Observational Study

Medicine®

OPEN

# Surgical performance of large loop excision of the transformation zone in a training model

## A prospective cohort study

Günther A. Rezniczek, PhD[a,*], Sofia Severin, MD[a], Ziad Hilal, MD[a], Askin Dogan, MD[a], Harald Krentel, MD[b], Bernd Buerkle, MD[c], Clemens B. Tempfer, MD, MBA[a]

**Abstract**

Large loop excision of the transformation zone (LLETZ) is one of the most common procedures in operative gynecology and it is a routine part of the surgical training program of residents. There is, however, no established and standardized method of teaching residents how to perform LLETZ. Here, we present a surgical training model and assessed the improvement of surgical skills during repeated hands-on trainings of LLETZ in this model.

Surgical novices and experts were recruited and were shown a LLETZ training video and then performed 3 LLETZ training sessions on consecutive days. Surgical skills were assessed by Objective Structured Assessment of Technical Skills (OSATS). Global rating scale (GRS), confidence (CON), fragmentation rate (FR), performance time (PT), and OSATS scores were calculated. Intra- and interobserver variabilities were determined. The construct validity of OSATS was assessed comparing metric scores of novices with those of experts.

Sixty-eight probands (58 novices, 10 experts) were recruited. GRS, $2.3 \pm 1.3$ (median ± SD) versus $1.4 \pm 0.6$, $P < .001$; CON, $2.7 \pm 0.9$ versus $1.6 \pm 0.6$, $P < .001$; FR, 81% versus 100%, $P < .001$; PT, $152 \pm 33$ versus $120 \pm 27$ seconds, $P = .006$; and OSATS scores, $18.8 \pm 1.3$ versus $19.1 \pm 1.1$, $P = .16$ of novices improved from session 1 to session 3. OSATS showed construct validity with metric scores (GRS, $1.1 \pm 0.3$ vs $2.3 \pm 0.8$, $P < .001$; CON, $1.0 \pm 0.0$ vs $2.7 \pm 0.9$, $P < .001$; PT $125 \pm 30$ vs $152 \pm 33$ seconds, $P = .02$; OSATS scores, $19.6 \pm 0.7$ vs. $18.8 \pm 1.3$, $P = .02$) reliably discriminating between experts and novices. Intra- and interobserver variabilities across probands were $0.99 \pm 0.03$ and $0.64 \pm 0.10$, respectively. OSATS scores were independent of handedness, sex, and regular sports activity in univariate and multivariate analyses.

Repeated hands-on trainings improve surgical performance of LLETZ in a surgical training model with construct validity.

**Abbreviations:** CON = confidence, FR = fragmentation rate, GRS = global rating scale, LLETZ = large loop excision of the transformation zone, OSATS = Objective Structured Assessment of Technical Skills, PT = performance time.

**Keywords:** conization, dysplasia, LLETZ, OSATS, technical skills, training model

## 1. Introduction

Large loop excision of the transformation zone (LLETZ) is a standard technique for the surgical treatment of high-grade dysplasia of the cervix.[1] LLETZ is one of the most common procedures in operative gynecology and it is a routine part of the surgical training program of residents.[2,3] There is, however, no established and standardized method of teaching residents how to perform LLETZ. Typically, residents gain experience in a learning-by-doing approach with experts demonstrating and assisting procedures in a master and apprentice setting. This method is subjective and has been criticized for methodological and ethical reasons.[4] Developing standardized, reliable, and efficient methods of teaching gynecology residents how to perform surgical procedures such as LLETZ is an important and challenging task. Solid evidence regarding the teaching and training of LLETZ based on prospective trials is lacking (PubMed search, January 22, 2016; search terms: conization, LLETZ/LEEP, training, model, teaching, and dummy). Therefore, we have constructed and tested a simple and inexpensive LLETZ training model using 2 porcine sausages in an easy to construct, yet rigid and maintainable plastic assembly. The feasibility and validity of this training model has been tested and previously published.[5] In the current study, we aimed to establish the construct validity of this model. Furthermore, we wanted to test the hypothesis that repeated hands-on trainings of LLETZ using this surgical training model will significantly improve the surgical skills of novices.

In previous studies, we used the Objective Structured Assessment of Technical Skills (OSATS) method to evaluate and score the technical performance of hysteroscopy,[6] vaginal operative delivery by vacuum extraction,[7] and resolving a shoulder dystocia.[8] We and others found that OSATS is a

Editor: Bernhard Schaller.

The authors report no conflicts of interest.

Supplemental Digital Content is available for this article.

[a] Department of Obstetrics and Gynecology, Ruhr-Universität Bochum, Bochum, [b] Department of Obstetrics and Gynecology, St. Anna Hospital, Herne, [c] Department of Obstetrics and Gynecology, Klinikum Oldenburg, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany.

[*] Correspondence: Günther A. Rezniczek, Department of Obstetrics and Gynecology, Ruhr-Universität Bochum - Marien Hospital Herne, Düngelstraße 33, D-44623 Herne, Germany (e-mail: guenther.rezniczek@rub.de).



Medicine (2017) 96:23(e7026)

Received: 25 January 2017 / Received in final form: 3 May 2017 / Accepted: 8 May 2017

http://dx.doi.org/10.1097/MD.0000000000007026

reliable and reproducible method to objectively assess the technical skills of both experts and novices.[6–10] For example, Martinez *et al*[9] used OSATS to study the long-term effects of a training workshop regarding the theoretical knowledge and the practical skills of gynecology residents to repair fourth-degree lacerations. They found that residents improved on theoretical knowledge and OSATS scores after the training workshop and maintained this improvement for 6 months. Antomarchi *et al*[10] assessed the reliability of OSATS for evaluating a vaginal delivery simulation device. In their study, OSATS was a reliable means to assess medical students' competence in procedural skills using a simulator for vertex presentation delivery. Based on this evidence supporting OSATS to measure surgical skills,[6–10] we used a 20-item OSATS checklist to assess and quantify the technical skills of surgical novices performing a LLETZ procedure.

Although OSATS seems to be a reliable means for structured analysis of technical skills, this does not necessarily mean that it also has construct validity. Construct validity is defined as the degree to which a test measures what it claims, or purports, to be measuring.[11] To establish the construct validity of our study setting, we compared metric parameters such as performance time (PT), global rating scale (GRS), confidence (CON), and OSATS scores between surgical novices and surgical experts. We defined construct validity of our porcine LLETZ training model as given if all metric parameters significantly differed between experts and novices favoring the expert group. In this respect, we wanted to make sure that the results of the performance assessments of our study are representative and clinically meaningful. Moreover, we were interested in identifying the specific steps of the LLETZ procedure best defining who is an expert and who is not.

We hypothesized that our training model has construct validity and will lead to a measurable improvement in surgical skills. To test this hypothesis, we designed a prospective cohort study measuring OSATS scores of surgical novices repeatedly training a defined LLETZ algorithm after an expert demonstration session using a training video. In addition, we compared the performances of novices and surgical experts to establish construct validity.

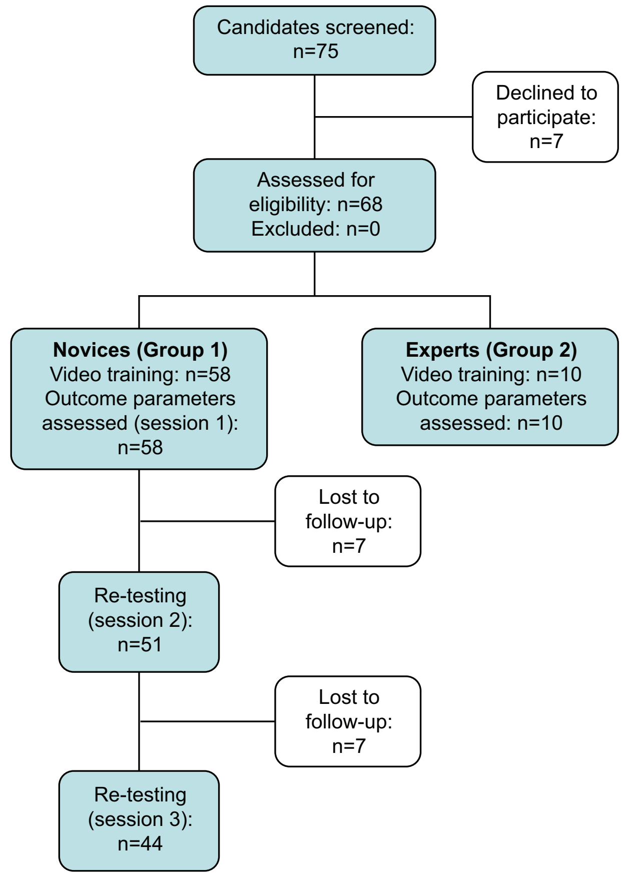


**Figure 1.** Diagram of the proband flow through the study.

## 2. Methods

This prospective cohort study was carried out at the Department of Obstetrics and Gynecology, Ruhr-Universität Bochum, Bochum, Germany, in a population of consecutive medical students, who took part in a gynecology rotation. In addition, expert surgeons (consultants or residents with at least 4 years of surgical experience) were recruited among the clinical staff of the Department. Approval for this study was obtained by the institutional review board of the Ruhr-Universität Bochum Medical Faculty (registration number 5200–15). Informed consent was obtained from all study probands. Figure 1 shows a diagram of the study probands' flow through the study. Inclusion criteria were informed consent and being a medical student without prior surgical training. Exclusion criteria were presence of a language barrier, previous exposure to LLETZ, and previous exposure to LLETZ training. Probands were shown a LLETZ training video (available as Supplemental Video 1, http://links.lww.com/MD/B717). This video was produced by our Department and is in use for the purpose of resident training. The video was originally recorded in German, but is available in a dubbed English version as an online supplementary file. In the video, one of the authors (CBT) acts as an expert instructor demonstrating all steps of performing a LLETZ according to the OSATS checklist used in this study. Table 1 shows all 20 OSATS items in detail. In the training video, we used the same LLETZ training model as the one used for the study. Figure 2 shows the model in detail. The feasibility and validity of a simpler, less robust version this sausage-based LLETZ training model has been previously published.[5] The key maneuvers of LLETZ were as follows: application of a 5% acetic acid solution to the cervix; identification of the acetowhite lesion; choice of the appropriate loop according to the size of the cervix, removal of the cone using adequate speed, and exploration of the cervical canal with a Hegar dilator. Additional tissue ("cowboy hat") was excised from the endocervix using a rectangular loop with a smaller diameter. Endocervical curettage was performed and hemostasis was obtained with a ball electrode sparing the cervical canal. We used high-frequency wire loops and ball electrodes (Erbe, Tübingen, Germany) for the procedures.

Probands were not allowed to film the procedures, to take photos, or to write notes. Immediately after the video instruction (session 1) and on consecutive days (sessions 2 and 3), probands were tested by 1 assessor (SS). All sessions were recorded on video. The outcomes GRS and CON were graded using a 5-item scale with lower values denominating better performance. Both GRS and CON were rated by the probands themselves and by the assessor directly after the procedure. PT was measured in seconds and any fragmentation of the cone was noted (fragmentation rate [FR], ie, the proportion of procedures where fragmentation occurred). In addition, the weights and heights of the excised

Table 1

**Objective structures assessment of technical skills checklist.**

| Steps: | Yes | No |
|---|---|---|
| 1. Application of a 5% acetic acid solution to the cervix | | |
| 2. Identification of the acetowhite lesion | | |
| 3. Grasping the cervix with a forceps outside of the acetowhite lesion | | |
| 4. Identification of the cervical canal using a Hegar dilator | | |
| 5. Choice of the appropriate loop size | | |
| 6. Proper holding of the loop's handle | | |
| 7. Excision of the cone using adequate speed | | |
| 8. Appropriate distance to vaginal wall is kept (heat damage) | | |
| 9. Removal of the cone using small forceps by maintaining specimen orientation; marking of orientation | | |
| 10. Check if the complete lesion is on the specimen | | |
| 11. Identification of the cervical canal using a Hegar dilator | | |
| 12. Choice of a rectangular loop with the smallest available diameter for the excision of additional endocervical tissue ('cowboy hat' configuration) | | |
| 13. Excision of the endocervical cone using adequate speed | | |
| 14. Removal of the endocone using small forceps | | |
| 15. Check if the cervical canal is identifiable on the specimen by placing a Hegar dilator through the canal | | |
| 16. Identification of the cervical canal using a Hegar dilator | | |
| 17. Endocervical curettage | | |
| 18. Achieving hemostasis with a ball electrode by taking care not to coagulate the cervical os | | |
| 19. Removal of forceps | | |
| 20. Coagulation of the forceps wounds | | |
| OSATS score (max. 20) | | |
| PT: seconds | | |
| GRS: 1 to 5* | | |
| CON: 1 to 5* | | |
| Additional data obtained: | | |
| Speed of cone excision is measured in steps 7 and 13. | | |
| Weights and depths of the cones are determined. | | |
| Cone fragmentation rates (FR) are noted. | | |

CON = confidence, GRS = global rating scale, OSATS = Objective Structured Assessment of Technical Skills, PT = performance time.
* Self-assessed and rater-assessed; see Fig. 3 for scale details.

cones were measured. To grade each proband's performance, the recorded videos were evaluated at a later time with the evaluator blinded to session number or group assignment. OSATS scores were calculated by adding points given for each of the 20 items on a task-specific check list (Table 1) with one point for correctly performing each item and zero points for not performing or not correctly performing the item. Thus, a higher score indicates greater proficiency.

We tested the intra- and interobserver variability of the OSATS score assessment by having 25 randomly chosen videos scored by four different raters with one of them rating all videos 3 times. Fleiss' kappa[12] was calculated based on the assumption that the mean proportion of agreement after random assignments by the raters was fixed at 0.5 in case of yes/no items and 0.2 in case of 5-point scales. In addition, the percentage of rater-agreement was calculated by pairwise comparisons of the ratings: for each disagreement, a disagreement score was increased by 1; agreement was then expressed as 1 – (disagreement score/maximally possible disagreement score).

The construct validity of the OSATS setting used in this study was assessed by comparing predefined metric scores between 10 experts and the 58 surgical novices recruited for this study. The metric scores were PT, GRS, CON, and OSATS scores. We defined construct validity as given, when all assessed metric scores were significantly different between the 2 groups favoring the experts.

We chose to recruit medical students, because they were naïve to any form of previous LLETZ training or practical experience with LLETZ. Thus, in our view, in contrast to residents, medical students were ideal subjects in the sense that they did not introduce bias in terms of different years of residency experience, past practical LLETZ experience, or previous LLETZ training experience.

When comparing different groups, categorical variables were analyzed by $\chi^2$ test and continuous variables were compared using the Mann–Whitney $U$ test with a significance level of .05. Wilcoxon signed rank or paired $t$ tests were used for the analysis of repeated measures data. We performed a multiple linear regression model to test whether the training effect, as measured by OSATS scores, was independent of potential confounders such as the probands' gender (male vs female) or handedness. Values are given as means. As this was not a comparative study, power and sample size calculations were not performed. We used the statistical software SigmaPlot 12.5 (Systat Software Inc., San Jose, CA) for statistical analysis.

## 3. Results

We recruited 68 of 75 screened probands for this study (7 declined to participate); 58 were novices and 10 were experts. Characteristics of the study population are given in Table 2. To assess whether repeated LLETZ trainings improved surgical performance of novices, we compared GRS, CON, FR, PT, and OSATS scores in 3 consecutive training sessions. We found that the surgical performance continuously improved over time and with increasing number of trainings. Table 3 shows all performance scores of novices in sessions 1 to 3. Specifically, rater-assessed GRS (rater-assessed: $2.3 \pm 1.3$ vs $1.4 \pm 0.6$, $P < .001$; self-assessed: $2.4 \pm 0.8$ vs $2.1 \pm 0.7$, $P = .001$), CON (rater-assessed: $2.7 \pm 0.9$ vs $1.6 \pm 0.6$, $P < .001$; self-assessed: 2.6

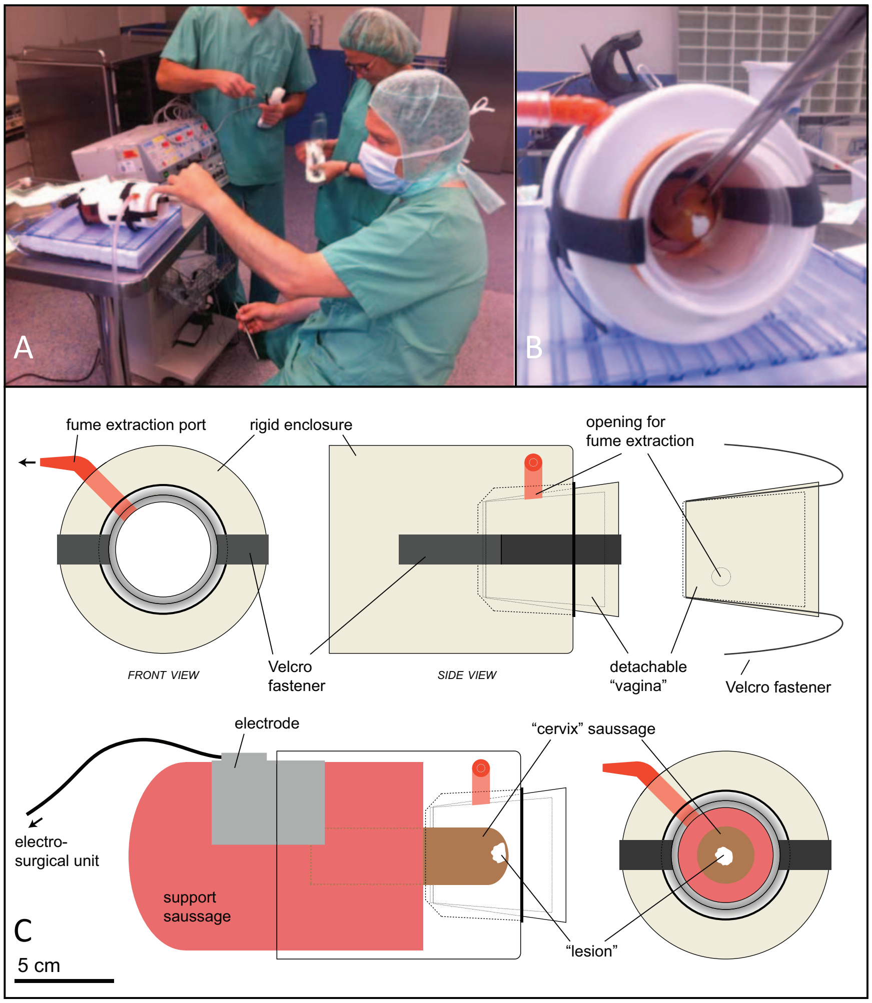


Figure 2. Training model. A, The model in use during a training session with gynecology residents. B, Close-up view of the model with an approx. 3-cm-diameter sausage representing the cervix with an “acetowhite lesion” (painted with white correction fluid), accessible through a cup representing the vagina held open by a speculum, and grasped with forceps. C, Schematic drawings of the model. The enclosure is made from a rigid plastic container. A hole is cut into the bottom and fitted with an inwall socket whose bottom was also cut out. Velcro strip are fastened to the sides of the enclosure and a hole is drilled in the side to mount a fitting for a vacuum line to suck out fumes. The “vagina” is detachable to facilitate access and cleaning (it will become quite greasy), and is made from 2 plastic cups (bottoms cut out) with Velcro strips glued in between the cups (aligned to fit the strips on the outer enclosure) and a matching hole for the fume exhaust. Two types of sausages are used: a large (approx. 9–10 cm diameter) support sausage, to which the return electrode is attached, and a smaller type of sausage with approx. 3 cm diameter that represents the cervix and is tightly inserted into a cavity scooped from the support sausage (conductivity between the sausages must be ensured). The sausage assembly is inserted from the back end of the rigid enclosure and both are fastened to a support stand using Velcro strips. After each procedure, the “vagina” is detached and the “cervix” can easily be replaced with a new one. The support sausage can be used for many procedures over several days when kept in a refrigerator in between sessions.

$\pm 0.8$ vs $2.1 \pm 0.9$, $P < .001$), PT ($152 \pm 33$ vs $120 \pm 27$ seconds; $P = .006$), and OSATS scores ($18.8 \pm 1.3$ vs $19.1 \pm 1.1$; $P = .16$) of novices improved from session 1 to session 3. Figure 3 demonstrates that OSATS scores, PT, self-, and rater-assessed CON and GRS scores of novices continuously improved during all 3 training sessions, but never reached the level of experts. Of note, both experts and novices scored better in rater-assessed GRS and CON scores than in self-assessed GRS and CON scores, indicating that both novices and experts tend to underrate their own performance.

**Table 2**

**Characteristics of study probands.**

| | Novices | Experts | *P* |
|---|---|---|---|
| N | 58 | 10 | |
| Age, y | 27.0 ± 4.6 (24.9; 22.4–41.4) | 42.2 ± 8.3 (41.6; 33.7–60.1) | <.001 |
| Sex, male/female | 16 (28%)/42 (72%) | 6 (60%)/4 (40%) | .046 |
| Right/left handed | 55 (95%)/3 (5%) | 10 (100%)/0 (0%) | .48 |
| Regular sports activity | 31/58 (53%) | 0/10 (0%) | .002 |
| Curriculum type (model/regular) | 4 (7%)/54 (93%) | – | |

Note: Values are reported as mean ± SD (median; range), absolute numbers (percentage), or fractions (percentage); *P*-values were calculated using the Mann–Whitney *U* test.

**Table 3**

**Comparison of performance scores in training sessions 1–3 in surgical novices.**

| | Session 1 | Session 2 | Session 3 | P* |
|---|---|---|---|---|
| N | 58 | 51 | 44 | |
| GRS | | | | |
| Self-assessed | 2.4±0.8 (2; 1–5) | 2.2±0.6 (2; 1–3) | 2.1±0.7 (2; 1–3) | .001[a] |
| Rater-assessed | 2.3±1.3 (2; 1–4) | 1.9±0.7 (2; 1–3) | 1.4±0.6 (1; 1–3) | <.001[a] |
| CON | | | | |
| Self-assessed | 2.6±0.8 (3; 1–4) | 2.4±0.8 (2; 1–4) | 2.1±0.9 (2; 1–4) | <.001[a] |
| Rater-assessed | 2.7±0.9 (3; 1–5) | 2.2±0.9 (2; 1–4) | 1.6±0.6 (2; 1–3) | <.001[a] |
| Ectocone | | | | |
| Complete removal of the acetowhite area | 47/58 (81%) | 48/51 (94%) | 44/44 (100%) | <.001[a] .006[c] |
| FR | 1/58 (2%) | 1/51 (2%) | 2/44 (5%) | .50[a] .84[c] |
| Weight, mg | 1053±606 | 968±658 | 1108±714 | .64[b] |
| Height, mm | 6.3±2.2 | 5.2±2.2 | 5.8±2.2 | .08[b] |
| Cut duration, s | 3.9±1.5 | 3.6±1.2 | 3.5±1.2 | .12[b] |
| Endocone | | | | |
| Cervical canal included | 51/58 (88%) | 47/51 (92%) | 41/44 (93%) | >.99[a] .58[c] |
| FR | 9/58 (16%) | 9/51 (18%) | 5/44 (11%) | .82[a] .75[c] |
| Weight, mg | 442±295 | 493±342 | 427±253 | .41[a] |
| Height, mm | 5.1±2.4 | 4.9±2.2 | 4.9±2.1 | .71[b] |
| Cut duration, s | 3.3±1.1 | 3.5±1.9 | 3.1±1.0 | .17[b] |
| PT, s | 152±33 | 133±32 | 120±27 | <.001[b] |
| OSATS score | 18.8±1.3 | 19.0±1.1 | 19.1±1.1 | .16[a] |

Note: Values are reported as mean±SD (median; range) or ratios (percentage).

* *P* values compared training session 1 versus 3 and were calculated using the Wilcoxon signed rank test,[a] or paired *t* test (two-tailed)[b] after testing for normality according to Shapiro–Wilk, or $\chi^2$ test (with Yates' correction).[c]

CON=confidence, FR=fragmentation rate, GRS=global rating scale, OSATS=Objective Structured Assessment of Technical Skills, PT=performance time.

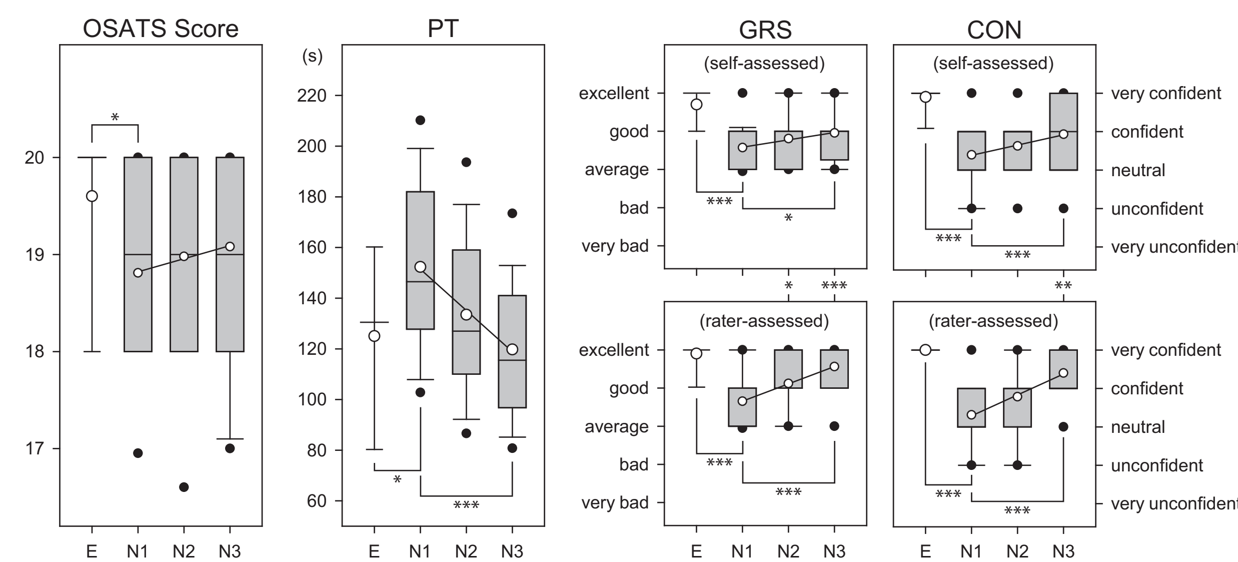


**Figure 3.** Main results. Total Objective Structured Assessment of Surgical Skills (OSATS) score, performance time (PT; in seconds), global rating scale (GRS) and confidence (CON) of the expert group (E) and the novices at training sessions 1, 2, and 3 (N1–3) are shown; numbers in parentheses indicate the number of probands. Box plots: thick lines indicate medians, boundaries, whiskers, and filled circles the 25th/10th/5th and the 75th/90th/95th percentiles, respectively; in case of the experts (N=10), the medians and ranges are indicated instead. White circles are means, lines between N1-N3 show linear regressions through the means. Statistical significant differences between groups (Mann–Whitney *U* test; PT: *t* test) or over time (Wilcoxon signed rank test; PT: paired *t* test) are indicated by brackets within panels (in case of GRS/CON also vertical lines connecting panels); levels of significance are as follows: *, $P<.05$; **, $P<.01$; ***, $P \leq .001$.

We tested the construct validity of the OSATS model used in this study by comparing metric and nonmetric scores of experts and novices. We found that our model reliably discriminated between experts and novices regarding GRS, CON, PT, and OSATS scores. Specifically, GRS ($1.1 \pm 0.3$ vs $2.3 \pm 0.8$; $P < .001$), CON ($1.0 \pm 0.0$ vs $2.7 \pm 0.9$; $P < .001$), PT ($125 \pm 30$ vs $152 \pm 33$ seconds; $P = .02$), and OSATS scores ($19.6 \pm 0.7$ vs $18.8 \pm 1.3$; $P = .02$) were significantly different between experts and novices, all favoring the expert group. Table 4 shows GRS, CON, PT, and OSATS scores and other scores broken down by surgical expertise group. In addition, we tested the intra- and interobserver variabilities of the OSATS score assessment by having a subset of 25 randomly chosen videos scored by 4 different raters with one of them rating all videos 3 times. We found that the intra- and interobserver variabilities were low with agreement scores of $0.99 \pm 0.03$ and $0.64 \pm 0.10$ across subjects, and $0.99 \pm 0.02$ and $0.63 \pm 0.32$ across items (OSATS including GRS and CON). Corresponding Fleiss' kappa values were $0.86 \pm 0.07$ and $0.49 \pm 0.13$, and $0.97 \pm 0.07$ and $0.49 \pm 0.44$, respectively. Figure 4 shows a graphical depiction of the agreement scores and Fleiss' kappa values broken down by each of the OSATS score items and GRS and CON, and broken down by study proband.

In surgical novices, we assessed sex (female; n = 42 vs male; n = 16), handedness (right-handed; n = 55 vs left-handed; n = 3), regular sports activities (defined as at least 3 times/week; yes; n = 31 vs no; n = 27), and the type of Medical School Curriculum (regular curriculum; n = 54 vs an experimental 'model' curriculum; n = 4). Assessing these potential confounders, we found that OSATS scores were independent of handedness, sex, regular sports activity, or type of curriculum in univariate and multivariate analyses (Table 5).

## 4. Discussion

In this prospective trial, we found that repeated hands-on trainings improved the surgical skills of LLETZ in a surgical training model. The OSATS used for assessing surgical performance showed construct validity. This study adds to the literature supporting the use of training models for improving the surgical performance of residents. In addition to other procedures in the field of Obstetrics and Gynecology such as hysteroscopy,[6] vaginal operative delivery by vacuum extraction,[7] or resolving a shoulder dystocia,[8] the current study suggests that LLETZ can also be effectively trained by using a dummy model. Based on the results of our study, we suggest that implementing this tool into clinical practice might improve the technical skills of residents, shorten their learning curve in the operating room, and might increase their motivation assuming that they feel more confident and have more satisfactory experiences during their first steps in the surgical theater. Furthermore, it may also improve patient safety.

Simulator trainings are time- and cost-intensive procedures. Raque *et al*,[13] for example, analyzed the cost-effectiveness of a virtual reality simulator program. They found that the use of an endoscopic simulator increased the costs of training without accelerating the learning curve or decreasing faculty time spent for instruction. Therefore, inexpensive and effective simulators should be developed and rigorously tested in clinical trials, if simulator trainings are to be implemented into the clinical routine. Therefore, in the present trial, we tested a low-tech, inexpensive, and easy to use LLETZ simulator, based on previous experience of our own group and of others.[5,14–16] We found that this low-tech tool is an effective means to train surgical novices in the technique of LLETZ. In contrast to other training gadgets such as virtual reality programs, the LLETZ training tool used in our study has a cost level < 10 USD to construct, can be easily assembled, and does not infer significant running costs.

One of the aims of our study was to establish the construct validity of the LLETZ training model used in the study. Therefore, we compared metric and nonmetric parameters between surgical novices and surgical experts. We found that construct validity of our porcine LLETZ training model was established, because parameters such as PT, GRS, CON, OSATS scores, and cutting time significantly differed between experts and novices favoring the expert group. Based on these results, we assume that this training model is representative of the real performance capacity of probands and is able to produce clinically meaningful results. In addition, we were interested in identifying the specific steps of the LLETZ procedure best

**Table 4**

**Construct validity of the training model.**

| | Experts | Novices (Session 1) | $P^*$ |
|---|---|---|---|
| N | 10 | 58 | |
| GRS (rater-assessed) | 1.1 ± 0.3 (1; 1–2) | 2.3 ± 0.8 (2; 1–4) | <.001[a] |
| CON (rater-assessed) | 1.0 ± 0.0 (1; 1–1) | 2.7 ± 0.9 (3; 1–5) | <.001[a] |
| Ectocone | | | |
| Complete removal of the acetowhite area | 9/10 (90%) | 47/58 (81%) | .68[b] |
| FR | 0/10 (0%) | 1/58 (2%) | >.99[b] |
| Weight, mg | 1255 ± 575 | 1053 ± 606 | .27[a] |
| Height, mm | 6.2 ± 2.7 | 6.3 ± 2.2 | .77[a] |
| Cut duration, s | 7.2 ± 3.0 | 3.9 ± 1.5 | <.001[a] |
| Endocone | | | |
| Cervical canal included | 8/10 (80%) | 51/58 (88%) | .61[b] |
| FR | 3/10 (30%) | 9/58 (16%) | .36[b] |
| Weight, mg | 345 ± 220 | 442 ± 295 | .30[a] |
| Height, mm | 3.8 ± 1.5 | 5.1 ± 2.4 | .12[a] |
| Cut duration, s | 4.6 ± 1.2 | 3.3 ± 1.1 | .001[a] |
| PT, s | 125 ± 30 | 152 ± 33 | .017[c] |
| OSATS score | 19.6 ± 0.7 (20; 18–20) | 18.8 ± 1.3 (19; 12–20) | .024[a] |

Note: Metric and nonmetric scores of surgical experts and surgical novices were compared. Values are reported as mean ± SD (median; range) or ratios (percentage).

[*] $P$ values: Mann–Whitney rank sum test,[a] Fisher exact test (two-tailed),[b] and $t$ test (two-tailed).[c]

CON = confidence, FR = fragmentation rate, GRS = global rating scale, OSATS = Objective Structured Assessment of Technical Skills, PT = performance time.

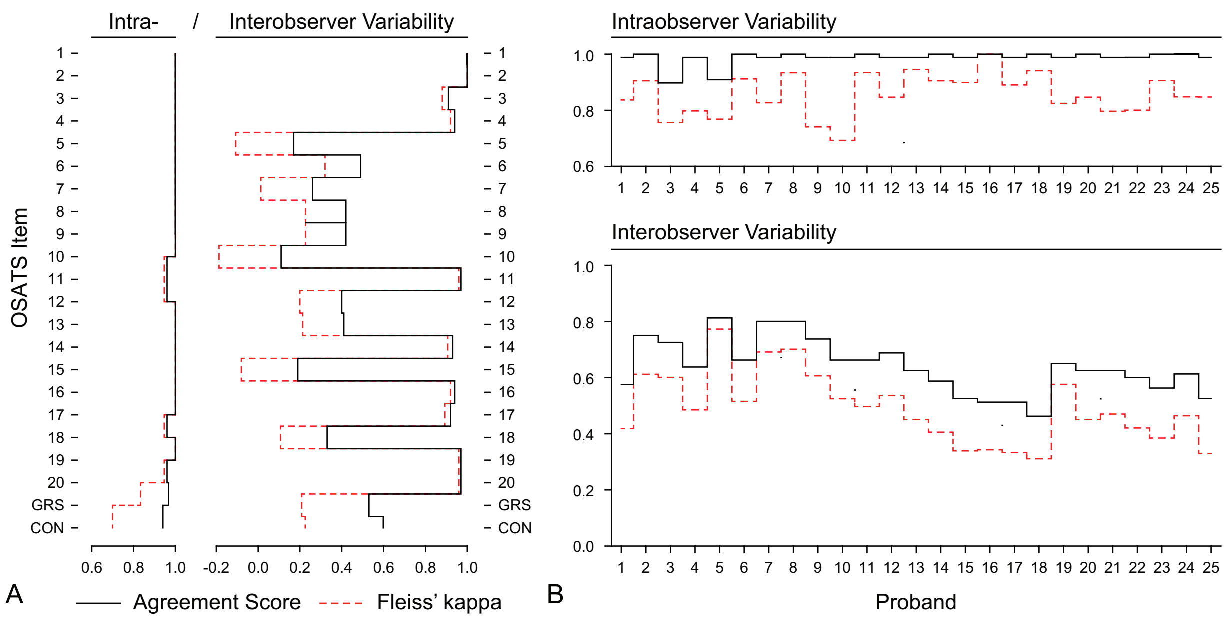


**Figure 4.** Intra- and interobserver variabilities. Agreement score (solid lines; as described in the methods section) and Fleiss' kappa (dotted lines) are shown broken down by each of the 20 Objective Structured Assessment of Surgical Skills (OSATS) score items, global rating scale (GRS), and confidence (CON) (panel A) and by proband (panel B).

**Table 5**

**Dependence of scores, global rating scale, and confidence on proband characteristics.**

| | Proband Characteristic | | | |
|---|---|---|---|---|
| **Sex** | **m (N=16)** | **f (N=42)** | ***P* (univariate)** | ***P* (multivariate)** |
| Total OSATS score | 18.5±2.0 | 18.9±0.9 | .70 | .21 |
| GRS | | | | |
| Self-assessed | 2.5 (1–3) | 2 (1–5) | .81 | .69 |
| Rater-assessed | 3 (1–4) | 2 (1–4) | .19 | .13 |
| CON | | | | |
| Self-assessed | 2.5 (1–4) | 3 (1–4) | .54 | .85 |
| Rater-assessed | 3 (2–4) | 2.5 (1–5) | .66 | .49 |
| **Handedness** | **right (N=55)** | **left (N=3)** | | |
| Total OSATS score | 18.8±1.3 | 18.3±0.6 | .17 | .40 |
| GRS | | | | |
| Self-assessed | 2 (1–5) | 3 (2–4) | .23 | .21 |
| Rater-assessed | 2 (1–4) | 3 (1–3) | .85 | .94 |
| CON | | | | |
| Self-assessed | 3 (1–4) | 4 (2–4) | .16 | .13 |
| Rater-assessed | 3 (1–5) | 4 (2–4) | .24 | .24 |
| **Curriculum** | **regular (N=54)** | **model (N=4)** | | |
| Total OSATS score | 18.8±1.3 | 18.5±0.6 | .24 | .57 |
| GRS | | | | |
| Self-assessed | 2 (1–5) | 2 (2–3) | .53 | .70 |
| Rater-assessed | 2 (1–4) | 2 (2–3) | .69 | .84 |
| CON | | | | |
| Self-assessed | 3 (1–4) | 2.5 (2–4) | .82 | .61 |
| Rater-assessed | 3 (1–5) | 2 (2–3) | .29 | .38 |
| **Sports Activity** | **no (N=27)** | **yes (N=31)** | | |
| Total OSATS score | 18.9±0.8 | 18.8±1.6 | .49 | .74 |
| GRS | | | | |
| Self-assessed | 2 (1–3) | 3 (1–5) | .18 | .12 |
| Rater-assessed | 2 (1–4) | 3 (1–4) | .15 | .10 |
| CON | | | | |
| Self-assessed | 2 (1–4) | 3 (1–4) | .11 | .10 |
| Rater-assessed | 2 (1–5) | 3 (1–4) | .06 | .06 |

Note: Values are reported as mean±SD or median (range); *P* values were calculated using Kruskal–Wallis one-way ANOVA on ranks for univariate assessment and multiple linear regressions for multivariate assessment. No statistically significant associations were found.
CON=confidence, GRS=global rating scale, OSATS=Objective Structured Assessment of Technical Skills.

defining who is an expert and who is not. GRS, CON, and cutting time were the items most significantly differing between experts and novices. Thus, one of the most significant characteristics of experts was that they used a well-dosed and slow cutting speed, whereas novices acted much more hastily. Thus, using an appropriate cutting speed should be emphasized when teaching novices how to perform LLETZ.

Based on a PubMed literature search (January 22, 2016; search terms: conization, LLETZ/LEEP, training, model, teaching, dummy), no data on the practical usefulness or the procedural efficacy of LLETZ training models have been published. Thus, our study is the first report indicating that LLETZ can be effectively trained before residents are allowed to perform this procedure in real patients in the operating room. Our study, however, has limitations. First, we have tested medical students, because they were naïve to LLETZ and other surgical procedures. On the other hand, obstetrics and gynecology residents might be different from this study population regarding motivation, self-selection, professional attitude, and procedure-specific theoretical knowledge. We have not assessed prior theoretical knowledge, a potential confounder, in our novice probands (eg, through a quiz taken before discussing any procedure-specific items or showing the training video), but it is unlikely that this plays a significant role as practical skills and (theoretical) knowledge do not necessarily correlate with each other and the effectiveness of the training model for increasing the practical skills of the probands was clearly demonstrated, not so much through an increase in the OSATS score, but by the significantly decreased procedure time, quality of conization (removal of the lesion), and rater and self-assessed global rating and confidence. Furthermore, our study does not guarantee that an improvement in LLETZ performance on a training model does also translate into an improved surgical performance in the operating room. Lastly, an improvement over 3 days might not be representative of a long-term improvement. Repeated LLETZ trainings over a long period might be necessary to reach the ultimate flattening of the learning curve.

Loss to follow up in our study was considerable with 24%. This might be because of the fact that medical students are less motivated to learn LLETZ compared with obstetrics and gynecology residents. Moreover, we have not used financial or other compensations for study participation for ethical reasons. We have not used systematic measures of evaluation to find out the reasons for nonadherence. A likely reason, however, was that because of a restriction imposed by our institutional review board we could not test students during the hours allotted to their rotation, but in their free time, and some decided to have better things to do. However, as participation was strictly voluntary, we believe that there was no motivation-related selection bias.

## 5. Conclusion

In summary, the results of our study indicate that repeated hands-on trainings using a LLETZ simulation model help to achieve a significant improvement of technical performance of surgical novices, and that the LLETZ training tool used in this study has construct validity. Therefore, this study supports the incorporation of LLETZ training models and OSATS into educational curriculums.

## Acknowledgments

The authors would like to acknowledge the support by the Open Access Publication Funds of the Ruhr-Universität Bochum, Germany.

## References

[1] Martin-Hirsch PPL, Paraskevaidis E, Bryant A, et al. Surgery for cervical intraepithelial neoplasia. Cochrane Database Syst Rev 2013;12: CD001318.
[2] World Health OrganizationWHO guidelines for treatment of cervical intraepithelial neoplasia 2-3 and adenocarcinoma in situ: Cryotherapy, large loop excision of the transformation zone, and cold knife conization. World Health Organization, Geneva:2014.
[3] Lindeque BG. Management of cervical premalignant lesions. Best Pract Res Clin Obstet Gynaecol 2005;19:545–61.
[4] Nagendran M, Gurusamy KS, Aggarwal R, et al. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev 2013;8:CD006575.
[5] Hefler L, Grimm C, Kueronya V, et al. A novel training model for the loop electrosurgical excision procedure: an innovative replica helped workshop participants improve their LEEP. Am J Obstet Gynecol 2012;206:535.
[6] Alici F, Buerkle B, Tempfer CB. Objective Structured Assessment of Technical Skills (OSATS) evaluation of hysteroscopy training: a prospective study. Eur J Obstet Gynecol Reprod Biol 2014;178:1–5.
[7] Buerkle B, Rueter K, Hefler LA, et al. Objective Structured Assessment of Technical Skills (OSATS) evaluation of theoretical versus hands-on training of vaginal breech delivery management: a randomized trial. Eur J Obstet Gynecol Reprod Biol 2013;171:252–6.
[8] Buerkle B, Pueth J, Hefler LA, et al. Objective structured assessment of technical skills evaluation of theoretical compared with hands-on training of shoulder dystocia management: a randomized controlled trial. Obstet Gynecol 2012;120:809–14.
[9] Martinez A, Cassling C, Keller J. Objective Structured Assessment of Technical Skills to teach and study retention of fourth-degree laceration repair skills. J Grad Med Educ 2015;7:32–5.
[10] Antomarchi J, Delotte J, Jordan A, et al. Development and validation of an objective structured assessment of technical skill tool for the practice of vertex presentation delivery. Arch Gynecol Obstet 2014;290:243–7.
[11] Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull 1955;52:281–302.
[12] Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76:378–82.
[13] Raque J, Goble A, Jones VM, et al. The relationship of endoscopic proficiency to educational expense for virtual reality simulator training amongst surgical trainees. Am Surg 2015;81:747–52.
[14] Reeves KO, Young AE, Kaufman RH. A simple, inexpensive device for teaching the loop electrosurgical excision procedure. Obstet Gynecol 1999;94:474–5.
[15] Walters CL, Whitworth JM, Tyra SL, et al. Constructing a Novel Simple LEEP Training Model. J Grad Med Educ 2013;5:320–2.
[16] Connor RS, Dizon AM, Kimball KJ. Loop electrosurgical excision procedure: an effective, inexpensive, and durable teaching model. Am J Obstet Gynecol 2014;211:706.e1–e3.