Stereoscopic versus monoscopic displays: Learning fine manual dexterity skills using a microsurgical task simulator

Journal Article

Author(s):
Wehr, Franka; Held, Jürgen

Publication date:
2019-05

Permanent link:
https://doi.org/10.3929/ethz-b-000320074

Rights / license:
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

Originally published in:
Stereoscopic versus monoscopic displays: Learning fine manual dexterity skills using a microsurgical task simulator

Franka Wehr, Jürgen Held*

University of Applied Sciences Schwäbisch Gmünd, Rektor-Klaus Str. 100, 73525 Schwäbisch Gmünd

* Corresponding author.
E-mail addresses: franka.wehr@hfg-gmuend.de (F. Wehr), juergen.held@hfg-gmuend.de (J. Held).

HIGHLIGHTS
• Easier initial learning process when using a stereoscopic display.
• Funnel-shaped task simulator for fine manual dexterity in a microsurgical context.
• Recording and analysis of learning curves extended to a plateau of best performance.
• Linear and non-linear mixed effects models to consider subject specific inferences.

KEYWORDS
Microsurgery, Stereoscopy, Skill Acquisition

ABSTRACT
We investigated the learning of fine manual dexterity with a microsurgical instrument and a new simulator in a context of microsurgery. 30 subjects were divided into two groups. One (3D group) interacted with a stereoscopic and the other (2D group) with a monoscopic display. Visual information for the displays was captured from a surgical stereomicroscope. In 20 trials, both groups performed the repetitive tasks of picking up small rods from a funnel-shaped cavity and placing them outside. In analysing learning curves, we found that the initial learning process for hand-eye coordination is easier with a 3D display, and that performance persists at a higher level of proficiency than with the 2D display option. Thus stereoscopic displays can be especially beneficial for novices, for those learning new procedures, or for providing orientation to operators facing a new or altered spatial situation. Simulators with few reliefs or spatial textures should not be used for comparison between 3D and 2D viewing conditions.
1. Introduction

For several years, stereoscopic or 3D displays have been available for laparoscopic or minimal invasive surgery and have been used to some extent in operating theatres. More recently, and for the field of microsurgery, manufacturers have been promoting stereoscopic displays for surgical stereomicroscopes in addition to the microscope’s optical binocular. The operating surgeon and assistants can use such a display instead of viewing, sometimes with awkward body postures, through the eyepieces of the optical binoculars. In addition, assistants, scrub nurses, learners, or other participants can use the 3D display instead of having a view on a 2D display (fig. 1).

![Fig. 1. a) Surgeon and assistant at the stereomicroscope. Scrub nurse uses a 2D display.
 b) Surgeon and awkward position of assistant at the stereomicroscope.](image)

For using the 3D display instead of the optical binoculars some laboratory studies compared the optical binoculars of a surgical microscope with the use of a stereoscopic display. In these investigations the binoculars ruled out the 3D display regarding the performance in hand-eye coordination (Nibourg et al. 2015, Yu et al. 2012, Yu et al. 2016). The few clinical studies in this field recorded no significant difference (Jianfen et al. 2014, Mendez et al. 2016) or a better performance for the binoculars (Kotsougiani et al. 2016).

When using the 3D display instead the view on a 2D display, the question arises of whether the investment in stereoscopic technology will lead to benefits in learning and performance compared to the less expensive and established standard of using monoscopic 2D displays in microsurgery.

On the one hand, this question has rarely been the topic of past investigations in microsurgery. For example Yu et al. (2016) compared a 3D display with a monoscopic 2D display and found no significant differences in task performance between both displays. On the other hand, numerous comparison studies between 3D and 2D displays can be found, but in the field of video endoscopy, for example in laparoscopy. In the past years, these studies have consistently indicated a better task performance for the 3D displays compared with monoscopic displays, as shown in laboratory experiments (Abdelrahman et al. 2018, Alaraimi et al. 2014, Ashraf et al. 2015, Chiu et al. 2015, Feng et al. 2015a, Guanà et al. 2017, Han et al. 2018, Hani et al. 2014, Harada et al. 2018a, Mashiach et al. 2014, Özsoy et al. 2015, Romero-Loera et al. 2016, Schoenthaler et al. 2015, Smith et al. 2012, Spille et al. 2017, Wilhelm et al. 2014), as well as clinical studies (Agrusa et al. 2018, Bove et al. 2015, Currò et al. 2016, Feng et al. 2015b, Kyriazis et al. 2014) and in systematic reviews (Fergo et al. 2017, Soerenson et al. 2016).
Many earlier studies did not show such consistency. For example, Cicione et al. (2013) conclude with “conflicting findings” in reviewing comparative studies in laparoscopy from the years 1998 to 2012, and Hani et al. (2014) summarized eight studies from 2009 to 2012 of 3D vs. 2D in neuroendoscopy and reported “mixed findings”. In a broader time span, McIntire et al. (2014) reviewed 56 medically-focused experiments of the last five decades and concluded that in 70% of them performance was better when using stereoscopic viewing conditions instead of monoscopic ones.

However, the transference of the recent and more consistent results from video endoscopy to the field of microsurgery can be subject to uncertainties because of at least four differences: (1) ports-guided endoscopes and related hand-finger movements differ from instrumentation and related motor skills used in conjunction with a surgical stereomicroscope; (2) the magnification of the operating area is usually much greater; (3) laparoscopy task simulators often show the topology of a board, whilst in open microsurgery the real operating area is seldom a flat surface and can often be a kind of funnel; (4) in video endoscopy the line-of-sight towards the monitor and the working-axis can be aligned with each other on the surgeon’s sagittal surface; this is unachievable with a surgical microscope because its housing obstructs such a line-of-sight (fig. 1).

Most of the aforementioned comparative studies avoid extensive learning in their experiments, and only some of them analyse learning curves. For example, Feng et al. (2015b), report a study of laparoscopic surgery in animal cadavers with better performance times when using the 3D display. In a with-in subject design, four surgeons performed the same procedure of anastomoses 12 - 14 times during the different display conditions. The authors did not present numerical or graphical data for the repetitions but expressed verbally that learning curves of task time in 3D and 2D conditions had a parallel course with no differences. More explicitly, Smith et al. (2012) showed the course of 3D and 2D learning in detail. They presented graphs of mean task times and mean number of errors covering 10 repetitions of four different simulator tasks for laparoscopy performed with-in subjects of novices. Their curves showed superior performance times and error reduction when using the 3D display. Their learning curves showed a parallel course too, while Jourdan et al. (2004) found curves of 10 repetitions for complex simulator tasks much flatter under 3D than those under 2D conditions. But the limit of 10 repetitions might not have allowed the authors to determine statistically if learning plateaus at the end of the curves had already been reached or not. For their suggestion that the advantages of the 3D display persist once proficient performance has been established, proof through more repetitions might be helpful.

Thus, in our study we were interested in investigating learning more extensively in terms of skill acquisition processes in microsurgery with stereoscopic displays. We wanted to compare two groups of learners. Both groups were to learn fine motoric hand-finger movements with a microsurgical instrument in a context comparable to the process and situation of a microsurgical procedure. The visual information for the hand-eye coordination of both groups was to be derived from a surgical stereomicroscope. One group (3D group) would interact visually with a stereoscopic and the other (2D group) with a monoscopic display. Comparing the 3D and 2D groups, we expected:

H1: Both groups to reach a plateau of best performance.
H2: The performance on the plateau level to be better for the 3D group.
H3: The 3D group to have an easier initial learning process.
2. Methods

2.1 Ethics
The authors’ university delegate for ethics approved the study design.

2.2 Subjects
The 30 subjects were recruited via email or were invited personally. All subjects were randomly divided into two groups of 15 (3D, 2D group), each group comprising 8 males and 7 females. They included students of industrial, communication, and interaction design, and employees of a university. The professional background of the two employees was in mechanical engineering. Mean age of the participants was 27 ± 4.2 years. They were all right-handed. The visual acuity of all subjects was tested, using a Landolt chart procedure (OCULUS Optical devices GmbH). Additionally the stereoscopic visual acuity of the 3D group was tested by applying the principle of the Howard-Dolman test (Howard 1919), i.e. testing the binocular parallax. In this test all subjects of the 3D group were asked to judge the relative positions of three pegs lying in different spatial depth and visualized through the stereoscopic display without provision of any other visual cues for depth, i.e. object occlusion, shadows, relative size or perspective views.

2.3 Laboratory setup
A surgical stereomicroscope (OPMI 1 FC, Carl Zeiss Meditec AG, stereo basis of 22 mm) was used in conjunction with a functional model of a surgical table (200 x 60 cm, adjustable height 60 - 130 cm). The optical binoculars and eyepieces of the microscope were dismounted. Left and right views of disparate images were recorded by two 3CCD remote head cameras (GP-US932A HD, Panasonic Inc.), mounted on the stereoscopic microscope and synchronized and amplified by the camera control unit (Panasonic GP-US932 CUSE). The 3DHD video processing unit (INITION StereoBrain 3D) generated interlaced video images, i.e. the camera signals were interwoven or interlaced with each row or line at a width of 1024 pixels and a height of 1 pixel was derived in alternation from the left or right camera image. Calibration of both cameras alignments (horizontal, vertical, rotational) was achieved using a scaled test pattern.

A LCD monitor (Hyundai S465D, 46 inch viewable diagonal, 16:10, 1920 x 1080 pixels, pixel pitch 0.53025 height x 0.53025 width, 144 Hz = 6.9 ms response rate) with circular polarized filter screen for line-by-line interlaced images was used as a stereoscopic display and in monoscopic mode as a 2D display. The participants of the 3D group wore lightweight, circular polarized passive filter glasses. One of the two stereo canals was used to display a monoscopic image in conjunction with the abovementioned LCD monitor. Neither optical apparatus nor camera settings were altered between 3D and 2D modes.

The subjects were placed in front of the microscope at a viewing distance of 200 cm to the display. The height of the display’s centre point was 130 cm. A shield was mounted on the surgical microscope so the subjects had no direct view of the simulator and their hand/finger movements, but only indirectly using the display (fig. 2).
2.4 Photometrical properties

Full-screen brightness and luminance ratio of the LCD monitor was measured using white- and black-level images. Intraocular cross-talk, i.e. visual information of a disparate stereo image leaking to the wrong eye, was measured through displaying pairs of white and black images for the both stereo video channels. Observed cross-talk ratios were calculated according to Weissman and Woods (2011). All measurements of luminance properties were taken using a luminance measurement camera (LMK mobile, TechnoTeam Illmenau).

![Laboratory setup with subject wearing polarized filter glasses. Computer station for video recording (1), video cameras (2), 3D display (3), stereomicroscope (4), simulator (5), subject (6).]

2.5 Time delay of video signal

Repeated measurements (N=20) of time differences between a real interaction stimulus and the response of displaying this interaction on the LCD screen were taken in stereoscopic and monoscopic signal processing mode. Therefore a stop watch application (iPad, iOS 10.2.1, 1/100 s, Apple Inc.) was placed as stimulus under the microscope, the LCD display was positioned beside it on the table, and comparative still images of the stop watch’s continuous numbers and their delayed video representation on the LCD display were taken photographically (S120, 1/2500 s, Canon Inc.).

2.6 Experimental task

The idea and structure of the task simulator originate with Rocker (2011), who investigated the use of microsurgical instruments through analysis of 43 min slow motion video films and 44 still photographs of arm, hand, and finger movements or postures during surgical procedures in brain and spine microsurgery. The simulator then was improved through trials with 13 experts, who were practicing surgeons in microsurgery with 5 to 15 years of professional experience (mean 9.7 years, SD = 8.8). The principle of the task simulator can be compared with the game "Mikado" or "Pick-A-Stick". Subjects had to use a micro-forceps (platform tip, straight, 110 mm length) to pick out 10 metal rods out of the 27 mm deep
funnel-shaped cavity of the task simulator. The rods were of 0.5 mm diameter and a section of each was wrapped with a 2 mm wide strip of plastic insulation. The forceps were connected through a thin and lightweight cable to a low-voltage electronic beeper. Contact between the uncovered part of the metal rod and another rod or the inner surface of the funnel triggered the beeper to emit an acoustic signal - heard by the subjects - and this was counted as an error. The rods were placed at four different spatial depth levels. Each level had a different alignment of rods. The rods were placed close to each other in predefined positions. On every level the insulation on the rods was a different colour (1st level: three rods with black insulation, 2nd level: three rods with red, 3rd level: three rods with yellow, 4th level: one rod with green) and the rods were of different lengths depending on their position in the funnel. The task was to pick out the rods as fast and as accurately as possible, ideally with no errors or with as few as possible (fig. 3). Each subject was asked directly after completion of all trials to describe verbally and retrospectively any sensations of discomfort or visual fatigue.

Fig. 3. Simulator for testing of fine manual dexterity in a microsurgical setting:
   a) illustration of forceps manipulation, b) geometrical design of the funnel.

2.7 Procedure
The subjects were instructed to take their place in front of the stereomicroscope and adjust the height of the table and the position of their chair so they all sat, relative to their body sizes, in an ergonomically favourable posture to see the display and reach the simulator on the table. The task was explained to the subjects verbally. Subjects were instructed to perform each trial as fast and accurately as possible. The picking task was repeated 20 times by each participant. A start signal was given before each of the 20 rounds. Between trials there were short breaks of approx. 30 s to replace the simulator with an identical one, i.e. one with 10 rods in the predefined positions inside the funnel.

2.8 Data analysis
The experiment was recorded by video. A video analysis was done with standard video software (QuickTime Player, Apple Inc.). The time was captured for each trial, starting with the start signal and finishing when the last ‘Mikado stick’ left the funnel of the simulator. Errors were counted for each trial.

Box plots were created for the time to complete a task (task time) and the number of errors. Whiskers showing the lowest data value within a distance of 1.5 x inter-quartile range (IQR) from the 1st quartile, respectively the highest value within 1.5 x IQR of from the 3rd quartile.
A non-linear mixed effects model, fit by maximum likelihood for the task time, (“nlme”, Pinheiro and Bates, 2016) was used to visualise the learning curves and their courses. The nlme-model has the formula: \( Y_{ij} = A_i + (B_i - A_i) \cdot \exp(-\exp(C_i) \cdot X) + V_{ij} \) whereas \( i = 1, \ldots, 15 \) the subjects in-between one group (3D, 2D) and \( j = 1, \ldots, 20 \) the number of trials. \( Y_{ij} \) is the calculated task time for subject \( i \) and trial \( j \). \( A_i \) is the subject-specific asymptote, \( B_i \) the subject-specific intersection with the y-axis and \( C_i \) the subject-specific logarithmic rate of reduced task time from one trial to the next one or along the repetitions. \( A_i, B_i, C_i \) are independent random variables, while \( V_{ij} \) is the random variable for the residual or remaining component.

Statistical analysis, in particular fitting a linear mixed effects model and an auto-regressive time series model to the data, was conducted with R version 3.3.2 (R foundation for statistical computing, R Core Team 2016). In all tests, values of \( p \) less than 0.05 were considered as an indication for statistical significance.

For hypothesis H1 (plateau of best performance reached) a linear mixed effects model (the statistical package „lme4“, Bates et al. 2015) was fitted by restricted maximum likelihood for performance results during the last five trials, i.e. trial number 16 to 20. It was used to gain slope coefficients in this section of the learning curve for each trial of the both groups (3D, 2D). These slope coefficients were then tested with the null hypothesis that they differ from zero, using the Wald statistical test.

For H2 (performance on the plateau level is better for 3D) the 20 trials and performance values for each subject are regarded as a time series. An auto-regressive model of range one (AR(1)) for each time series was adapted using the equation: \( Y_{n+1} - m = a \cdot (Y_n - m) + e_{n+1} \) whereas \( n = 1, \ldots, 19 \) and \( Y_n \) the performance value for trial number \( n \), \( m \) the central value, \( e_n \) the residual terms and \( a \) the auto-regressive coefficient which represent the subject-specific progress of improvement or learning from trial to trial. The 15 central values \( m \) indicating the asymptotes are compared in the last five trials between the both groups (3D, 2D) using the alternative hypothesis that the 3D performance values are smaller (shorter time duration, less number of errors) and the one-sided Wilcoxon rank-sum test for independent samples.

For H3 (3D group will have an easier initial learning process) the one-sided Wilcoxon rank-sum test for independent samples was used for comparing between the 3D and 2D group the step-down from beginning of learning to the plateau of best performance, i.e. the differences between the means of the performance values for trial number 1 to 3 (beginning) with the means of the performance values for trial number 18 to 20 (plateau).

The described statistical analysis to investigate the hypotheses H1 - H3 was applied to the performance criteria of times to complete the tasks and to the criteria of errors occurring during task completing.

3. Results
All of the 2 x 15 subjects reached a visual acuity value of at least 1.25, which is the best possible value of the used Landolt chart. In the 3D group all 15 subjects passed the stereo acuity test. Full-screen brightness of the LCD monitor was 248 cd/m², luminance ratio of a full screen white versus black image was 91:1. Observed cross-talk ratio from left to right eye:
2.0 % and from right to left: 1.5 % respectively. Measurement of video time lag revealed mean delays of 105 ms (SD=10) for the 3D mode and 137 ms (SD=19) for 2D respectively.

All subjects completed the 20 trials with the task simulator. The total mean time spent in interaction with the 3D or 2D display was 13:08 min:s (SD=04:36) for the 3D group and 17:48 min:s (SD=02:32) for 2D respectively. Maximum/minimum of the total interaction time was 25:07/07:52 min:s for the 3D group and 24:01/14:35 for 2D respectively. Fig. 4 presents the subject-/trial-specific data of task time and the non-linear mixed effects model. Box plots for task time along the 20 trials resulted with 13 outliners for 3D viewing condition versus 18 for 2D. The box plots for number of errors along the 20 trials showed 16 outliers in 3D versus 11 in 2D (fig. 5). The interrogation afterwards revealed that 14 subjects in the 3D group did not experience any discomfort or visual fatigue. One subject stated retrospectively that he had had to make strenuous efforts and rated the viewing process with the stereoscopic display as unpleasant.

3.1 Plateau of best performance (H1)
Both groups reached a plateau of best performance for the task time. The linear mixed effects models for the two groups (3D, 2D) showed slope coefficients for the last five trials (trial number 16 to 20) that do not significantly differ from zero (3D: p = 0.118 and 2D: p = 0.983, Wald-test). For the number of errors in the last five trials, the slope coefficients of both 3D and 2D group were not significantly different from zero and indicated a plateau (3D: p = 0.610 and 2D: p = 0.446, Wald-test).

3.2 Performance on plateau level was better for 3D (H2)
Mean task time for one trial at plateau level, i.e. level of best performance was 33.6 s (SD=13.4) for the 3D group and 43.7 s (SD=8.1) for the 2D group respectively. The central values of the AR(1)-model for the task times of the 3D group were significantly smaller compared with the central values of the 2D group (p = 0.001, one-sided Wilcoxon rank-sum test). Mean number of errors for one trial at level of best performance was 1.1 (SD=1.4) for the 3D group and 2.6 (SD=2.5) for the 2D group respectively. The 3D group error plateau level, represented by the central values of the AR(1)-model for the number of errors, was significantly smaller compared with the 3D group (p = 0.004, one-sided Wilcoxon rank-sum test).

3.3 Easier initial learning process for 3D group (H3)
The step-down for the task time, i.e. the difference between mean task time in the beginning (first three trials) and the plateau (last three trials), of the 3D group was significantly smaller compared to the 2D group (p = 0.012, one-sided Wilcoxon rank-sum test). No significance was found for the number of errors and the hypothesis that the error step-down of the 3D group was smaller compared to the 2D group (p = 0.434, one-sided Wilcoxon rank-sum test).
Fig 4. Task time values and nlme-models of the 2D group (1-15) and 3D group (16-30) along the 20 trials.
Fig. 5. Box plots for task time and errors of the 2D group (white boxes) and 3D group (grey boxes).
4. Discussion

4.1 Different plateaus of best performance for 3D, 2D
The findings of different plateaus of best performance and better performance levels for the 3D group were contrary to Yu et al. (2016), one of the few 3D vs. 2D investigations using a laboratory setting of microsurgery, i.e. simulated tasks under a microscope. In their study of with-in subject design, no significance was found for the difference between 3D vs 2D.

One explanation for this might be that we used an experimental design of between-group instead of a with-in subject design. In the latter, one may consider carryover effects, despite alternations of task sequences. This may appear especially when having relatively long durations for manual dexterity tasks and short or no washout phases. Otherwise, in our between-group design one may argue about differences between the two separated groups due to individual variability, for example variability in fine motor skills or spatial ability. However regarding these skills we assume there was a high level of homogeneity. This is because 28 of our 30 subjects were design students while the two other subjects were proficient in mechanical engineering. All design students passed the university admission exam, during which spatial thinking and in contrast to entrance examinations of students in medicine also expert skills in fine hand drawing were both major criteria and a matter of qualitative evaluation. We assume hereby that expert skills in fine hand drawing correlate with fine motor skills. Nevertheless we expected significant differences when comparing the Mikado simulator performance of design students to experienced surgeons. During an international medical conference we invited 13 experienced surgeons to perform three trials with the Mikado simulator and the stereoscopic display under similar conditions to our student experiment. Unexpectedly the performance did not differ significantly when compared with trial number three of our student group (Mann-Whitney, two-tailed, p = 0.960).

4.2 Easier initial learning process for 3D
In studies of 3D vs. 2D comparison in microsurgery and video endoscopy, for example in the references cited in the introduction, the overall task time or number of trials give reason to assume that performance criteria are often compared at the beginning or in the first third of the learning curves. This may neglect the fact that, when comparing different conditions, learning curves can have different courses and outcomes (fig. 6).

![Skill acquisition courses through task repetitions: (a) parallel, (b) convergence, (c) widening, (d) intersection.](image-url)
The characteristics of those learning curves seem to be of importance for applications where assistants or participants have the opportunity to work with a stereoscopic display during a microsurgical procedure over a prolonged time. An initial benefit in one condition may then be outlined during the course of skill acquisition as for example shown in fig. 6 c) and d).

In our experiment the step-down for the task time, i.e. the difference between mean task time of the first three trials and the last three trials of the 3D group, was significantly smaller compared to the 2D group. This, and the better performance at proficiency level, can be explained by an easier initial learning process under 3D condition compared to 2D, which is represented by the course of type b) in fig. 6. Despite having a 3D technology with lower visual qualities than today, Jourdan et al. (2004) found also lower initial levels and relatively flat courses of learning curves for the 3D mode. This can support the findings from laparoscopy, that “3D will benefit rather novices” as Feng et al. 2015a and other authors suggested, for example Ashraf et al. 2015, Chiu et al. 2015, Özsoy et al. 2015, Spille et al. 2017. But here it should be taken into consideration that in the past most 3D versus 2D studies in the medical domain took place in laparoscopy or video endoscopy. Thus experienced surgeons in monoscopic video laparoscopy may not benefit as much from 3D as novices because of their pre-existing skills from practising with the monoscopic displays they have been accustomed to using in video endoscopy for a long time. In microsurgery, experienced surgeons are used to using the binoculars of the stereomicroscope. Thus working with monoscopic or stereoscopic displays instead is a new condition for them.

It is then of interest to discuss the scenario of beginners trained using stereoscopic displays but in practice required to perform surgical tasks with monoscopic displays. We did not investigate such a 3D-to-2D sequence, and only a few studies exist for this crossover aspect for education in laparoscopy. For example Ashraf (2015) found significantly better performance between groups when first trained on 3D and switching to 2D compared to the 2D performance of the group first trained in 2D, while Özsoy et al. (2015) found only slightly better performance for the 3D-to-2D sequence. Poudel et al. (2017) looked into skill transferability between 3D and 2D after a proficiency level was reached. For simple peg transfer tasks they revealed no significant differences in 2D performance between 3D-to-2D group and the group first trained in 2D, while for suturing tasks, which require more spatial perception, the 3D-to-2D group performed slower in 2D compared to the proficiency level of the group first trained in 2D. In a similar study, Harada et al. (2018b) also reported a drop in performance from 3D-to-2D for more complex tasks. Interestingly Dawidek et al. (2017) used this performance drop in a training plan and reached proficiency in the shortest time, when the first part of the training was done in 3D and was finished in 2D.

Especially for education in microsurgery it is of interest to discuss whether it might be advantageous to do initial training with a 3D simulator followed by a switch to the binocular optics of a stereomicroscope. Unger et al. (2016) found no benefit for this, but their results still remain unsupported due to the lack of similar studies.

4.3 Learning curve models
To test the three hypotheses of our study we used a non-linear mixed effects model (nlme) to reach subject-specific learning curve fitting by maximum likelihood estimation (H1), an autoregressive time series model (H2), and split groups (H3). Ramsay et al. (2000) reviewed studies of such learning curve assessment in relation to new health technologies. They revealed that curve fitting is seldom used, in only 5 of 272 reviewed studies, while descriptive
plotting of data points and splitting the learning curves into two or three groups of trials are the two most frequent methods used. In their content analysis Ramsay et al. (2000) found no application of nlme models, nor did Anzanello et al. (2011) in reviewing the state-of-the-art of learning curve models in the field of industrial engineering. Thus only a few studies in the field of surgical skill training with simulators used learning curve fitting. For example Feldman et al. (2009) or Kotsougiani et al. (2016), both were using a linear model to estimate parameters of inverse curve \( y = a - b/x \) fitting. At first sight such linear models appear simple to apply, but then parametric extensions are required to provide an appropriate covariance structure for subject-specific interferences and increase the complexity of the curve fitting, which is then a nonlinear model for parameter estimation (Papachristofi et al. 2016). Thus we estimated that we noted the same complexity in using the nlme approach for describing change over time in repeated measurements, especially individual differences in response, as discussed and recommended by Cudeck and Harring (2007).

4.4 Simulator Design
The design of our Mikado-simulator derives from microsurgical task analysis and hands-on reviews through surgical expert heuristics. We decided not to copy laparoscopic simulator tasks, for example the frequently used peg transfer, as we found most of them less demanding when using microsurgical forceps instead of endoscopic instruments. Furthermore, in analysing the studies referred to in the introduction, we observed that in most of the former studies simulator targets are placed on neutrally painted surfaces or boards surrounded by no, or very few, other reliefs and textures at different height levels. Whereas in real open microsurgery the surrounding area of the operating focus provides textures at different height levels and is often topologically a kind of funnel, as we tried to simulate in our study (fig. 3). Thus the use of such simulators seems unsuitable for comparisons between 2D and 3D viewing conditions, because it suppresses the utilisation of monocular cues for depth perception. As a limitation of our simulator design, we did not assess predictive validity nor concurrent validity as proposed by Evgeniou et al. (2018) in their review of 44 simulator models in microsurgical training.

4.5 Speed and accuracy
The task in our experiment leads to a speed and accuracy trade-off (SATO) as we informed the subjects about concurrent scoring of time and errors. Because finger and wrist movements, distances or amplitudes of movements, tools and target sizes are similar to a micro miniature assembly task investigated by Hancock et al. (1973), we can refer to their description of speed and accuracy trade-off (SATO) in such stereomicroscopic work procedures. They were able to predict movement times using an index of difficulty, i.e. a ratio of target size and movement amplitude as proposed from Fitts (1954) and known as Fitts’ Law.

Concurrent monitoring or scoring of speed and accuracy is common performance measurement in medical simulator training in relation with hand-eye coordination. This is in contrast to real microsurgical procedures, where trading accuracy for speed is not a matter of professional practices, except in emergency cases. The design of the simulator or the simulated task can take this into consideration, for example when the overshooting or undershooting of the target will have an immediate feedback, as in our study an audio signal. Thus the performing subject will be reminded to maintain accuracy during the task. In addition our simulator design supports a ‘better-avoid-any-haste’ performance, because inaccurate manipulations with the forceps will lead to the fall down of a ‘Mikado stick’ and in a
kind of chain reaction impacting other ‘Mikado sticks’ to fall too. This forces new errors when subjects are urged to recover from the occurred clutter of rods.

4.6 Delayed visual feedback
We measured a significant (Mann-Whitney, p < 0.00001) greater time delay between manual instrument operation and depicted video image in 2D mode (137 SD 19 ms) than in 3D (105 SD 10 ms). We expected the contrary, because for 2D mode we bypassed the 3DHD video unit and therefore shortened the signal processing chain.

From telemanipulation and target acquisition studies (Kim et al. 2005, Maida et al. 2016), we can estimate that our delay differences between 2D and 3D viewing conditions hardly affect task performance differences. However, such values of delayed visual feedback can have adverse side-effects, for example when the surgeons’ adaptation to the delayed system is disrupted by frequently changing from display view to direct view at the operating area and vice versa. As example for this impact, MacKenzie and Colin (1993) found a linear performance degradation of up to 64% in movement time and an increment of up to 214% in error rate when delay time changed from 8.3 to 225 ms. For comparison we applied the method described in chapter 2.5 to measure time delay in two surgical stereomicroscopes in a showroom of the manufacturer. In a stereomicroscope with integrated stereo cameras and processing unit we found a delay of 77 SD 18 ms for 3D mode and 63 SD 16 ms for 2D respectively. But in a microscope in which stereo cameras and processing unit are not integrated, but added on devices, we measured a time delay of 201 SD 9 ms for 3D mode.

4.7 Image quality
The stereoscopic display used in our experiment is similar in resolution and luminance to state-of-the-art displays in medicine. We used a stereoscopic display with passive glasses providing circular polarized filter functions to separate the two disparate images, which - as is the status quo for so-called 3DHD displays in medicine - were presented simultaneously at the 3D display. Hence our full HD display resolution of width x height = 1920 x 1080 pixels is used to present two images of 2 x (1920 x 540) pixels, interlaced line-by-line. Both display manufacturers and the overwhelming majority of previous 3D versus 2D studies use the terms “stereoscopic Full HD” or “3DHD”. Unfortunately they often do not explain that this means a presentation of two disparate 2D images derived from HD cameras but interlaced through deleting half of the lines of each image. Thus Full HD or 3DHD end up in fact with a resolution of one half in height i.e. half the resolution the display can offer. Paradoxically, in some actual clinical assessments, 4K or Ultra HD (3840 x 2160 pixels) resolution cameras are used in conjunction with a Full HD 3D display, subsequently diminishing the camera resolution to one fourth in height and one half in width, i.e. the total image resolution to one eighth (Beez et al. 2018, Rossine et al. 2017, Oertel et al. 2017).

Furthermore, by virtue of a pixel’s height, any line-by-line interlacing causes a disruption, because for each second pixel line of the left image the corresponding line of the disparate right image is missing, and vice versa. Thus with a passive polarization technique, 3D versus 2D comparisons are often comparisons between full and half image resolutions. But then in most previous simulator studies high resolution of display images seems not to be an issue, because the artificial and neutral simulator design does not require the operator to discriminate among a wide range of fine structures. Interestingly, when comparing ex vivo 3D-half-Full HD images of resected specimens from tumours with 2D Full HD, Nomura et al.
(2018) found significantly higher diagnostic accuracy in 3D mode among novices but not among experts or trainees.

4.8 Discomfort
The overall time spent in interaction with the stereoscopic display varied in the 3D group from 8 to 25 min, which are relatively short durations compared to microsurgical procedures. For example, Rocker (2011) analysed four surgical cases in microsurgery and found time durations of surgeons’ hand-eye coordination in conjunction with the microscope of 40, 65, 115 and 275 min. Thus, in our study, the lack of extensive interaction and the relatively short task time between 18 to 151 s for each trial can be the reason why only one of the 15 subjects in the 3D group experienced visual fatigue. We identified visual fatigue or discomfort qualitatively, and this could be improved through the use of standardized subjective measurement tools (McIntire et al. 2014).

Generally, all stereoscopic display technologies like passive polarization filters, shutter or auto-stereoscopic displays can provoke visual discomfort and fatigue. Related perceptual and human factor risks are known as: inter-ocular differences in luminance and contrast, accommodation-vergence mismatch, stereo-anomaly, spatio-temporal frequency effects, distance scaling of disparity, and high-level cue conflict (Patterson 2012). International standards address these risks (ISO 9241-392, 2015), which may not occur representatively in laboratory studies due to the small amount of interaction time, but in clinical settings with longer 3D interaction periods (Wanibuchi et al. 2018). Nevertheless, experimental work should focus on factors involved in the causalities of visual impairments and discomfort and ideally should use a consistent framework for communicating them. McIntire et al. (2014) proposed such a framework and revealed the lack of information about such factors as for example the type of stereo display, viewing distance, visible monocular clues, or disparity values across studies in 3D/2D comparison. We can support these findings, for example of the 28 studies of the years 2014 to 2018 cited in chapter 1, only four provide the viewing distance. None of them reports about inter-ocular cross talk, although it is rated as probably the most serious human factor issue in stereoscopy and Patterson (2012) recommended limiting cross talk to less than 2% to avoid risks of viewing discomfort.

The recommendation given by international standards is to reduce cross talk to as low a level as reachable (ISO 9241-392, 2015) and that cross talk should not exceed 5% and shall not exceed 10% (ISO 9241-333, 2017). However, it seems of importance for the field of microsurgery, where discrimination of colours and tissue structures is a demanding visual task, that smaller values such as 2 - 7% of cross talk can disturb binocular fusion and degrade image quality (Yeh and Silverstein, 1990). Thus cross talk should be considered as mandatory information in the manufacturer’s display specification.

5. Conclusion
In the experimental setting of this study, hand-eye coordination with fine manual dexterity and a stereoscopic display will have an easier initial learning process, and persist at proficiency plateau level in better performance than the monoscopic display option. From this characteristic it can be concluded that stereoscopic displays in microsurgery can improve the process of achieving spatial orientation and performing with fine hand-eye coordination. This is especially the case when new visual information arises, in the sense of not matching an
already existing mental representation, or to established motor skill routines of the surgeon, assistants, scrub nurses or participating learners in the operating room.

Thus, the findings suggest that stereoscopic displays can be especially beneficial for novices, or for learning new procedures, or to find orientation and motor control when confronted with a new or changed spatial situations. As the latter may occur in many microsurgical procedures, it can be assumed that experts will also benefit from this advantage in learning. However, regarding the effect of performance improvement for skilled experts, further research should be undertaken to determine whether it is relevant enough to outweigh the added expense and the disadvantages stereoscopic displays still exhibit in terms of probable visual impairments and side-effects. For these studies simulators without any or only little surrounding relief or spatial texture, as employed in nearly all of the former studies, should not be used for comparison of 2D versus 3D viewing conditions.

Acknowledgements
The authors would like to acknowledge the contributions from surgeons given as attendees during the 15th international Course on Spinal Advanced Microsurgery in Hamburg/Germany to the development of the “Mikado” task simulator. Also the authors would like to acknowledge Jürgen Dippon, PhD, Institute of Stochastics and Applications, University of Stuttgart, Germany for his valuable support in the statistical analysis.

Disclaimer
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

http://dx.doi.org/10.1097/GOX.0000000000000727.


https://dx.doi.org/10.1055/a-0577-3009.


http://dx.doi.org/10.1007/s10103-015-1739-0.


http://dx.doi.org/10.1007/s00464-010-3134-9.


http://dx.doi.org/10.1016/j.circir.2015.06.032.


http://dx.doi.org/10.1016/j.wneu.2017.08.083.


http://dx.doi.org/10.1007/s00464-014-3487-9.


http://dx.doi.org/10.1177/00187208903200104.


http://dx.doi.org/10.3233/WOR-2012-0901-5634.