HAL open science

Validation of virtual reality arthroscopy simulator relevance in characterising experienced surgeons

Alexandre Tronchot, Julien Berthelemy, Hervé Thomazeau, Arnaud Huaulmé, Paul Walbron, François Sirveaux, Pierre Jannin

▶ **To cite this version:**

Alexandre Tronchot, Julien Berthelemy, Hervé Thomazeau, Arnaud Huaulmé, Paul Walbron, et al.. Validation of virtual reality arthroscopy simulator relevance in characterising experienced surgeons. Orthopaedics & Traumatology: Surgery & Research, 2021, 107 (8), pp.103079. 10.1016/j.otsr.2021.103079 . hal-03414455

HAL Id: hal-03414455

https://hal.science/hal-03414455v1

Submitted on 16 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Journal Pre-proof

Original article

Validation of virtual reality arthroscopy simulator relevance in characterising experienced surgeons

Alexandre TRONCHOT [1,2] ; Julien BERTHELEMY [1] ; Hervé THOMAZEAU [1,2] ; Arnaud HUAULMÉ [1] ; Paul WALBRON [3] ; François SIRVEAUX [3] ; Pierre JANNIN [1]

1. Univ. Rennes, Inserm, LTSI – UMR 1099, F-35000 Rennes, France
2. Orthopaedics and Trauma Department, Rennes University Hospital, 2 rue Henri Le Guilloux, 35000, Rennes, France
3. Orthopaedics Department, Nancy University Hospital, Centre Chirurgical Emile Gallé, 49 rue Hermite, 54000 Nancy, France

Corresponding author:

Alexandre Tronchot, MSc

Univ. Rennes, CHU Rennes, Inserm, LTSI, Equipe MediCIS- UMR 1099, F-35000, Rennes, France
Phone: Mobile +33 6 38 40 73 33 Work + 33 99 28 43 21
E-mail: Alexandre.TRONCHOT@etudiant.univ-rennes1.fr
ORCID: 0000-0002-3092-3282

Abstract

Background Virtual reality (VR) simulation is particularly suitable for learning arthroscopy skills. Despite significant research, one drawback often outlined is the difficulty in distinguishing performance levels (Construct Validity) in experienced surgeons. Therefore, it seems adequate to search new methods of performance measurements using probe trajectories instead of commonly used metrics.

Hypothesis It was hypothesized that a larger experience in surgical shoulder arthroscopy would be correlated with better performance on a VR shoulder arthroscopy simulator and that experienced operators would share similar probe trajectories.

Materials & Methods After answering to standardized questionnaires, 104 trajectories from 52 surgeons divided into 2 cohorts (26 intermediates and 26 experts) were recorded on a shoulder arthroscopy simulator. The procedure analysed was the "loose body removal" in a right shoulder joint. 10 metrics were computed on the trajectories including procedure duration, overall path length, economy of motion and smoothness. Additionally, Dynamic Time Warping (DTW) was computed on the trajectories for unsupervised hierarchical clustering of the surgeons.

Results Experts were significantly faster (Median 70.9s Interquartile range [56.4-86.3] vs. 116.1s [82.8-154.2], $p<0.01$), more fluid ($4.6.10^5 mm.s^{-3}$ [$3.1.10^5$-$7.2.10^5$] vs.

Journal Pre-proof

$1.5.10^6$mm.s^{-3} [$2.6.10^6$-$3.5.10^6$], p=0.05), and economical in their motion (19.3mm² [9.1-25.9] vs. 33.8mm² [14.8-50.5], p<0.01), but there was no significant difference in performance for path length (671.4mm [503.8-846.1] vs 694.6mm [467.0-1090.1], p=0.62). The DTW clustering differentiates two expertise related groups of trajectories with performance similarities, respectively including 48 expert trajectories for the first group and 52 intermediates and 4 expert trajectories for the second group (Sensitivity of 92%, Specificity of 100%). Hierarchical clustering with DTW significantly identified expert operators from intermediate operators and found trajectory similarities among 24/26 experts.

Conclusion This study demonstrated the Construct Validity of the VR shoulder arthroscopy simulator within groups of experienced surgeons. With new types of metrics simply based on the simulator's raw trajectories, it was possible to significantly distinguish levels of expertise. We demonstrated that clustering analysis with Dynamic Time Warping was able to reliably discriminate between expert operators and intermediate operators.

Clinical Relevance The results have implications for the future of arthroscopic surgical training or post-graduate accreditation programs using virtual reality simulation.

Level of evidence: III; prospective comparative study.

Keywords VR Simulation - Shoulder Arthroscopy - Training - Arthroscopic skills

1. Introduction

Surgical simulation aims to teach practical skills outside the operating room. It is necessary for safety[1] and organizational reasons related to the functioning of health facilities[2] and the time constraints of teachers and students. In France, procedural simulation has been officially included in the medical and surgical training resident national curriculum since 2017[3]. Virtual reality (VR) simulation is a tool that complements traditional techniques on models[4], or cadaveric subjects[5]. It is particularly suitable for learning arthroscopy skills[6–8]. Numerous researches have been dedicated to the development of this tool and its use. In two recent meta-analysis[9,10], one of the main outlined drawbacks was the difficulty in distinguishing performance levels (Construct Validity) with traditional metrics (such as duration and

Journal Pre-proof

path length) in homogeneous experienced populations. Therefore, it seems adequate to search for new metrics or new methods of performance measurements.

In data mining, hierarchical clustering is a method of analysis which seeks to build a hierarchy of clusters with a linkage criterion. In time series analysis, dynamic time warping (DTW) is one of the algorithms for measuring similarity between two temporal sequences, which may vary in speed. Hierarchical clustering with DTW was used with success to distinguish different levels of expertise on studies about VR simulation in robotic surgery training[11] but have not yet been analysed in arthroscopic training.

It was hypothesized that a larger experience in surgical shoulder arthroscopy would be correlated with better performance on a virtual reality shoulder arthroscopy simulator and that experienced operators would share similar probe trajectories. The objective of this study was to analyse instrument trajectories and to determine the ability of the simulator to distinguish individual performance levels within cohorts of experienced surgeons of different expertise levels using innovative metrics and hierarchical clustering with DTW.

2. Methods

2.1. Population

The trajectories of two cohorts have been prospectively collected during international orthopaedic surgery congresses, French Society of Arthroscopy (SFA) held in Strasbourg-December 2018, and the International Shoulder Course, Val d'Isère-January 2019. Each participant answered a standardized questionnaire about his/her expertise in arthroscopy, his/her hobbies (**Table 1**). Informed consent was obtained for all the participants. Questions regarding the number of actual surgical procedures ever done before doing the exercise served to establish base-line experience levels for each cohort. The first cohort called INTERMEDIATE is composed of 29 orthopaedic surgeons with an interest in arthroscopic surgery (surgeons at the end of the resident training phase and fellows), who have performed between one and 99 arthroscopic surgeries as principal operator possibly under supervision of a senior surgeon. The second cohort called EXPERT is composed of 28 orthopaedic surgeons with a subspecialisation in arthroscopic shoulder surgery who have performed more than 100 arthroscopic

Journal Pre-proof

surgeries as principal operator. This experience threshold was previously used in a similar study[12]. Only actual surgical practice was assessed to determine the expertise level, therefore previous practice on any surgical simulator has not been take into account. Three INTERMEDIATE and two EXPERT trajectories data were incomplete. The remaining 26 INTERMEDIATE and 26 EXPERTS participant trajectories data were analysed for the study. A two-sided power calculation was performed for the null hypothesis ("performance of experts is equal to that intermediate") with an acceptable type 1 error, $\alpha = 0.05$ and acceptable type 2 error $\beta = 0.20$, power = 0.8. For the power calculation, data from previous shoulder arthroscopy validation study were used [13]. This resulted in a minimum number of 14 subjects in each group, which is in line with previously reported validation studies[12–15].

2.2. Methods of assessment

The VirtaMed AG (Zurich) ArthroS™ arthroscopy simulator was used to collect trajectories. Face and construct validity of the simulator and its different modules have been assessed in previous studies for novice and experts population [12,13,16]. The *FAST* and *SHOULDER* modules were used for the study. The exercise corresponded to a basic procedure in shoulder surgery: the extraction of foreign bodies in a shoulder joint. "*Catch the Stars Glenohumeral*" is a semi-realistic immersion exercise with modelling of a pathology-free shoulder joint where the goal is to grasp and extract five foreign bodies (i.e., five stars).

The data collection of study participants was done according to a standardized model (**Figure 1**). First, the participant performed exercises on the *FAST* module using the arthroscopy simulator (**1.1, 1.2 and 1.3 in figure 1**). Second, the participant was introduced to the modelling of the right shoulder joint by performing a video-guided identification of the different structures of the joint in the *SHOULDER* module, procedure done in lateral decubitus position for all the participants. (**2.1 in figure 1**). Third, the users performed the *Catch the stars glenohumeral* exercise in the SHOULDER module (**2.2 in figure 1**) in lateral decubitus which he/she performed 3 times. The first trial corresponding as the learning curve was removed from the study analysis.

2.4. Statistical Analysis

Journal Pre-proof

The Matlab-based program *TRACAS (TRajectory Analysis for Computer Assisted Surgery)* was adapted to the arthroscopy simulator data and used for trajectory analysis. In the first step, the data extracted from the simulator were used to compute ten metrics (defined in **Figure 2**): *duration, path length, velocity, acceleration, jerk, volume, economy of movement, motion smoothness, spectral arc length smoothness* and *depth*. The analysis was done following previously published techniques[17–20]. Computation of all metrics and trajectory analysis were run on a dedicated computer (Xeon E5-1650V4 @3.60GHz with 32Go RAM) using Matlab_R2018b. Non-parametric test (two-sided Mann–Whitney U test) was used for these metrics between the different manipulators in order to observe which of these metrics showed statistical differences ($p < 0.05$). Box-plots representations were modelled. The second step of the study consisted of clustering the trajectories of participant using the ascending hierarchical classification with Dynamic Time Warping (DTW). DTW is relevant for the time series indexer because it has the ability to recognize shapes regardless of their time distortion or their phase shift[21].

3. Results

3.1. Metric Analysis

Analysis of the probe trajectory metrics revealed significant differences between the INTERMEDIATE and EXPERT groups (**Table 2**). Box-plots are represented in (**Figure 3**) and (**Figure 4**). The expert group significantly outperformed the intermediate group for the following metrics (**Figure 3**): *Duration (s)*, $p<0.01$ *(***Figure 3.1***); Average velocity (mm.s^{-1})*, $p<0.01$ *(***Figure 3.2***); Economy (mm^{2})*, $p<0.01$ *(***Figure 3.3***)*; *Motion Smoothness (mm.s^{-3})*, $p=0.05$ *(***Figure 3.4***)*; *Spectral Arc Length* Smoothness, $p<0.01$ *(***Figure 3.5***)* and *Working Volume (mm^{3})*, $p<0.01$ *(***Figure 3.6***)*. There were no significant differences between the groups for the following metrics **(Figure 4)**: *Path length (mm)*, $p=0.08$ *(***Figure 4.1***); Average Acceleration (mm.s^{-2})*, $p=0.39$ *(***Figure 4.2***)*; *Average Jerk (mm.s^{-3})*, $p=0.41$ *(***Figure 4.3**) and *Depth (mm)*, $p=0.11$ *(***Figure 4.4***)*.

3.2. Trajectory Analysis

Journal Pre-proof

The results of unsupervised clustering made from the DTW are given in (**Figure 5**). On this dendrogram, each datum is represented in abscissa and they all join one by one until forming a single cluster. The lower the groupings of clusters, the lower the DTW distance between the aggregated data and the more similar the trajectories. Since two categories are studied here, it was decided to compare the two largest clusters only. The first cluster was composed of 48 similar trajectories, 100% of the data from the first category was derived from the EXPERT cohort (24/26 experts) while 93% of the second cluster was composed of data from the INTERMEDIATE cohort (26/26 intermediaries) and 7% from the EXPERT cohort (2 experts) for a total of 56 similar trajectories (Sensitivity of 92%, Specificity of 100%).

4.Discussion

About this loose body removal arthroscopic procedure in a virtual-reality simulated shoulder, the results confirmed that six among the ten metrics computed on the trajectories were able to distinguish between expert and non-expert operators: experts had more fluid (better result for spectral arc length smoothness), more economical (better results for economy and working volume), and more efficient (better results for time and speed) gesture. However, the acceleration and path length did not statistically differ between groups. In the same way, Rahm et al.,[13] had noticed that the path length metric was not as discriminative as the others used in this simulator. Although often used by VR simulators, path length does not seem to be correlated with expertise. The Motion Smoothness metric was calculated with different formulas in the past [17,18,20]. For this study, three formulas for motion smoothness were computed. The Hofstad computation [18] was not significant and we choose to not represent it. The Cotin computation [17] was just significant for this study. More interesting, our results of Spectral Arc Length Motion Smoothness were very significant and consistent with the results found by Balasubramanian et al., [20]. With these new types of metrics simply based on the simulator's raw trajectories, it was possible to significantly distinguish levels of expertise.

We demonstrated that clustering analysis with DTW was able to reliably discriminate between expert operators and intermediate operators in an unsupervised way. We suggest that this whole trajectory-based approach could be used to better

Journal Pre-proof

classify the performance level in a simulator. Cluster 2 is composed of the entire intermediate cohort and two experts. The declared ages of these two experts are near the median age of the intermediate cohort (31 years). Their presence in this cluster could be explained by a non-age-related experience. We did not find any impact of nationality on simulator performance. On the contrary, the intermediate cohort composed entirely of French participants is distinguished by the heterogeneity of the trajectories in the DTW classification.

Construct validity of the simulator had already been demonstrated for novices and experts[13],[22]. But, Stunt et al.,[22] faced difficulties to find significant differences in performance for the intermediate group (less than sixty arthroscopic surgeries as principal operator). Roberts et al.,[16] did not distinguish between surgical experience for a similar shoulder arthroscopic exercise with the performance metrics proposed by the simulator, questioning its interest for middle-grade trainees and consultant. Our study shown that it is possible for a VR arthroscopy simulator to distinguish between more levels of expertise by using new metrics or with clustering analysis with DTW.

The majority of the intermediate cohort had similar theoretical training in arthroscopy (84% had the French national diploma of arthroscopy). Nearly half of them had already received simulator training (46% vs. 27% for the experts). However, this group stands out from the experts because of the heterogeneity of their trajectories and their lower performance. It is therefore essential to combine practical simulator training with theoretical training. But, Manoharan et al.,[23] recently demonstrated the difficulty in the United Kingdom in creating a national training program for arthroscopic surgery. Walbron et al.,[6,24] reported on the efforts of the French College of Orthopaedic Surgery and Arthroscopic Society to propose a nationwide program of arthroscopic surgery learning with virtual reality simulator. First results were interesting but reached some limits due to the necessary but complicated organizational presence of a trainer with the student. Thus, creating a standardized nationwide program with regular monitoring of student performance remains complex in practice with only few successful examples in the world[25]. Immediate and direct feedback on performance seems mandatory with ongoing guidance rather than having the student work autonomously with occasional feedback during spaced sessions with the trainer.

Journal Pre-proof

We still need to demonstrate that the results for this method can be extrapolated to other exercises and simulators. Indeed, the *Catch the Stars Glenohumeral* exercise is identical for each attempt: the position as well as the order of appearance of the stars are the same; similarly, the anatomy and dimensions of the simulated shoulder are always the same. This invariance of the exercise makes the measurement of DTW very effective because each participant will have to use the same path. The experience level was only assessed by the number of arthroscopies performed as the principal operator, in line with previous studies' experience thresholds [13,22]. Although the exercise was conducted under similar conditions for the two cohorts (at a congress and not in an OR environment), the results may have been distorted by inherent conference distractions. This may explain the presence of some outliers in the study. Also mentioned by Fucentese et al,,[12] collecting data during congress might cause a selection bias by attracting participants who are highly interested in medical education and virtual reality training. Subgroup analysis for sports, video game and musical instrument practices were limited by the small sample size and the heterogeneity of the subgroups, we choose to not represent the results. Finally, this study is about technical skills of the surgeon only. Studying non-technical skills including cognitive and interpersonal skills is also of high importance[26].

5. Conclusion

This study demonstrated the Construct Validity of the VR shoulder arthroscopy simulator within groups of experienced surgeons. With new types of metrics simply based on the simulator's raw trajectories, it was possible to significantly distinguish levels of expertise. We demonstrated that clustering analysis with Dynamic Time Warping was able to reliably discriminate between expert operators and intermediate operators.

6. Acknowledgements

We thank all participants of the study, and the VirtaMed team, Martina Vitz and Benjamin Grauer for the technical support.

Journal Pre-proof

Conflicts of interest: The author(s) do not declare any inherent conflicts of interest in the study.
F. Sirveaux is a consultant for Tornier Wright Medical outside the submitted work.

Sources of funding: No funding was received to support of this work.

Author contribution:

Tronchot: Designed the methodology, collected, analysed, and visualized the data, wrote the manuscript. Acted as a project lead supervisor.

J. Berthelemy: Analysed, and Visualized the data, helped to write the manuscript.

H. Thomazeau: Conceptualized the study and supervised the study, designed the methodology, reviewed the paper.

A. Huaulmé: Designed the methodology, helped in the analysis of the data collected, reviewed the paper.

P. Walbron: Performed the investigation, collected the data, reviewed the paper.

F. Sirveaux: Conceptualized the study, performed the investigation, reviewed the manuscript.

P. Jannin: Conceptualized and supervised the study, designed the methodology, reviewed the paper

Journal Pre-proof

Journal Pre-proof

References

[1] Barger LK, Rosner B. Extended Work Shifts and the Risk of Motor Vehicle Crashes among Interns. N Engl J Med 2005:10.
[2] Erivan R, Chaput T, Villatte G, Ollivier M, Descamps S, Boisgard S. Ten-year epidemiological study in an orthopaedic and trauma surgery centre: Are there risks involved in increasing scheduled arthroplasty volume without increasing resources? Orthop Traumatol Surg Res 2018;104:1283–9. https://doi.org/10.1016/j.otsr.2018.08.009.
[3] Bonnafous S, Amiot-Chanal H, Rouanet P. Arrêté du 12 avril 2017 portant organisation du troisième cycle des études de médecine. NOR: MENS1708241A, n.d.
[4] Banaszek D, You D, Chang J, Pickell M, Hesse D, Hopman WM, et al. Virtual Reality Compared with Bench-Top Simulation in the Acquisition of Arthroscopic Skill: A Randomized Controlled Trial. J Bone Jt Surg 2017;99:e34. https://doi.org/10.2106/JBJS.16.00324.
[5] Henn RF, Shah N, Warner JJP, Gomoll AH. Shoulder Arthroscopy Simulator Training Improves Shoulder Arthroscopy Performance in a Cadaveric Model. Arthrosc J Arthrosc Relat Surg 2013;29:982–5. https://doi.org/10.1016/j.arthro.2013.02.013.
[6] Walbron P, Common H, Thomazeau H, Hosseini K, Peduzzi L, Bulaid Y, et al. Virtual reality simulator improves the acquisition of basic arthroscopy skills in first-year orthopedic surgery residents. Orthop Traumatol Surg Res 2020:S187705682030116X. https://doi.org/10.1016/j.otsr.2020.03.009.
[7] Marcheix P-S, Vergnenegre G, Dalmay F, Mabit C, Charissoux J-L. Learning the skills needed to perform shoulder arthroscopy by simulation. Orthop Traumatol Surg Res 2017;103:483–8. https://doi.org/10.1016/j.otsr.2017.02.008.
[8] Baumann Q, Hardy A, Courage O, Lacombes P, Accadbled F. Lessons taught by a knee arthroscopy simulator about participants in a European arthroscopy training programme. Orthop Traumatol Surg Res 2019;105:S287–91. https://doi.org/10.1016/j.otsr.2019.09.008.
[9] Frank RM, Wang KC, Davey A, Cotter EJ, Cole BJ, Romeo AA, et al. Utility of Modern Arthroscopic Simulator Training Models: A Meta-analysis and Updated Systematic Review. Arthrosc J Arthrosc Relat Surg 2018;34:1650–77. https://doi.org/10.1016/j.arthro.2017.10.048.
[10] Aïm F, Lonjon G, Hannouche D, Nizard R. Effectiveness of Virtual Reality Training in Orthopaedic Surgery. Arthrosc J Arthrosc Relat Surg 2016;32:224–32. https://doi.org/10.1016/j.arthro.2015.07.023.
[11] Despinoy F, Zemiti N, Forestier G, Sánchez A, Jannin P, Poignet P. Evaluation of contactless human–machine interface for robotic surgical training. Int J Comput Assist Radiol Surg 2018;13:13–24. https://doi.org/10.1007/s11548-017-1666-6.
[12] Fucentese SF, Rahm S, Wieser K, Spillmann J, Harders M, Koch PP. Evaluation of a virtual-reality-based simulator using passive haptic feedback for knee arthroscopy. Knee Surg Sports Traumatol Arthrosc 2015;23:1077–85. https://doi.org/10.1007/s00167-014-2888-6.
[13] Rahm S, Germann M, Hingsammer A, Wieser K, Gerber C. Validation of a virtual reality-based simulator for shoulder arthroscopy. Knee Surg Sports Traumatol Arthrosc 2016;24:1730–7. https://doi.org/10.1007/s00167-016-4022-4.
[14] Bayona S, Fernández-Arroyo JM, Martín I, Bayona P. Assessment study of

Journal Pre-proof

insight ARTHRO VR ® arthroscopy virtual training simulator: face, content, and construct validities. J Robot Surg 2008;2:151–8. https://doi.org/10.1007/s11701-008-0101-y.
[15] Gomoll AH, O'Toole RV, Czarnecki J, Warner JJP. Surgical Experience Correlates with Performance on a Virtual Reality Simulator for Shoulder Arthroscopy. Am J Sports Med 2007;35:883–8. https://doi.org/10.1177/0363546506296521.
[16] Roberts PG, Guyver P, Baldwin M, Akhtar A, Alvand A, Price AJ, et al. Validation of the updated ArthroS simulator: face and construct validity of a passive haptic virtual reality simulator with novel performance metrics. Knee Surg Sports Traumatol Arthrosc 2017;25(2):616–25. https://doi.org/10.1007/s00167-016-4114-1.
[17] Cotin S, Stylopoulos N, Ottensmeyer M, Neumann P, Rattner D, Dawson S. Metrics for Laparoscopic Skills Trainers: The Weakest Link! In: Dohi T, Kikinis R, editors. Med. Image Comput. Comput.-Assist. Interv. — MICCAI 2002, vol. 2488, Berlin, Heidelberg: Springer Berlin Heidelberg; 2002, p. 35–43. https://doi.org/10.1007/3-540-45786-0_5.
[18] Hofstad EF, Våpenstad C, Chmarra MK, Langø T, Kuhry E, Mårvik R. A study of psychomotor skills in minimally invasive surgery: what differentiates expert and nonexpert performance. Surg Endosc 2013;27:854–63. https://doi.org/10.1007/s00464-012-2524-9.
[19] Oropesa I, Sánchez-González P, Lamata P, Chmarra MK, Pagador JB, Sánchez-Margallo JA, et al. Methods and Tools for Objective Assessment of Psychomotor Skills in Laparoscopic Surgery. J Surg Res 2011;171:e81–95. https://doi.org/10.1016/j.jss.2011.06.034.
[20] Balasubramanian S, Melendez-Calderon A, Burdet E. A Robust and Sensitive Metric for Quantifying Movement Smoothness. IEEE Trans Biomed Eng 2012;59:2126–36. https://doi.org/10.1109/TBME.2011.2179545.
[21] Keogh E, Ratanamahatana CA. Exact indexing of dynamic time warping. Knowl Inf Syst 2005;7:358–86. https://doi.org/10.1007/s10115-004-0154-9.
[22] Stunt JJ, Kerkhoffs GMMJ, van Dijk CN, Tuijthof GJM. Validation of the ArthroS virtual reality simulator for arthroscopic skills. Knee Surg Sports Traumatol Arthrosc 2015;23:3436–42. https://doi.org/10.1007/s00167-014-3101-7.
[23] Manoharan G, Sharma N, Gallacher P. Competence in using the arthroscopy stack system: a national survey of orthopaedic trainees in the UK. Ann R Coll Surg Engl 2020;102:149–52. https://doi.org/10.1308/rcsann.2019.0131.
[24] Walbron P, Common H, Thomazeau H, Sirveaux F. Evaluation of arthroscopic skills with a virtual reality simulator in first-year orthopaedic residents. Int Orthop 2020. https://doi.org/10.1007/s00264-020-04520-1.
[25] Strandbygaard J, Bjerrum F, Maagaard M, Rifbjerg Larsen C, Ottesen B, Sorensen JL. A structured four-step curriculum in basic laparoscopy: development and validation. Acta Obstet Gynecol Scand 2014;93:359–66. https://doi.org/10.1111/aogs.12330.
[26] Yule S, Flin R, Maran N, Rowley D, Youngson G, Paterson-Brown S. Surgeons' Non-technical Skills in the Operating Room: Reliability Testing of the NOTSS Behavior Rating System. World J Surg 2008;32:548–56. https://doi.org/10.1007/s00268-007-9320-z.

Journal Pre-proof

Figure

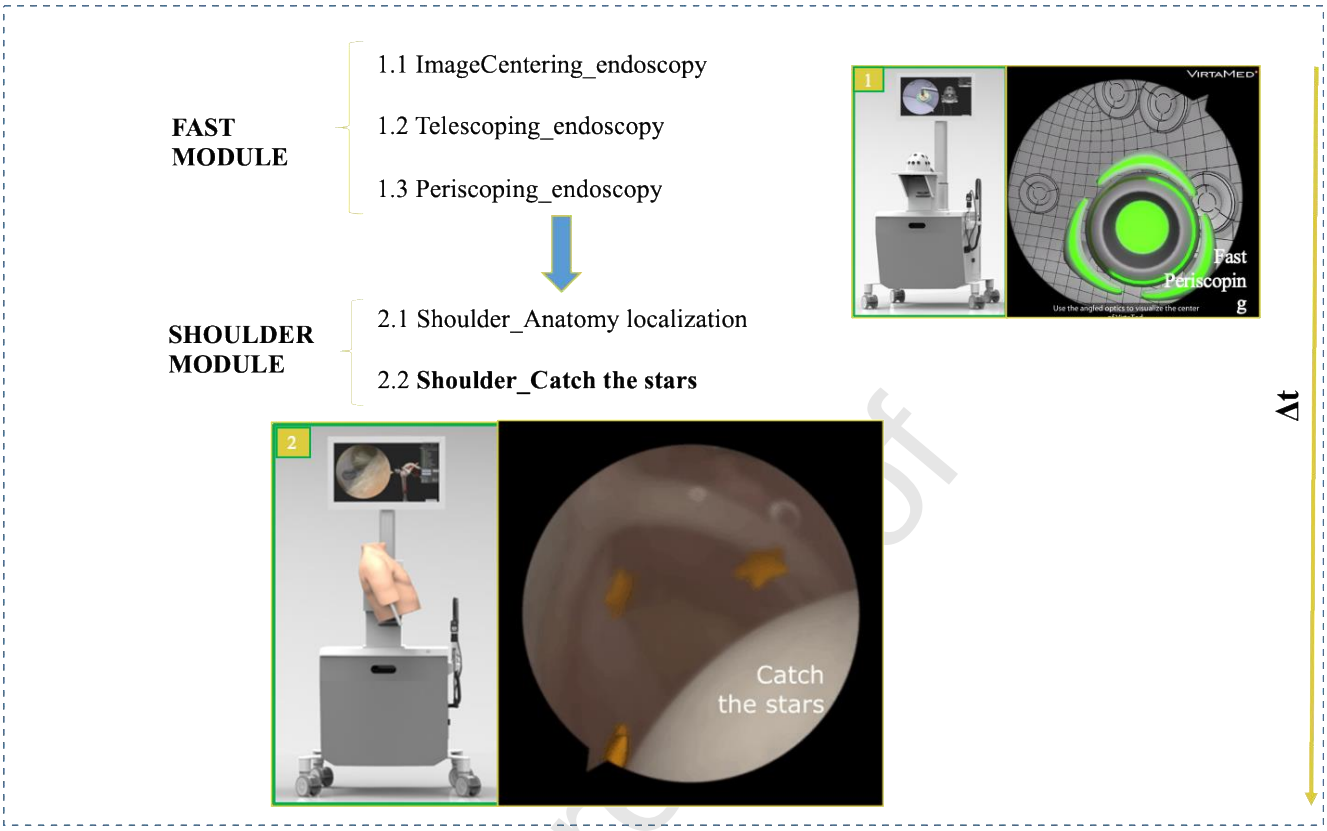

Figure 1. Standardized protocol (*1. The trainee is introduced to the simulator with the FAST (Fundamentals of Arthroscopy Surgery Training) module with 3 different exercices 1.1, 1.2 and 1.3; 2.1 The trainee is introduced to the shoulder anatomy of the SHOULDER module in lateral decubitus, 2.2. The trainee performs the task 3 times in lateral decubitus*)

	METRICS DEFINITION
Duration ***(DUR)***	Execution time between the first time the hand moves the probe, until the probe has been released at the end of the task. It is measured in seconds (s).
Path Length ***(PL)***	Sum of the total distances travelled by the probes (camera and grasper) during the execution of the task. It is measured in millimetres (mm).
Average Velocity ***(AV)***	Average linear speed of the probe during the task. It is measured in millimetres per second ($mm.s^{-1}$).
Average Acceleration ***(AA)***	Average instantaneous acceleration of the probe during the task. It is measured in millimetres per second squared ($mm.s^{-2}$).

Journal Pre-proof

Average Jerk ***(AJ)***	Average jerk (derivative of the acceleration) during the task, also known as "smoothness". It is measured in millimetres per second cubed ($mm.s^{-3}$).
Working volume ***(WV)***	Volume of the convex hull for each trajectory. It is measured in millimetres cubed (mm^3).
Economy of motion ***(ECO)***	Ratio between the working volume and the path length. It is measured in mm^2.
Motion Smoothness ***(HMS)***	Total change in acceleration of the tip of the instrument. It is measured in millimetres per second cubed ($mm.s^{-3}$) and is normalized by the duration of the task.
Spectral Arc Length Smoothness ***(SPARC)***	Smoothness of the linear movement speed profile using the spectral arc length method. Unitless measure.
Depth ***(D)***	Total distance travelled by the tip of the instrument along the z-axis. It is measured in millimetres (mm).

Figure 2. Metric definitions

Journal Pre-proof

Journal Pre-proof

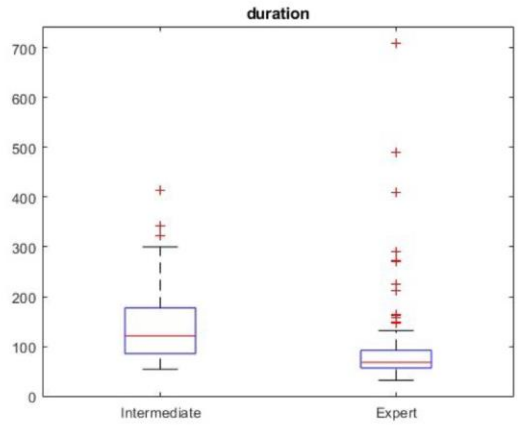

Figure 3.1. Duration

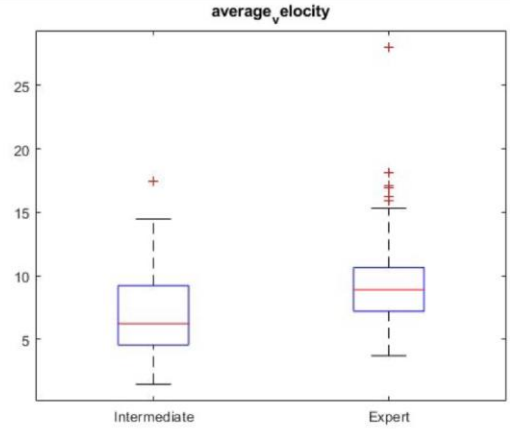

Figure 3.2. Velocity

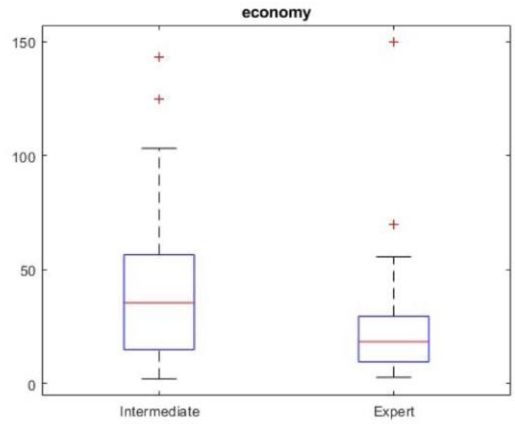

Figure 3.3. Economy

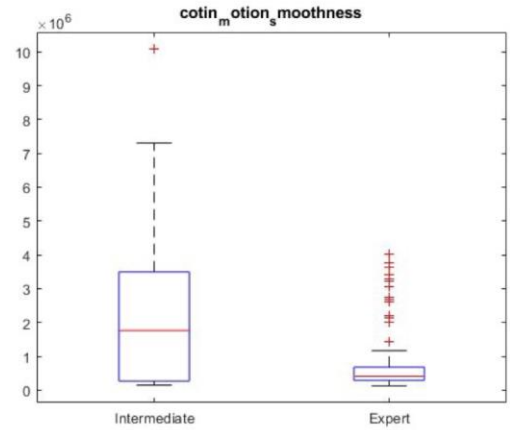

Figure 3.4. Motion Smoothness

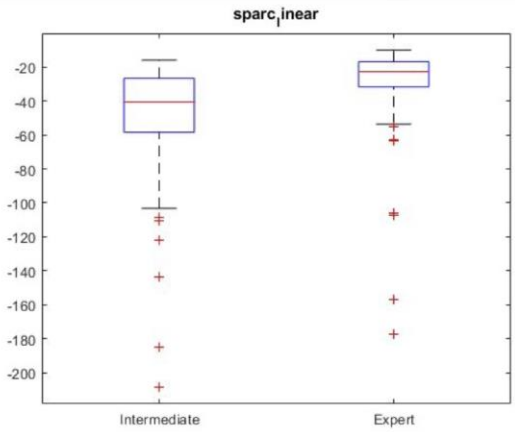

Figure 3.5. Spectral Arc Length Smoothness

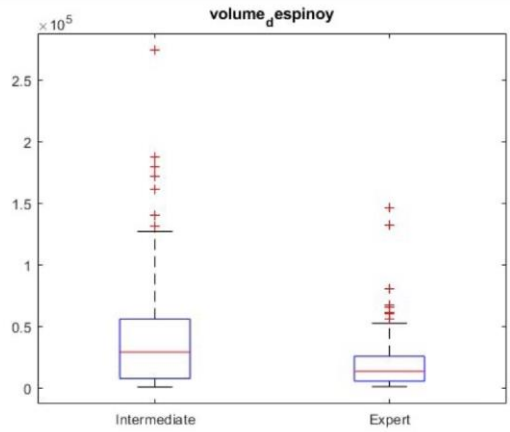

Figure 3.6. Working Volume

Figure 3. Metrics with significant differences box-plot (*3.1. Duration, 3.2. Velocity, 3.3. Working Volume, 3.4. Economy, 3.5. Motion Smoothness,* and *3.6. Spectral Arc Length Smoothness)*

Journal Pre-proof

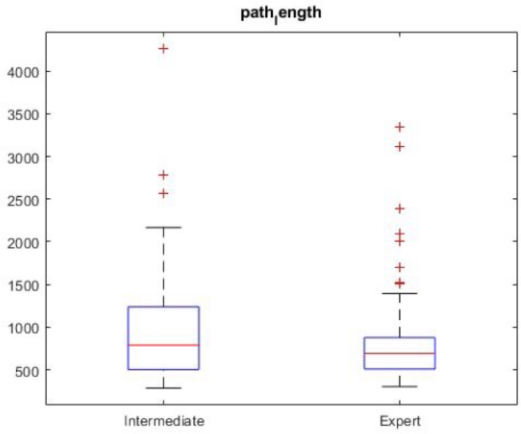

Figure 4.1. Path Length

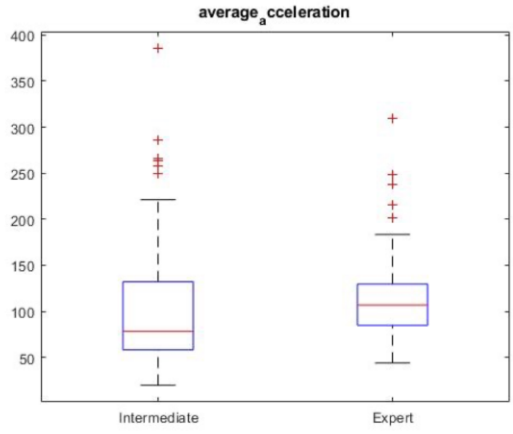

Figure 4.2. Acceleration

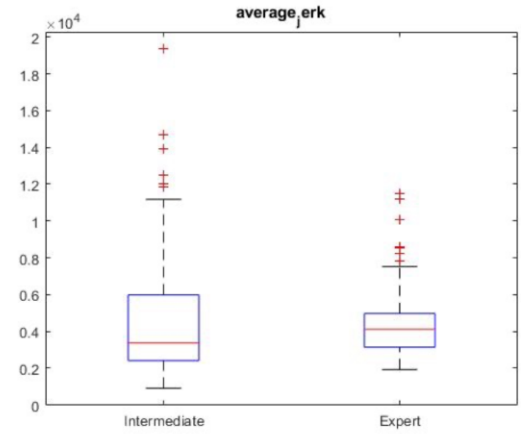

Figure 4.3. Jerk

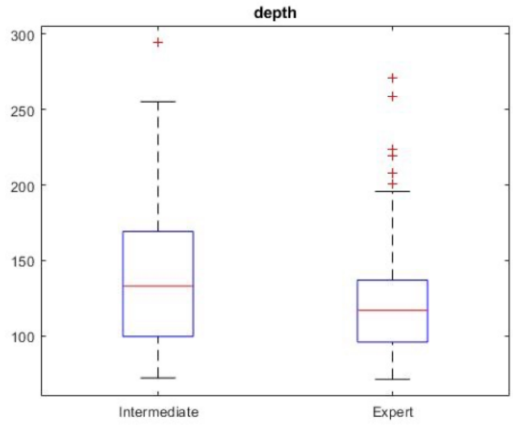

Figure 4.4. Depth

Figure 4. Metrics with no significant differences box-plot (*4.1. Path length, 4.2. Acceleration, 4.3. Jerk,* and *4.4. Depth)*

Journal Pre-proof

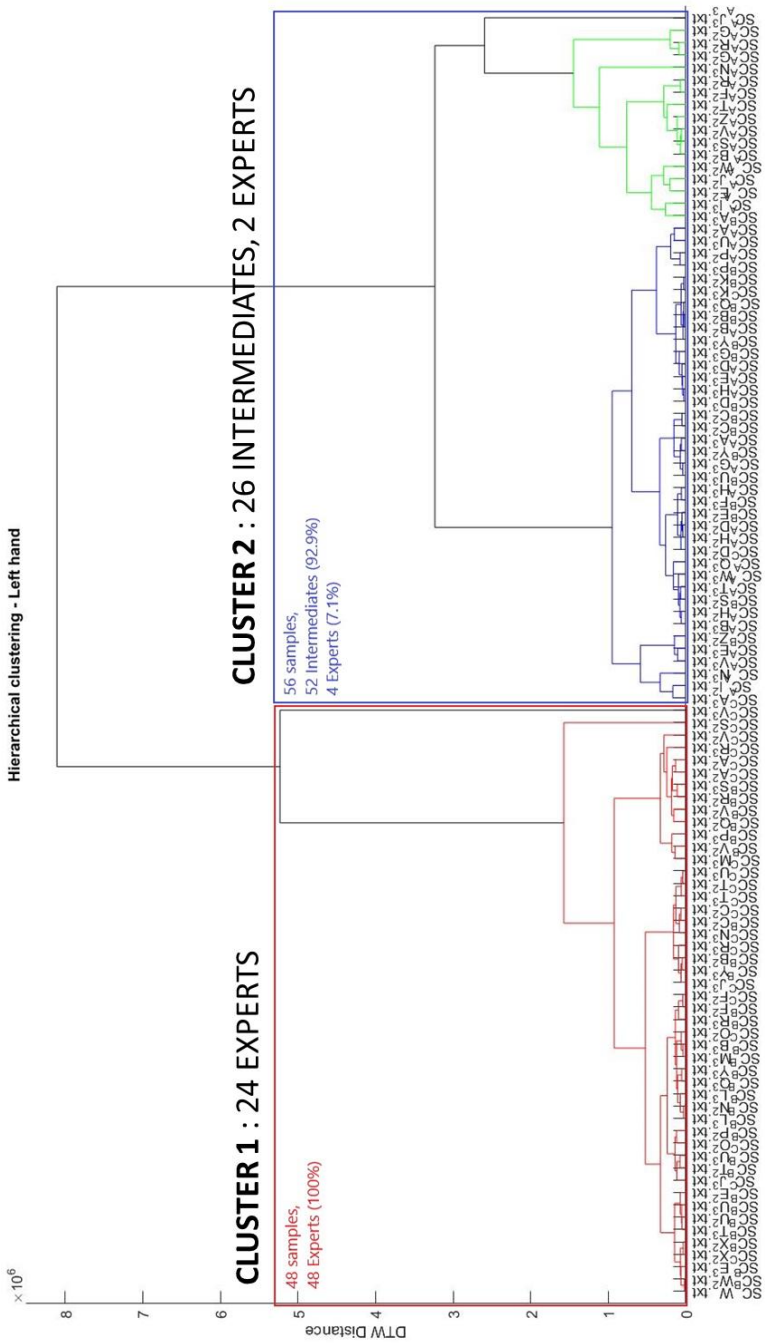

Figure 5. Hierarchical Clustering with Dynamic Time Warping Dendrogram (*The abscissa axis indicates the trajectories and the ordinate axis indicates the distance between merged clusters. The two main clusters are represented, cluster 1 is only composed of experts with low DTW distance, cluster 2 is composed of the intermediaries and 2 experts with more heterogeneous DTW distance.*)

Journal Pre-proof

	INTERMEDIATE (n=26)	EXPERT (n=26)	TOTAL (n=52)	P-VALUE
AGE, yr	30 [28-31]	41 [36-46]	33 [29-41]	< 0.001*
GENDER				1†
Male	23 (88.5%)	24 (92.3%)	47 (90.4%)	
Female	3 (11.5%)	2 (7.7%)	5 (9.6%)	
DOMINANT SIDE				0.6†
Right	24 (92.4%)	25 (96.2%)	49 (94.3%)	
Left	1 (3.8%)	0 (0.0%)	1 (1.9%)	
Ambidexter	1 (3.8%)	1 (3.8%)	2 (3.8%)	
COUNTRY				< 0.001†
France	26 (100.0%)	5 (19.2%)	31 (59.6%)	
Other	0 (0%)	21 (80.8%)	21 (40.4%)	
SPORT PRACTICE				0.01†
Yes	15 (57.7%)	24 (92.3%)	39 (75.0%)	-
No	11 (42.3%)	2 (7.7%)	13 (25.0%)	
VIDEO GAME PRACTICE				0.22†
Yes	10 (38.5%)	5 (19.2%)	15 (28.8%)	
No	16 (61.5%)	21 (80.8%)	37 (71.2%)	
MUSICAL INSTRUMENT PRACTICE				1†
Yes	7 (26.9%)	6 (23.1%)	13 (25%)	
No	19 (73.1%)	20 (76.9%)	39 (75.0%)	
SIMULATOR TRAINING EXPERIENCE				0.25†
Yes	12 (46.2%)	7 (26.9%)	19 (36.6%)	
No	14 (53.8%)	19 (73.1%)	33 (63.4%)	

Table 1. Baseline Characteristics of participants
NOTE. Data are reported as median [interquartile range] or number (percentage).
* *Independent-samples Student's t-test*
† *Pearson's chi-squared test*

Journal Pre-proof

METRICS	*Values (Median [Interquartile range (IQR), 25% - 75%]*		*p*†
	INTERMEDIATE	***EXPERT***	
Path_Length ***PL***	695mm [IQR, 467-1090]	671mm [504-846]	0.62
Duration ***DUR ****	116.1s [82.8-154.2]	70.9s [56.4-86.3]	<0.01
Average_Velocity ***AV ****	6.9mm.s^{-1} [4.7-9.5]	10.0mm.s^{-1} [7.4-11.6]	<0.01
Average_Acceleration ***AA***	92.3mm.s^{-2} [61.0-141.4]	116.6mm.s^{-2} [79.2-144.2]	0.09
Average_Jerk ***AJ***	3717mm.s^{-3} [2431-6087]	4211mm.s^{-3} [3017-5604]	0.31
Economy ***ECO ****	33.8mm^{2} [14.8-50.5]	19.3mm^{2} [9.1-25.6]	<0.01
Motion Smoothness ***HMS ****	1.5.10^{6}mm.s^{-3} [2.6E+5-3.5E+6]	4.6.10^{6}mm.s^{-3} [3.1E+5-7.2E+5]	0.05
Spectral Arc Length Smoothness ***SPARC ****	-37.8 [-60.7 - -26.5]	-21.2 [-31.0 - -16.8]	<0.01
Depth ***D***	123.8mm [97.1-155.0]	116.6mm [92.1-138.8]	0.11
Working Volume ***WV ****	28236mm^{3} [7193-52229]	12687mm^{3} [5753-22309]	<0.01

Table 2. Metric results for INTERMEDIATE and EXPERT categories, p-values

*Significant p-values are highlighted in grey and with an * mark.*

† Mann-Whitney U p-value

Journal Pre-proof