology or statistical analysis incompatible with the research question of the meta-analysis or because of technical difficulties during aptitude testing (see Appendix S2 online for details).[28–35] Some studies evaluated the correlation between a simulator-based assessment and a subsequent performance on a simulator or performance in the OR. These studies were analysed separately from those using other aptitude measurements. In total, 34 studies were eligible for inclusion on the basis of the present criteria and could be used for further analysis; six of these studies included an OR performance measurement. A limited series of studies reported cut-off scores for the classification of candidates and their corresponding sensitivity and specificity.[1,36,37]

Quality assessment

The quality of the 34 included studies is shown in Appendix S3. Seven studies reported correlations that were not at risk of bias and six studies used an OR performance measurement and therefore did not raise concerns regarding applicability.[13,37–48]

Visual–spatial ability and laparoscopic skills

Twenty-seven studies containing 36 participant groups were included in the analysis of visual–spatial ability (Table 1). An overview of the aptitude tests used to measure visual–spatial ability is shown in Appendix S4. In five groups of participants, laparoscopic skills were measured in the OR.[13,48] The mean correlation in DS_0 (dataset in which unreported non-significant correlations are coded as 0) was 0.32 (95% CI 0.25–0.39; $p < 0.001$), Q was statistically significant (74.55; $p < 0.001$) and I^2 was 53%, indicating above-moderate heterogeneity (Fig. 2). When inclusion was limited to studies that used an OR performance to measure laparoscopic skills, the mean correlation in DS_0 increased to 0.50 (95% CI 0.07–0.77; $p = 0.024$).

Table 1 Overview of studies measuring the predictive power of visual–spatial ability

	Study	Year	Country	*n*	Level of training	VSA*	MM	BL	LR	LP	Sign	Total	Correlation[†]	PS[‡]	Remarks
1	Risucci *et al.*[68]	2000	USA	39	20 beginning PGY 1 19 PGY 3–5 (Np > 30)	1, 2b	VT	NA	NA	NA	3	12	0.41–0.71	p	
2	Eyal & Tendick[42]	2001	USA	27	Undergraduates	2a, 2b, 4	LCN	NA	NA	NA	4	4	0.39–0.58	p	
3	Risucci *et al.*[69]	2001	USA	94	23 PGY > 3 71 attending surgeons	1, 2b	VT	+	NA	+	10	12	0.21–0.51	p	Partial correlation corrected for acquired knowledge during course also significant
4	Haluck *et al.*[70]	2002	USA	25	No simulator experience	2a, 2b, 3	LCN	NA	NA	NA	NA	3	0.30–0.39	p	
5a	Keehner *et al.*[13]	2004a	USA	48	Low experience (median Np = 13)	2b	OR	NA	NA	NA	1	1	0.39	p	Performance on live animals
5b	Keehner *et al.*[13]	2004b	USA	45	High experience (median Np = 302)	2b	OR	NA	NA	NA	0	1	0.02	p	Performance on live animals
6	Schijven *et al.*[1]	2004	Netherlands	28	Hospital residents and final-year interns	2b	VRS	NA	NA	NA	0	1	0.40	p	Kendall tau correlation
7	McClusky *et al.*[12]	2005	USA	11	Year 4 medical students	2a, 2b, 3	VRS	NA	NA	NA	0	3	–	p	
8	Stefanidis *et al.*[36]	2006	USA	20	Year 1 surgical residents (median Np = 0)	1, 2a, 2b, 3	VT, VRS, LCN	+	NA	NA	5	30	0.44–0.64	p	
9a	Hedman *et al.*[43]	2006a	Sweden	54	Medical students (*n* = 0)	2b	VRS	+	NA	NA	8	15	0.28–0.40	p	
9b	Hedman *et al.*[43]	2006b	Sweden	25	Medical students (*n* = 0)	2b	VRS	+	NA	NA	8	15	0.43–0.49	p	Correlation significant when compensated for BASIQ-general test scores
10	Keehner *et al.*[38]	2006	USA	22	Non-medical students	5	LCN	+	NA	+	3	4	0.19–0.46	p	Correlation corrected for general intelligence was significant

Table 1 (Continued)

	Study	**Year**	**Country**	***n***	**Level of training**	**VSA***	**MM**	**BL**	**LR**	**LP**	**Sign**	**Total**	**Correlation**[†]	**PS**[‡]	**Remarks**
11	Andalib *et al.*[71]	2006	Canada	32	Medical and dental students (Ns = 0)	2a, 2b, 3, 5	VRS	NA	+	+	3	6	0.36–0.46	a	
12	Birbas *et al.*[73]	2006	Greece	21	Minimal experience	5	VRS	NA	NA	+	1	NA	0.72	a	
13	Hassan *et al.*[72]	2007	Germany	16	NA	2b	VRS	NA	NA	NA	NA	NA	NA	p	Mann Whitney *U* test
14	Enochsson *et al.*[74]	2008	Sweden	9	Gynaecological consultants	2b	VRS	NA	NA	NA	4	4	0.72–0.82	a	
15	Rosenthal *et al.*[31]	2010	Switzerland	56	Novice (Np = 0) to expert (Np > 100)	2b	LCN	NA	NA	NA	4	3	0.28–0.45	p	
16	Sliwinski[75]	2010	Netherlands	7	Surgical and gynaecologic trainees	1, 2a, 2b	VRS	–	NA	+	2	20	0.78–0.88	t	
17	Kolozsvari *et al.*[14]	2011	Canada	32	Medical and dental students (Ns = 0)	2a, 2b, 3	VT	–	–	–	0	9	–	p	Only study to measure surgical interest
18	Jungmann *et al.*[44]	2011	Germany	40	Medical students (Ns = 0)	2b	VRS	NA	NA	NA	2	3	0.38–0.56	p	
19	Ahlborg *et al.*[45]	2011	Sweden	13	Gynaecological consultants	2b	VRS	NA	NA	NA	7	15	0.57–0.64	p	
20	Schlickum *et al.*[76]	2011	Sweden	25	Medical students	2b	VRS	NA	NA	NA	1	1	0.45	p	
21	Luursema *et al.*[11]	2012	Netherlands	23	Technical medicine students	1, 2a, 5	VRS	–	NA	–	0	3	–	p	
22a	Ahlborg *et al.*[39]	2012a	Sweden	28	Gynaecological trainees (Ns = 0, Np = 10)	2b	VRS	+	NA	NA	1	2	0.40	p	
22b	Ahlborg *et al.*[39]	2012b	Sweden	19	Gynaecological trainees with VRS training	2b	VRS	NA	NA	NA	1	2	–	p	
23a	Nugent[46]	2012a	Ireland	40	Pre-clinical medical students Years 1–3	2a, 2b, 3	VRS	+	NA	NA	9	12	0.34–0.48	d	
23b	Nugent[46]	2012b	Ireland	20	12 PGY 1 basic surgical trainees 8 PGY 2 basic surgical trainees	2a, 2b, 3	VRS	NA	NA	NA	3	12	0.45–0.59	d	
23c	Nugent[46]	2012c	Ireland	8	Higher surgical trainees Years 1–3	2a, 2b, 3	VRS	NA	NA	NA	3	12	0.75–0.80	d	

Table 1 (Continued)

	Study	Year	Country	*n*	Level of training	VSA*	MM	BL	LR	LP	Sign	Total	Correlation[†]	PS[‡]	Remarks
23d	Nugent[46]	2012d	Ireland	12	Higher surgical trainees Years 4–6	2a, 2b, 3	VRS	NA	NA	NA	0	12	–	d	
23e	Nugent[46]	2012e	Ireland	26	Pre-clinical medical students Years 1–3	0	VRS	NA	NA	NA	3	4	0.54–0.94	d	
24	Nugent[46]	2012	Ireland	67	General and plastic surgery trainees	0	VRS, bench models	NA	NA	NA	0	2	–	d	
25	Nugent *et al.*[51]	2012	Ireland	10	Surgical trainees (Nbl > 20, Nal < 5)	0	VRS	+	NA	NA	6	13	0.67–0.78	p	
26a	Ahlborg *et al.*[48]	2013	Sweden	28	Gynaecological trainees (Ns = 0, Np < 10)	2b	OR	+	NA	NA	0	1	0.33	p	Performance on humans
26b	Ahlborg *et al.*[48]	2013	Sweden	7	Gynaecological trainees no VRS training	2b	OR	+	NA	NA	1	1	0.98	p	Performance on humans
26c	Ahlborg *et al.*[48]	2013	Sweden	13	Gynaecological trainees with VRS training	2b	OR	NA	NA	NA	0	1	0.13	p	Performance on humans
27	Groenier *et al.*[10]	2014	Netherlands	53	Technical medicine students	1, 2b	VRS	NA	NA	NA	0	1	–	p	

± indicates significant/not significant
BL = baseline performance; LCN = laparoscopic camera navigation; LP = learning plateau; LR = learning rate; MM = measurement method; NA = not addressed; *n* = number of participants; Nal = number of advanced laparoscopic procedures performed; Nbl = number of basic laparoscopic procedures performed; Np = number of procedures performed; Ns = number of simulator tasks performed; OR = animal/human operating room performance; PGY = postgraduate year; PS = publication status; VSA = visual–spatial ability; VRS = virtual reality simulator; VT = video trainer
*0 = composite score static and dynamic; 1 = intrinsic static; 2a = intrinsic dynamic two-dimensional; 2b = intrinsic dynamic three-dimensional; 3 = extrinsic static; 4 = extrinsic dynamic; 5 = composite score intrinsic dynamic and extrinsic dynamic
[†] Correlation: minimum–maximum significant correlation found in the study
[‡] a = abstract; d = dissertation; t = thesis; p = published in peer-reviewed journal

Perceptual ability and laparoscopic skills

In all of the studies included, perceptual ability was assessed with the Pictorial Surface Orientation test (PicSOr). The PicSOr was developed to measure the ability to recognise the 3-D orientation of a virtual object from a 2-D screen.[12,49] Seven studies containing 15 participant groups were included in the correlation analysis of perceptual ability (Table 2). The mean correlation for perceptual ability in DS_0 was 0.31 (95% CI 0.22–0.39; $p < 0.001$); Q was 21.30 ($p = 0.128$), indicating no heterogeneity among studies (Fig. 2).

Psychomotor ability and laparoscopic skills

Four studies containing eight participant groups were included in the correlation analysis of

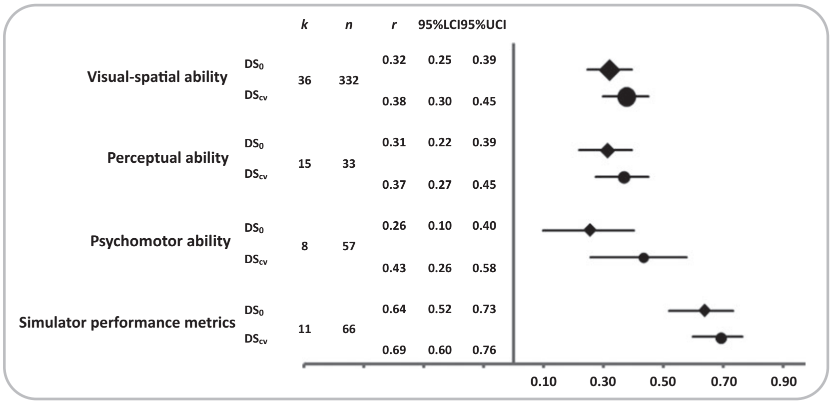

Figure 2 Forest plot of the summary correlations of the different aptitude measurements with 95% confidence intervals. DS_0 = dataset with non-reported non-significant correlations coded as 0; DS_{cv} = dataset with non-reported non-significant correlations coded as the critical Pearson correlation coefficient at $\alpha_2 = 0.05$ and df = $n - 2$; k = number of participant groups; n = number of correlations; r = summary correlation; LCI = lower confidence interval; UCI = upper confidence interval

psychomotor ability (Table 3). An overview of the encountered aptitude tests for psychomotor ability is shown in Appendix S5. Of the seven different psychomotor ability tests used, only the finger tap test and the grooved pegboard test showed significant correlations with laparoscopy performance.[36,46] The mean correlation for psychomotor ability in DS_0 was 0.26 (95% CI 0.10–0.40; $p < 0.003$); Q was 9.85 ($p = 0.197$), indicating no statistically significant heterogeneity among studies (Fig. 2).

Simulator-based assessment

Nine studies containing 11 participant groups were included in the analysis of the predictive validity of simulator-based assessment of aptitude (Table 4). In five of the nine studies, laparoscopic skills training parameters were correlated with OR performance measurements.[37,40,41,47,50] The mean correlation for simulated MIS performance in DS_0 was 0.64 (95% CI 0.52–0.73; $p < 0.001$); Q was 13.78 ($p = 0.183$), indicating no statistically significant heterogeneity among studies (Fig. 2). When the studies included were limited to those with a correlation between simulator performance and a subsequent OR performance, the mean correlation in DS_0 decreased to 0.61 (95% CI 0.42–0.75; $p < 0.001$).

Visual–spatial ability moderator analysis

2 × 2 classification of visual–spatial ability

There was significant heterogeneity in the summary correlation of visual–spatial ability ($p < 0.001$). A moderator analysis was performed to investigate whether the heterogeneity among studies was caused by differences in methodology. The results of the moderator analysis for visual–spatial ability are shown in Table 5. Three studies were excluded from the analysis of the 2 × 2 classification of visual–spatial ability because they used a composite measure of static and dynamic visual–spatial ability tests.[46,51] The subgroup *extrinsic dynamic* contained only one study.[38] The subgroups *intrinsic dynamic* and *extrinsic dynamic* were thus combined into the subgroup *dynamic visual–spatial ability*. Close inspection of this subgroup showed a subdivision of 2-D visual–spatial ability tests, often of low complexity. The subgroup *dynamic visual–spatial ability* was consequently divided into *2-D dynamic* and *3-D dynamic* to create an adjusted 2 × 2 classification.

The subgroup *intrinsic static* showed no statistically significant correlation in DS_0 (dataset with unreported non-significant correlations coded as 0) and DS_{cv} (dataset with unreported non-significant correlations coded as maximum achievable correlation) ($p = 0.069$ and $p = 0.100$, respectively) and the subgroup *extrinsic static* showed no significant correlation only in DS_0 ($p = 0.075$). A significant difference was observed between subgroups ($p = 0.024$). The unknown size of non-significant correlations led to a substantial difference between subgroups in DS_0 and DS_{cv} in the subgroup *extrinsic static* (DS_0: $r = 0.14$, 95% CI −0.01 to 0.28; DS_{cv}: $r = 0.34$, 95% CI 0.19–0.48). Consequently, only the subgroups *3-D dynamic* and *intrinsic static* were mutually compared. Comparison of the *3-D dynamic* and *intrinsic static* subgroups showed a statistically significant difference ($p = 0.024$).

Table 2 Overview of studies measuring the predictive power of perceptual ability

	Study	Year	Country	*n*	Level of training	MM	BL	LR	LP	Sign	Total	Correlation*	PS[†]
1	Haluck *et al.*[70]	2002	USA	25	No simulator experience	LCN	NA	NA	NA	1	1	0.59	p
2a	Gallagher *et al.*[49]	2003	Ireland, UK	48	Laparoscopic novices	VT	NA	NA	NA	1	1	0.50	p
2b	Gallagher *et al.*[49]	2003	Ireland, UK	32	Laparoscopic novices	VT	NA	NA	NA	1	1	0.50	p
2c	Gallagher *et al.*[49]	2003	Ireland, UK	34	Laparoscopic novices and experienced surgeons	VT	NA	NA	NA	1	1	0.42	p
2d	Gallagher *et al.*[49]	2003	Ireland, UK	18	Experienced laparoscopic surgeons	VT	NA	NA	NA	1	1	0.54	p
3	McClusky *et al.*[12]	2005	USA	11	Year 4 medical students	VRS	NA	NA	NA	1	1	0.76	p
4	Stefanidis *et al.*[36]	2006	USA	20	Year 1 surgical residents (median Np = 0)	VT, VRS, LCN	–	NA	NA	0	5	NS	p
5	Kolozsvari *et al.*[14]	2011	Canada	32	Medical and dental students (Ns = 0)	VT	+	–	–	1	3	0.38	p
6a	Nugent[46]	2012a	Ireland	40	Pre-clinical medical students Years 1–3	VRS	+	NA	NA	1	3	0.49	d
6b	Nugent[46]	2012b	Ireland	20	12 PGY 1 basic surgical trainees 8 PGY 2 basic surgical trainees	VRS	NA	NA	NA	1	3	0.52	d
6c	Nugent[46]	2012c	Ireland	8	Higher surgical trainees Years 1–3	VRS	NA	NA	NA	1	3	0.80	d
6d	Nugent[46]	2012d	Ireland	12	Higher surgical trainees Years 4–6	VRS	NA	NA	NA	2	3	0.70–0.73	d
6e	Nugent[46]	2012e	Ireland	26	Pre-clinical medical students Years 1–3	VRS	NA	NA	NA	1	4	0.56	d
7	Nugent[46]	2012	Ireland	67	General and plastic surgery trainees	VRS Bench models	NA	NA	NA	1	2	0.31	d

± indicates significant/not significant
BL = baseline performance; LCN = laparoscopic camera navigation; LP = learning plateau; LR = learning rate; MM = measurement method; NA = not addressed; *n* = number of participants; Nal = number of advanced laparoscopic procedures performed; Nbl = number of basic laparoscopic procedures performed; Np = number of procedures performed; Ns = number of simulator tasks performed; PGY = postgraduate year; PS = publication status; VRS = virtual reality simulator; VSA = visual–spatial ability; VT = video trainer
*Correlation: minimum–maximum significant correlation found in the study
[†] a = abstract; d = dissertation; t = thesis; p = published in peer-reviewed journal

Other moderators

In the moderators *measurement method*, *participant characteristics* and *learning curve*, no significant difference was observed between subgroups (p = 0.553, p = 0.271 and p = 0.507, respectively) (Table 5). Interestingly, a significant correlation was observed in the subgroup *learning plateau* in the moderator *learning curve* and in the subgroup *trained participants* in the moderator *participant characteristics*.

Publication bias

Visual inspection of the funnel plot of visual–spatial ability showed an asymmetrical distribution of the participant group effect sizes in DS_0 (Fig. 3a). The

Table 3 Overview of studies measuring the predictive power of psychomotor ability (PMA)

	Study	Year	Country	*n*	Level of training	PMA test	MM	BL	LR	LP	Sign	Total	Correlation*	PS[†]
1	Schijven *et al.*[1]	2004	Netherlands	28	Hospital residents and final-year interns	GSM, CSPDT	VRS	NA	NA	NA	0	3	NS	p
2	Stefanidis *et al.*[36]	2006	USA	20	Year 1 surgical residents (median Np = 0)	Tremor, Reaction time, Finger tap, Purdue pegboard, grooved pegboard	VT, VRS, LCN	–	NA	NA	3	25	0.56–0.67	p
3a	Nugent[46]	2012a	Ireland	40	Pre-clinical medical students Years 1–3	Grooved pegboard	VRS	NA	NA	NA	2	3	0.38–0.45	d
3b	Nugent[46]	2012b	Ireland	20	12 PGY 1 basic surgical trainees 8 PGY 2 basic surgical trainees	Grooved pegboard	VRS	NA	NA	NA	2	3	0.48–0.69	d
3c	Nugent[46]	2012c	Ireland	8	Higher surgical trainees Years 1–3	Grooved pegboard	VRS	NA	NA	NA	3	3	0.75–0.78	d
3d	Nugent[46]	2012d	Ireland	12	Higher surgical trainees Years 4–6	Grooved pegboard	VRS	NA	NA	NA	1	3	0.7	d
3e	Nugent[46]	2012e	Ireland	26	Pre-clinical medical students Years 1–3	Grooved pegboard	VRS	NA	NA	NA	0	4	NS	d
4	Nugent *et al.*[51]	2012	Ireland	10	Surgical trainees (Nbl > 20, Nal < 5)	Grooved pegboard	VRS	+	NA	NA	5	13	0.77–0.87	p

± indicates significant/not significant
BL = baseline performance; CSPDT = Crawford Small Parts Dexterity Tester; GSM = Gibson Spiral Maze; LCN = laparoscopic camera navigation; LP = learning plateau; LR = learning rate; MM = measurement method; NA = not addressed; *n* = number of participants; Nal = number of advanced laparoscopic procedures performed; Nbl = number of basic laparoscopic procedures performed; Np = number of procedures performed; Ns = number of simulator tasks performed; PGY = postgraduate year; PS = publication status; VSA = visual–spatial ability; VRS = virtual reality simulator; VT = video trainer
*Correlation: minimum–maximum significant correlation found in the study
[†] a = abstract; d = dissertation; t = thesis; p = published in peer-reviewed journal

Begg test and Egger test were both significant (p = 0.014 and p = 0.006, respectively), indicating the possibility of publication bias. The most evident outliers were the studies by Nugent[46] (left upper quadrant) and Ahlborg *et al.*[48] (right lower quadrant). The small participant group in Ahlborg *et al.*[48] was the only group of novice trainees to perform in the OR without prior simulator training. The higher level of difficulty of the task in comparison with the more commonly used simulator tasks may have enlarged the measurable range in skill level, leading to a higher correlation. The large study by Nugent[46] was the only study to measure basic laparoscopic skills in a subgroup of the highest scoring trainees after full basic surgery training. The pre-selection of the highest scoring trainees may have restricted the range of laparoscopic ability and, in consequence, led to the observation of a low correlation in this study. Thus, study methodology probably had opposite effects on the measurable range of laparoscopic skills in the small and large participant groups, visually identified as outliers. When these outliers were removed, the funnel plot was symmetrical and the Begg test and Egger test were not significant (p = 0.075 and p = 0.067, respectively). Therefore, the evidence for publication bias was

Table 4 Overview of studies measuring the predictive power of simulation-based assessment

	Study	Year	Country	*n*	Level of training	Aptitude test	MM	BL	LR	LP	Sign	Total	Correlation*	PS†
1	Macmillan & Cuschieri[47]	1999	Scotland	10	Higher surgical trainees	ADEPT	OR	NA	NA	NA	3	2	0.74–0.79	p
2a	Chaudhry *et al.*[77]	1999	UK	7	Hospital staff	VRS, BL	VRS	NA	NA	+	4	6	0.01–1.00	p
2b	Chaudhry *et al.*[77]	1999	UK	11	Basic surgical trainees and above	VRS, BL	VRS	NA	NA	+	2	6	0.61–0.89	p
2c	Chaudhry *et al.*[77]	1999	UK	17	Medical students	VRS, BL	VRS	NA	NA	+	6	6	0.56–0.98	p
3	Ahlberg *et al.*[40]	2002	Sweden	12	Medical student	VRS	OR	+	NA	NA	2	2	0.33–0.64	p
4	McClusky *et al.*[12]	2005	USA	11	Year 4 medical students	VRS	Duration of training	NA	NA	NA	2	2	0.62–0.73	p
5	Stefanidis *et al.*[36]	2006	USA	20	Year 1 surgical residents (median Np = 0)	VT, VRS, LCN BL	Duration of training	NA	NA	NA	4	6	0.55–0.66	p
6	McCluney *et al.*[37]	2007	Canada	40	Surgical trainees PGY 1–5, surgical fellows and consultant surgeons	FLS	OR (GOALS)	NA	NA	NA	1	1	0.77	p
7	Hogle *et al.*[41]	2008	Canada	10	Surgical trainees PGY 1	VRS	OR (GOALS)	NA	NA	NA	0	1	NS	p
8	Kundhal & Grantcharov[50]	2009	Canada	10	Surgical trainees Np = 5	VRS, LCN	OR (OSATS)	+	NA	NA	19	28	0.67–0.98	p
9	Nugent[46]	2012	Ireland	10	Surgical trainees (Nbl > 20, Nal < 5)	VRS basic tasks	VRS colectomy	+	NA	NA	3	6	0.77–0.92	d

± indicates significant/not significant
ADEPT = Advanced Dundee Endoscopic Psychomotor Tester; BL = baseline performance; FLS = fundamentals of laparoscopic surgery; GOALS = global operative assessment of laparoscopic skills; LCN = laparoscopic camera navigation; LP = learning plateau; LR = learning rate; MM = measurement method; NA = not addressed; *n* = number of participants; Nal = number of advanced laparoscopic procedures performed; Nbl = number of basic laparoscopic procedures performed; Np = number of procedures performed; Ns = number of simulator tasks performed; OR = animal/human operating room performance; OSATS = objective surgical assessment of technical skills; PGY = post-graduate year; PS = publication status; VRS = virtual reality simulator; VT = video trainer
*Correlation: minimum–maximum significant correlation found in the study
† a = abstract; d = dissertation; t = thesis; p = published in peer-reviewed journal

probably caused by the differences in methodology between these two studies.

In the evaluation of publication bias for perceptual ability (PicSOr) the funnel plot showed the absence of participant group effect sizes in the left lower quadrant (Fig. 3b), although the Begg test and Egger test did not indicate the presence of publication bias (p = 0.171 and p = 0.090, respectively). To exclude methodology as a potential cause of bias, the characteristics of the studies in the right lower quadrant of the funnel plot were inspected.[27] No common

Table 5 Results of the moderator analysis for visuospatial ability

Moderator	Subgroup	k	*n*	*r*	95% CI	p_r	*Q*	p_Q	I^2	p_{mod}
Adjusted 2 × 2 classification	Intrinsic static	6	74	0.14	−0.01–0.29	0.069	4.61	0.466	0	0.024
	Extrinsic static	9	29	0.14	−0.01–0.28	0.075	6.21	0.624	0	
	Dynamic 2-D	12	59	0.21	0.08–0.34	0.002	9.11	0.612	0	
	Dynamic 3-D	32	151	0.33	0.26–0.39	0.000	59.91	0.001	48	
Measurement method	VT	5	51	0.21	0.01–0.38	0.044	5.29	0.381	24	0.553
	VRS	22	231	0.32	0.22–0.42	0.000	44.81	0.003	53	
	LCN	7	32	0.34	0.18–0.49	0.000	2.67	0.914	0	
	OR	5	5	0.40	0.17–0.59	0.002	23.93	0.000	83	
Participant characteristics*	Non-medical students	3	102	0.19	−0.08–0.42	0.154	1.71	0.635	0	0.271
	Medical students	10	56	0.33	0.19–0.45	0.000	18.85	0.042	52	
	Novice trainees	10	85	0.40	0.25–0.54	0.000	20.78	0.023	57	
	Trained participants	9	66	0.21	0.04–0.37	0.021	16.49	0.057	51	
Learning curve	BL	13	69	0.23	0.10–0.36	0.002	34.16	0.001	65	0.507
	LR	2	6	0.05	−0.26–0.35	0.381	2.48	0.289	60	
	LP	7	66	0.26	0.08–0.41	0.007	3.41	0.844	0	

Between-group variance among moderators was evaluated with heterogeneity *Q* in a mixed-effects model. A pooled τ^2 among subgroups was used to estimate random-effects model summary estimates within subgroups. A p-value of < 0.05 was considered to indicate statistical significance

95% CI = 95% confidence interval; BL = baseline performance; LCN = laparoscopic camera navigation; LP = learning plateau; LR = learning rate; k = number of groups of participants; *n* = number of correlations in subgroup; OR = operating room; p_{mod} = p-value of between-group variance; p_r = p-value of *z*-score of mean correlation; *Q* = heterogeneity *Q*; p_Q = p-value of heterogeneity *Q*; VRS = virtual reality simulator; VT = video trainer

*Participant characteristics: trained participants = trainees with training in laparoscopic surgery and consultant specialists

difference in methodology was observed in these studies. However, retrospective evaluation of the three excluded articles that addressed perceptual ability showed that these studies reported low correlations.[30,34,35] These studies were not included in the meta-analysis because they reported technical errors during data acquisition. Therefore, the present exclusion criteria may have induced bias across studies examining the PicSOr. Other causes of the asymmetrical shape of the funnel plot that could not be excluded are publication bias and coincidence.[37]

The Begg test and Egger test were not significant for psychomotor ability (p = 0.083 and p = 0.086, respectively), but the funnel plot showed a small participant group with a large effect size. No difference in methodology could be identified in this participant group (Fig. 3c).[46] Exclusion of this outlier led to a smaller, but still significant mean correlation of 0.22 (95% CI 0.08–0.35; p = 0.004).

Neither the funnel plot nor the Begg and Egger tests indicated publication bias for simulator-based assessment (p = 0.756 and p = 0.408, respectively) (Fig. 3d).

DISCUSSION

Multiple studies have shown that medical students with interest in pursuing surgical careers display an equal variety of aptitude as medical students who are not interested in surgery.[52–54] Currently used methods for assessing suitability for medical specialties that require laparoscopy do not provide information about the potential to learn and perform laparoscopic skills to faculty members responsible for the assessment of trainees.[55] It has long been recognised in psychology that visual–spatial ability, perceptual ability and psychomotor ability determine levels of performance in the technical professions to some extent.[56–58] The results of this meta-analysis demonstrate that aptitude tests can be used to predict part of the individual differences in laparoscopic skills. Aptitude tests can therefore be considered to represent a useful adjunct to the currently used assessment methods. A laparoscopy aptitude test could also help low aptitude trainees make the right career decision and support surgical educators in the early recommendation to opt for an alternative

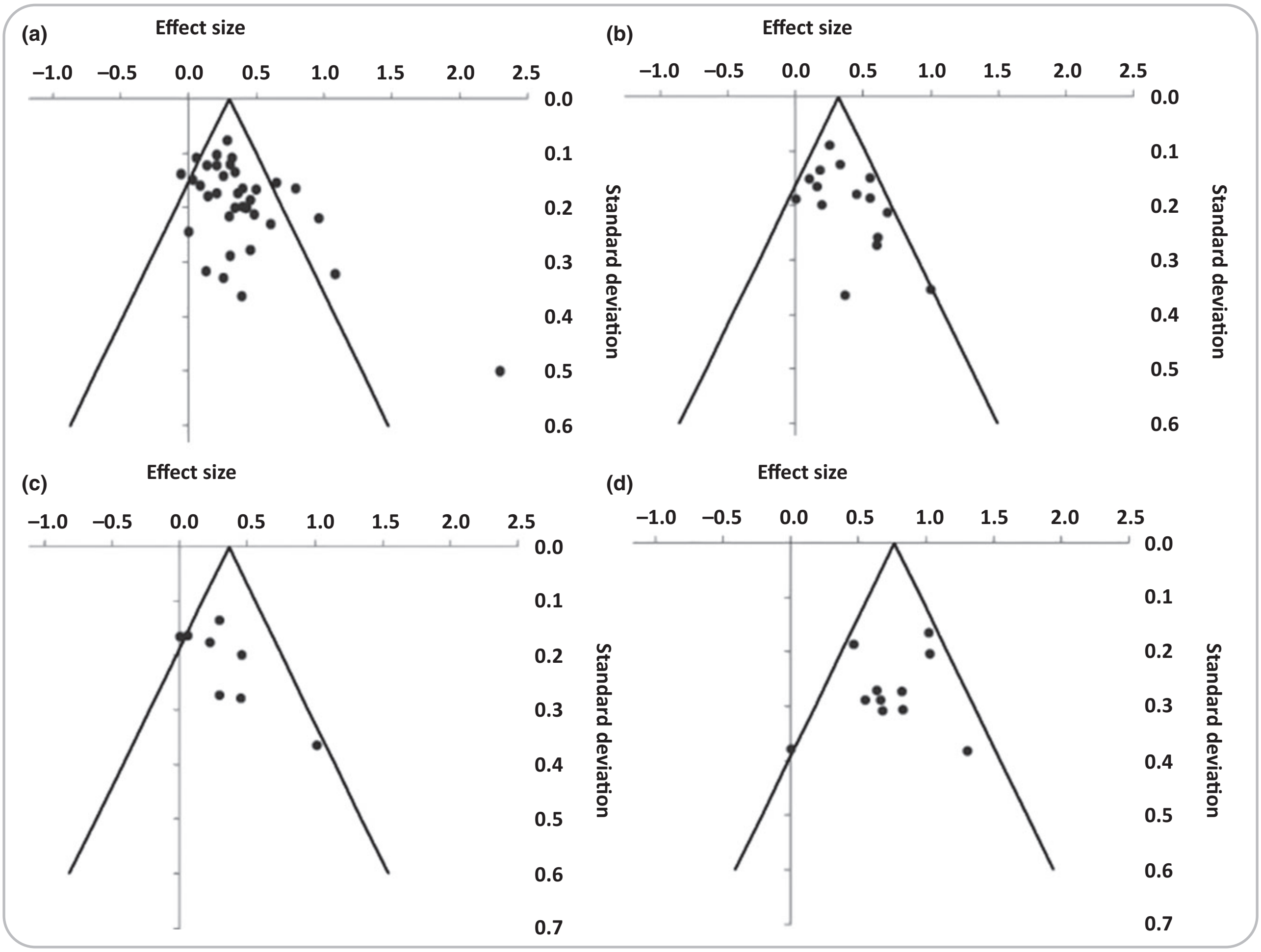

Figure 3 Funnel plot of studies measuring the correlation of laparoscopic performance with: (a) visual–spatial ability; (b) perceptual ability; (c) psychomotor ability, and (d) simulator-based assessment of aptitude

differentiation in surgery or area of medicine. Trainees then have the opportunity to invest their valuable time and energy in a specialty or differentiation program that more closely matches their talent. It is important to note that a laparoscopy aptitude test would not only be beneficial in abdominal surgery, but also in other specialties that depend heavily on laparoscopic skills, namely gynaecology and urology.

Simulator-based assessment of aptitude

If the logistical and organisational burdens are perceived as acceptable and the decision is made to incorporate aptitude testing into the assessment of candidates for laparoscopic surgery, the most appropriate aptitude tests will have to be chosen. All things taken together, we would say that simulator-based assessment is the most viable option for two reasons: (i) simulator-based assessment of aptitude appears to have a relatively high correlation with future laparoscopic skills, accounting for approximately 37–41% of variability in performance between individuals, and (ii) simulators are widely available in surgery departments involved in surgical training and can therefore facilitate laparoscopy aptitude testing with a minimum of additional financial investment and organisational effort.

Although simulator-based assessment shows a high potency in predictive value and practical applicability, it has some important downsides that should be mentioned. Firstly, according to the current quality assessment, all studies that included simulator-based assessment were at risk for bias. Although this does not mean the studies included were indeed all biased, the results as presented should be interpreted with caution because the quality of the studies is low. Secondly, as the number of published cut-off scores to classify candidates according to their potential is limited and based on low sample sizes[1,36,37], a norm-referenced scoring system, based on the ranking of candidates, would currently be more practical than a criterion-referenced scoring system based on cut-off scores. Thirdly, performance level is influenced by previous training and prior experience with video games.[59] The increasing availability of home laparoscopy trainers thus introduces the possibility that what is being assessed is the degree of adaptation to a human–computer interface rather than aptitude.

Visual–spatial, perceptual and psychomotor assessment of aptitude

The summary correlations for visual–spatial ability, perceptual ability and psychomotor ability were all statistically significant. However, they account for only a small proportion (r^2: 6–20%) of the variance in laparoscopic skills. Consequently, if these aptitude measurements are used instead of simulator-based assessment, they can best be used in combination within a laparoscopy aptitude test battery. This would increase organisational burdens, but would optimise the predictive value of a laparoscopy aptitude test by evaluating multiple aspects of potential.

Notably, correlations in this order of magnitude have not always been perceived as barriers to implementation. The correlations between pilot aptitude testing and training and flying performance after training, and between the North American dental aptitude test and practical hands-on dentistry performance levels are reported to be between 0.20 and 0.40.[60,61] Despite controversy around the size of these correlations, many directors of training programmes in aviation and dental education have determined that aptitude testing is of contributive value and therefore continue to implement these tests in selection procedures to optimise the distribution of talent in the workforce and to increase training efficiency.

If the choice is made to use visual–spatial, perceptual and psychomotor abilities rather than simulator-based assessment, one should be aware that the majority of the aptitude tests used in the studies included in the present analysis were developed in the fourth to seventh decades of the 20th century (see references of Appendices S4 and S5) and some have predominantly been evaluated for the ability to identify cognitive or psychomotor deficits in patient populations.[62] The aptitude tests are therefore not optimally adjusted to the challenges imposed by the MIS environment and it may be useful to consider composing cross-functional teams to develop new aptitude tests that reflect the demands of the MIS work environment to a higher extent than do currently available tests.

Moderator analysis

The calculation of the summary correlation of visual–spatial ability, heterogeneity Q and the amount of variance (I^2) indicated there was significant heterogeneity within the sample of included studies. This indicates that the size of the correlation between visual–spatial ability and laparoscopic skills may depend on the methodologies used in the included studies. A moderator analysis was conducted to identify differences in methodology that might have caused heterogeneity among studies.

In the moderator analysis of visual–spatial ability, a significant difference was observed between the mean correlations of the *3-D dynamic* and *intrinsic static* subgroups. This finding seems logical as laparoscopy requires rotation, translation and manipulation of mental representations of intra-abdominal structures. The significant difference in correlation could also have been caused by the use in the majority of studies of a simulator to measure performance level because, in general, simulator tasks do not challenge the ability of an individual to recognise objects on the basis of their characteristics. It is imaginable that *in vivo* laparoscopy, in contrast with simulator tasks, does require the ability to distinguish or recognise relevant structures on the basis of their intrinsic characteristics. For example, obvious visual signals to identify the cystic artery, such as a pulsation, can be absent during dissection of Calot's triangle. Surgeons must then rely on more subtle visual signals to identify the cystic artery and intrinsic static visual–spatial ability may become more relevant. Such nuances can only become visible if aptitude test scores for intrinsic static visual–spatial ability are correlated with the level of performance in these kind of subtask. Task need analysis of laparoscopic procedures, such as described by Tjiam *et al.*[63] could be used to further explore the predictive validity of the different forms of visual–spatial ability and, perhaps, also different forms of perceptual and psychomotor ability in the performance the steps of a procedure.

Further, in the analyses of the moderators *learning curve* and *participant characteristics* there arose some interesting observations. The significant correlations in the subgroup *learning plateau* observed in the moderator *learning curve*, and in the subgroup *trained participants* in the moderator *participant characteristics* both indicate the existence of a difference in capability to perform laparoscopic tasks that cannot be compensated by repetitive task training. Although this indicates an innate component to task performance level after training, we have not evaluated whether there are trainees with a low aptitude score that are actually unable to learn to perform laparoscopic tasks to a proficiency level. We therefore discourage the use of these study results to harness a deterministic perspective on laparoscopic skills training, because it may be that the effects of focused 'deliberate practice', consisting of training on a well-defined task, the provision of detailed immediate feedback and opportunities for practice tailored to individual needs, enable all candidates to eventually achieve the criteria for proficiency in laparoscopy surgery training.[64,65] Aptitude assessment should not become a self-fulfilling prophecy, wherein those who do not perform as well on an aptitude test misattribute their inability to reach proficiency levels to a lack of talent. Motivation, perseverance and deliberate practice are probably greater determinants of technical performance than a score on an aptitude test.

Limitations

Some important limitations should be kept in mind when interpreting the results of this study. Firstly, the methodological weaknesses of this study refer to the risk for bias and concerns of applicability in the included studies, potential bias across studies that were used to estimate the summary correlations of perceptual ability and psychomotor ability, the possibility of inadequate statistical power to identify differences in some parts of the moderator analysis, and the fact that data were extracted from the studies by a single author.

Secondly, as in any non-experimental design, in order to establish the existence of a causal relationship between two variables, this relationship must be shown not to be caused by the action of other variables. Two studies that addressed general intelligence as a possible confounding variable showed that the correlation with visual–spatial ability remains significant and may even increase when the correlation is corrected for general intelligence.[38,43] Although these studies support the hypothesis of aptitude as a determinant of laparoscopic skills independent of general intelligence, further research is necessary to identify the contributions of other confounding factors.

Thirdly, some studies addressing the advantages of binocular imaging systems have shown the improved quality of vision to be beneficial for novices and expert surgeons, inside and outside training centres.[66,67] Thus, although 3-D laparoscopy still has not yet become mainstream, the financial barriers to its widespread implementation may disappear as technology develops and hence some of the findings should be re-evaluated.

Fourthly, medical knowledge, communication skills, decision-making skills and clinical judgement are core clinical competencies that should always be considered in conjunction with technical abilities when surgical competence is addressed. Careful selection of trainees includes a holistic perspective of competency and a thorough assessment of all the technical and non-technical skills required in a surgeon. As stated, a laparoscopy aptitude test can therefore be considered only as an additional source of information that helps to attain a more complete picture of surgical potential.

CONCLUSIONS

In this study, the available evidence has been synthesised to provide programme directors in laparoscopy-related medical disciplines with the most important information for the assessment of aptitude for laparoscopic surgery among candidates. The summary correlations indicate that visual–spatial, perceptual and psychomotor abilities account for part of the variance in laparoscopic skills and that simulator-based assessment appears to have the highest predictive value because it can represent a job sample wherein all aptitudes for laparoscopy can be measured at once. Because of the wide availability of simulators, simulator-based assessment is also the most feasible assessment instrument of aptitude. Considering the importance of technical skills in laparoscopic surgery and the current lack of methods with which to assess the technical potential of candidates, aptitude assessment can be of contributive value for specialties that require laparoscopic skills.

Contributors: KHK contributed to the conception and design of the study, acquisition and analysis of data, and the interpretation of results. MJvD contributed to the study conception and the interpretation of results. NJGMV contributed to the study design, analysis of data and interpretation of results. CH and HOtCH contributed to the interpretation of results. J-PENP contributed to the study conception and design, and to the interpretation of results. All authors contributed to the drafting and critical revision of the paper, and approved the final manuscript for publication.
Acknowledgements: the authors would like to thank Tineke Bouwkamp-Timmer, Department of Education, University Medical Centre Groningen, the Netherlands, and Maarten Jalink, Department of surgery, Academic Hospital Para-

maribo, Suriname, for reviewing the manuscript. They would also like to thank Jan-Maarten Luursema, Liv Ahlborg and Jacek Sliwinski for generously providing the research data from their publications.
Funding: none.
Conflicts of interest: none.
Ethical approval: not applicable.

REFERENCES

1 Schijven MP, Jakimowicz JJ, Carter FJ. How to select aspirant laparoscopic surgical trainees: establishing concurrent validity comparing Xitact LS500 index performance scores with standardised psychomotor aptitude test battery scores. *J Surg Res* 2004;**121** (1):112–9.
2 Grantcharov TP, Funch-Jensen P. Can everyone achieve proficiency with the laparoscopic technique? Learning curve patterns in technical skills acquisition. *Am J Surg* 2009;**197** (4):447–9.
3 Bosker R, Groen H, Hoff C, Totte E, Ploeg R, Pierie J-PP. Early learning effect of residents for laparoscopic sigmoid resection. *J Surg Educ* 2013;**70** (2):200–5.
4 Maan Z, Maan I, Darzi A, Aggarwal R. Systematic review of predictors of surgical performance. *Br J Surg* 2012;**99** (12):1610–21.
5 Paice A, Aggarwal R, Darzi A. Safety in surgery: is selection the missing link? *World J Surg* 2010;**34** (9):1993–2000.
6 Andriole D, Jeffe D, Whelan A. What predicts surgical internship performance? *Am J Surg* 2004;**188** (2):161–4.
7 Bishawi M, Pryor A. Should technical aptitude evaluation become part of resident selection for surgical residency? *Surg Endosc* 2014;**28** (10):2761–2.
8 Gallagher A, Leonard G, Traynor O. Role and feasibility of psychomotor and dexterity testing in selection for surgical training. *ANZ J Surg* 2009;**79** (3):108–13.
9 Müller M. Safety lessons taken from the airlines. *Br J Surg* 2004;**91** (4):393–4.
10 Groenier M, Schraagen J, Miedema H, Broeders I. The role of cognitive abilities in laparoscopic simulator training. *Adv Health Sci Educ Theory Pract* 2014;**19** (2):203–17.
11 Luursema J-M, Verwey W, Burie R. Visuospatial ability factors and performance variables in laparoscopic simulator training. *Learn Individ Differ* 2012;**22** (5):632–8.
12 McClusky DA, Ritter EM, Lederman AB, Gallagher AG, Smith CD. Correlation between perceptual, visuo-spatial, and psychomotor aptitude to duration of training required to reach performance goals on the MIST-VR surgical simulator. *Am Surg* 2005;**71** (1):13–20.
13 Keehner MM, Tendick F, Meng MV, Anwar HP, Hegarty M, Stoller ML, Duh QY. Spatial ability, experience, and skill in laparoscopic surgery. *Am J Surg* 2004;**188** (1):71–5.
14 Kolozsvari N, Andalib A, Kaneva P, Cao J, Vassiliou M, Fried G, Feldman LS. Sex is not everything: the role of gender in early performance of a fundamental laparoscopic skill. *Surg Endosc* 2011;**25** (4):1037–42.
15 Louridas M, Szasz P, de Montbrun S, Harris KA, Grantcharov TP. Can we predict technical aptitude? A systematic review *Ann Surg* 2015; doi: 10.1097/SLA.0000000000001283 [Epub ahead of print.]
16 Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leefland MM, Sterne JA, Bossuyt PM. QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Int Med* 2011;**155** (8):529–36.
17 Borenstein M, Hedges LV, Higgins J, Rothstein HR. *Introduction to Meta-Analysis*. Chichester: John Wiley & Sons 2011.
18 Huedo-Medina T, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: *Q* statistic or I^2 index? *Psychol Methods* 2006;**11** (2):193–206.
19 Uttal D, Meadow N, Tipton E, Hand L, Alden A, Warren C, Newcombe NS. The malleability of spatial skills: a meta-analysis of training studies. *Psychol Bull* 2013;**139** (2):352–402.
20 Study NE. An overview of tests of cognitive spatial ability. In: 66th Engineering Design Graphic Division Mid-Year Conference Proceedings, 22–24 January 2012, Galveston, TX;92–7
21 Kozhevnikov M, Kosslyn S, Shephard J. Spatial versus object visualisers: a new characterisation of visual cognitive style. *Mem Cognit* 2005;**33** (4):710–26.
22 Blazhenkova O, Kozhevnikov M. Visual–object ability: a new dimension of non-verbal intelligence. *Cognition* 2010;**117** (3):276–301.
23 Feldman L, Cao J, Andalib A, Fraser S, Fried G. A method to characterise the learning curve for performance of a fundamental laparoscopic simulator task: defining 'learning plateau' and 'learning rate'. *Surgery* 2009;**146** (2):381–6.
24 Choi Y, Kim Z, Hur K. Learning curve for laparoscopic totally extraperitoneal repair of inguinal hernia. *Can J Surg* 2012;**55** (1):33–6.
25 Ritter FE, Baxter GD, Kim JW, Srinivasmurthy S. Learning and retention. In: Lee JD, Kirlik A, eds. *The Oxford Handbook of Cognitive Engineering*. New York, NY: Oxford University Press 2013;125–2.
26 Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315** (7109):629–34.
27 Sterne J, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;**53** (11):1119–29.
28 Buckley C, Kavanagh D, Gallagher T, Conroy R, Traynor O, Neary P. Does aptitude influence the rate at which proficiency is achieved for laparoscopic appendectomy? *J Am Coll Surg* 2013;**217** (6):1020–7.
29 Buckley C, Kavanagh D, Nugent E, Ryan D, Traynor O, Neary P. The impact of aptitude on the learning

curve for laparoscopic suturing. *Am J Surg* 2013;**207** (2):263–70.

30 Bartenbach H. *Leercurves op de LapSim in relatie tot cognitieve vaardigheden. BSc thesis*. Tilburg, Noord-Brabant: Tilburg University 2014 http://essay.utwente.nl/65616/. [Accessed 7 January 2015.]

31 Rosenthal, R., *et al*. Performance on a virtual reality angled laparoscope task correlates with spatial ability of trainees. *Indian J Surg* 2010: **72** (4): 327–30.

32 Van Bruwaene S, De Win G, Schijven M, De Leyn P, Miserez M. Effect of a short preclinical laparoscopy course for interns in surgery: a randomised controlled trial. *J Surg Educ* 2014;**71** (2):187–92.

33 Cadeddu JA, Kondraske GV. Human performance testing and simulators. *J Endourol* 2007;**21** (3):300–4.

34 Utesch T.*Effects of cognitive aptitude on the initial performance on a laparoscopic simulator. BSc thesis*. Tilburg, Noord-Brabant: Tilburg University 2014. http://essay.utwente.nl/65809/. [Accessed 7 January 2015.]

35 Hilgerink J. *De rol van cognitieve vaardigheden op een laparoscopische simulator. BSc thesis*. Tilburg, Noord-Brabant: Tilburg University 2014. http://essay.utwente.nl/64791/. [Accessed 7 January 2015.]

36 Stefanidis D, Korndorffer J, Black F, Dunne J, Sierra R, Touchard C, Rice DA, Markert RJ, Kastl PR, Scott DJ. Psychomotor testing predicts rate of skill acquisition for proficiency-based laparoscopic skills training. *Surgery* 2006;**140** (2):252–62.

37 McCluney A, Vassiliou M, Kaneva P, Cao J, Stanbridge D, Feldman L, Fried GM. FLS simulator performance predicts intraoperative laparoscopic skill. *Surg Endosc* 2007;**21** (11):1991–5.

38 Keehner M, Lippa Y, Montello D, Tendick F, Hegarty M. Learning a spatial skill for surgery: how the contributions of abilities change with practice. *Appl Cognit Psychol* 2006;**20** (4):487–503.

39 Ahlborg L, Hedman L, Rasmussen C, Felländer-Tsai L, Enochsson L. Non-technical factors influence laparoscopic simulator performance among OBGYN residents. *Gynecol Surg* 2012;**9** (4):415–20.

40 Ahlberg G, Heikkinen T, Iselius L, Leijonmarck C-E, Rutqvist J, Arvidsson D. Does training in a virtual reality simulator improve surgical performance? *Surg Endosc* 2002;**16** (1):126–9.

41 Hogle N, Widmann W, Ude A, Hardy M, Fowler D. Does training novices to criteria and does rapid acquisition of skills on laparoscopic simulators have predictive validity or are we just playing video games? *J Surg Educ* 2008;**65** (6):431–5.

42 Eyal R, Tendick F. Spatial ability and learning the use of an angled laparoscope in a virtual environment. *Stud Health Technol Inform* 2001;**81**: 146–52.

43 Hedman L, Ström P, Andersson P, Kjellin A, Wredmark T, Felländer-Tsai L. High-level visual–spatial ability for novices correlates with performance in a visual–spatial complex surgical simulator task. *Surg Endosc* 2006;**20** (8):1275–80.

44 Jungmann F, Gockel I, Hecht H, Kuhr K, Räsänen J, Sihvo E, Lang H. Impact of perceptual ability and mental imagery training on simulated laparoscopic knot-tying in surgical novices using a Nissen fundoplication model. *Scand J Surg* 2011;**100** (2):78–85.

45 Ahlborg L, Hedman L, Murkes D, Westman B, Kjellin A, Felländer-Tsai L, Enochsson L. Visuospatial ability correlates with performance in simulated gynaecological laparoscopy. *Eur J Obstet Gynecol Reprod Biol* 2011;**157** (1):73–7.

46 Nugent E. *The Evaluation of Fundamental Ability in Acquiring Minimally Invasive Surgical Skill Sets. MD thesis*. Dublin: Royal College of Surgeons in Ireland 2012 http://epubs.rcsi.ie/mdtheses/32/. [Accessed 6 January 2015.]

47 Macmillan AI, Cuschieri A. Assessment of innate ability and skills for endoscopic manipulations by the Advanced Dundee Endoscopic Psychomotor Tester: predictive and concurrent validity. *Am J Surg* 1999;**177** (3):274–7.

48 Ahlborg L, Hedman L, Nisell H, Felländer-Tsai L, Enochsson L. Simulator training and non-technical factors improve laparoscopic performance among OBGYN trainees. *Acta Obstet Gynecol Scand* 2013;**92** (10):1194–201.

49 Gallagher A, Cowie R, Crothers I, Jordan-Black J, Satava R. PicSOr: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance. *Surg Endosc* 2003;**17** (9):1468–71.

50 Kundhal PS, Grantcharov TP. Psychomotor performance measured in a virtual environment correlates with technical skills in the operating room. *Surg Endosc* 2009;**23** (3):645–9.

51 Nugent E, Hseino H, Boyle E, Mehigan B, Ryan K, Traynor O, Neary P. Assessment of the role of aptitude in the acquisition of advanced laparoscopic surgical skill sets. *Int J Colorectal Dis* 2012;**27** (9):1207–14.

52 Cope DH, Fenton-Lee D. Assessment of laparoscopic psychomotor skills in interns using the MIST Virtual Reality Simulator: a prerequisite for those considering surgical training? *ANZ J Surg* 2008;**78** (4):291–6.

53 Langlois J, Wells G, Lecourtois M, Bergeron G, Yetisir E, Martin M. Spatial abilities of medical graduates and choice of residency programmes. *Anat Sci Educ* 2015;**8** (2):111–9.

54 Panait L, Larios J, Brenes R, Fancher T, Ajemian M, Dudrick S, Sanchez JA. Surgical skills assessment of applicants to general surgery residency. *J Surg Res* 2011;**170** (2):189–94.

55 Makdisi G, Takeuchi T, Rodriguez J, Rucinski J, Wise L. How we select our residents – a survey of selection criteria in general surgery residents. *J Surg Educ* 2011;**68** (1):67–72.

56 Griffin GR, Koonce JM. Review of psychomotor skills in pilot selection research of the US military services. *Int J Aviat Psychol* 1996;**6** (2):125–47.

57 Hsi S, Marcia CL, John EB. The role of spatial reasoning in engineering and the design of spatial instruction. *J Eng Educ* 1997;**6** (2):151–8.

58 Ackerman PL, Cianciolo AT. Cognitive, perceptual-speed, and psychomotor determinants of individual differences during skill acquisition. *J Exp Psychol Appl* 2000;**6** (4):259–90.

59 Jalink MB, Goris J, Heineman E, Pierie J-PENP, ten Cate Hoedemaker HO. The effects of video games on laparoscopic simulator skills. *Am J Surg* 2014;**208** (1):151–6.

60 Ranney RR, Wilson MB, Bennett RB. Evaluation of applicants to predoctoral dental education programmes: review of the literature. *J Dent Educ* 2005;**69** (10):1095–106.

61 Hegarty M, Keehner M, Khooshabeh P, Montello D. How spatial abilities enhance, and are enhanced by, dental education. *Learn Individ Differ* 2009;**19** (1):61–70.

62 Causby R, Reed L, McDonnell M, Hillier S. Use of objective psychomotor tests in health professionals. *Percept Mot Skills* 2014;**118** (3):765–804.

63 Tjiam IM, Schout BM, Hendrikx AJ, Scherpbier AJ, Witjes JA, van Merriënboer JJ. Designing simulator-based training: an approach integrating cognitive task analysis and four-component instructional design. *Med Teach* 2012;**34** (10):e698–707.

64 Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004;**79** (10 Suppl):70–81.

65 Wanzel KR, Hamstra SJ, Anastakis DJ, Matsumoto ED, Cusimano MD. Effect of visual–spatial ability on learning of spatially complex surgical skills. *Lancet* 2002;**359** (9302):230–1.

66 Lusch A, Bucur P, Menhadji A, Okhunov Z, Liss M, Perez-Lanzac A, McDougall EM, Landman J. Evaluation of the impact of three-dimensional vision on laparoscopic performance. *J Endourol* 2014;**28** (2):261–6.

67 Sahu D, Mathew M, Reddy P. 3D Laparoscopy – help or hype: initial experience of a tertiary health centre. *J Clin Diagn Res* 2014;**8** (7):NC01–3.

68 Risucci D, Geiss A, Gellman L, Pinard B, Rosser JC. Experience and visual perception in resident acquisition of laparoscopic skills. *Curr Surg* 2000;**57** (4):368–72.

69 Risucci D, Geiss A, Gellman L, Pinard B, Rosser J. Surgeon-specific factors in the acquisition of laparoscopic surgical skills. *Am J Surg* 2001;**181** (4):289–93.

70 Haluck RS, Gallagher AG, Satava RM. Reliability and validity of Endotower, a virtual reality trainer for angled endoscope navigation. *Stud Health Technol Inform* 2002;**85**:179–84.

71 Andalib A, Feldman LS, Cao J, McCluney AL, Fried GM. *Can Innate Visuospatial Abilities Predict the Learning Curve for Acquisition of Technical Skills in Laparoscopy? Scientific Session of the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES)*. Surg Endosc (2006) 20: S283–S291.

72 Hassan I, Gerdes B, Koller M, Dick B, Hellwig D, Rothmund M, Zielke A. Spatial perception predicts laparoscopic skills on virtual reality laparoscopy simulator. *Childs Nerv Syst* 2007;**23** (6):685–9.

73 Birbas KN, Tzafestas CS, Kaklamanos IG, Vezakis AA, Polymeneas G, Bonatsos G. Spatial ability can predict laparoscopy skill performance of novice surgeons. 10th World Congress of Endoscopic Surgery, 14th International Congress of the European Association for Endoscopic Surgery (EAES), 13–16 September 2006. Berlin, Germany.

74 Enochsson L, Ahlborg L, Murkes D, Westman B, Hedman L, Kjellin A, Tsai-Felländer L. Visuospatial ability affects the performance of gynaecological simulation in the LapSimGyn® VR simulator. 16th International Congress of the European Association for Endoscopic Surgery (EAES), 11–14 June 2008, Stockholm.

75 Sliwinski J. *Visuo-Spatial Ability and Damage in Laparoscopic Simulator Training. BSc Thesis.* Tilburg, Noord-Brabant: Tilburg University 2010 http://essay.utwente.nl/60117/. [Accessed 7 January 2015.]

76 Schlickum M, Hedman L, Enochsson L, Henningsohn L, Kjellin A, Felländer-Tsai L. Surgical simulation tasks challenge visual working memory and visual–spatial ability differently. *World J Surg* 2011;**35** (4):710–5.

77 Chaudhry A, Sutton C, Wood J, Stone R, McCloy R. Learning rate for laparoscopic surgical skills on MIST VR, a virtual reality simulator: quality of human–computer interface. *Ann R Coll Surg Engl* 1999;**81** (4):281–6.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Formulae used in meta-analyses.

Appendix S2. Excluded Studies.

Appendix S3. QUADAS-2 Quality assessment of the 34 included studies.

Appendix S4. Aptitude tests used to measure visual–spatial ability.

Appendix S5. Aptitude tests used to measure psychomotor ability.

Received 16 April 2015; editorial comments to author 15 June 2015, 21 September 2015; accepted for publication 24 September 2015