

Adjusting Training Task Difficulty to Enhance Skill Transfer in Venipuncture: Evidence from Video Analysis

Jinling Yang¹, Minghui Yi¹, Fei Wu¹, Menghui Hu¹, Qinxi Yao², Hangdi Zhou², Ying Xiao², Yao Zhang³, Yun Wu³, Bin Zheng³

¹Medical Integration and Practice Center, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, People's Republic of China; ²School of Basic Medical Sciences, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, People's Republic of China; ³Surgical Simulation Research Lab, Department of Surgery, University of Alberta, Edmonton, Alberta, Canada

Correspondence: Jinling Yang, Medical Integration and Practice Center, Cheeloo College of Medicine, Shandong University, No. 44 Wenhua Xi Road, Jinan, Shandong, People's Republic of China, Email yjlxixi@126.com; Bin Zheng, Surgical Simulation Research Lab, Department of Surgery, University of Alberta, 162 Heritage Medical Research Centre, 11207 – 87 Ave NW, Edmonton, Alberta, T6G 2S2, Canada, Email bin.zheng@ualberta.ca

Introduction: We examine the impact of modifying the difficulty level of venipuncture at different training stages on the enhancement of skill transfer.

Methods: Medical students were recruited and assigned into either a control or an experimental group, both underwent 4 days of training. The control group practices venipuncture tasks under quiet and consistent environment. The experimental group were given pressures by the standardized patients or exposure to noise environment in the late training stages. The performance was videotaped. All trainees were assessed one day (Test 1) and one week (Test 2) post-training for skill transfer.

Results: At Test 1, the experimental group showed significantly shorter puncture time ($P = 0.042$) and a near-significant trend for skin disinfection time ($P = 0.055$), while total task time did not differ significantly. At Test 2, the experimental group demonstrated significantly shorter total task time ($P = 0.039$), skin disinfection time ($P = 0.023$), puncture time ($P = 0.035$), and tourniquet banding time ($P < 0.001$). No significant differences were observed in overall performance scores, but doctor-patient communications showed significant differences between groups at Test 1 ($P = 0.000$) and within-group changes in the control group ($P = 0.004$).

Conclusion: Evidence from video analysis suggests that adjusting the difficulty of training can enhance the skill transfer in the venipuncture task, primarily by improving procedural efficiency, which inspires us to explore more effective way to improve the training outcomes.

Keywords: venipuncture, skill acquisition, skill transfer, surgical training, simulation

Introduction

For health trainees to practice patient care with confidence, their training needs to encompass not only their theoretical knowledge but also mastering of a multitude of health procedures, many of which consist of multiple inseparable steps.^{1,2} Students should learn and complete each step gradually to become proficient in these procedures.³⁻⁵ Skillful performance of these steps is crucial to avoid unexpected harm to patients in clinical practice.⁶

To improve trainees' skills and prevent patient harm, it is imperative to structure the training syllabus in a way that optimizes trainee learning outcomes for multi-step health procedures using simulation training.^{7,8} Typically, simulators are available for health trainees to practice health skills repeatedly during their training.⁹⁻¹¹ Through repeated practice, trainees are expected to learn skill in a simulated training environment. The question is how much skills can be transferred to clinical setting because the skills learned in the simulation environment may not transfer effectively.^{12,13}

When trainees practice under an unchanging training condition in simulation, environmental feedback to them is relatively consistent.¹⁴ When trainees enter a new clinical environment, the external feedback and environment are



different, which impede their skill performance. Although previous research has demonstrated that varying training conditions enhances skill learning and adaptation compared to static practice.^{15–18} the optimal timing for introducing such variations remains poorly understood. Skill acquisition progresses through multiple stages, and it is unknown whether increasing task difficulty during the early stage (when learners are acquiring fundamental steps) or the late stage (after basic skills are consolidated) yields better transfer outcomes. This study addresses this gap by comparing a staged difficulty protocol – where challenge is introduced after basic skill consolidation – with a standard training protocol.

A brief review of the basic process of skill acquisition is necessary here. According to Fitts & Posner's three-stage theory, the early stage of skill learning (Cognitive) requires extensive decomposition of a complex skill, whole-body motor units are extensively activated and required a large amount of energy. Motor coordination is not yet established at this stage and cognitive load is relatively high. In the middle stage (Associative), a large amount of feedback needs to be absorbed, the motor units gradually establish effective connections and cooperative couplings, and the tool-body coupling is gradually strengthened. Cognitive load at this stage is level off and a trainee is observed an upward curve in this skill learning stage. In the later stage (Autonomous), the coupling of the trainee's motor units becomes robust and clear; performance is often automatic and required minimal cognitive engagement.^{19,20} Based on above theory, we believe that the task difficulty should reduce in the early stage and increase in the later stage, which may yield improved training outcomes, especially for skill transfer.

In this study, we chose the procedure of venipuncture and blood collection as the training task. Venipuncture was selected for several reasons. First, it is a skill required by most health trainees and is also a component of the General Practitioners' Qualifying Examination. Second, it is moderately difficult compared to other procedures—more complex than simple operations such as oxygen administration and sputum suction, yet less complex than advanced procedures like thoracentesis. Third, Venipuncture has multiple steps that can be quantified at multiple points in data collection, which enables quantification of training outcomes, particularly through video analysis. Specifically in this study, we reduced the training difficulty for trainees in their early training phase, where they were practiced with simulation in a quiet and distraction-free environment for two days. On day three, the venipuncture simulator was attached to the arm of a standardized patient (SP) who did not intervene trainee's practice. On day four, the training difficulty was increased by introducing noise to the training environment. SP in this phase occasionally rebuked or questioned trainee's performance to give them external interference. Venipuncture performance was performed by each trainee on one day and one week after training to verify skill transfer short- and mid-termly.

The hypotheses for this study were as follows: Trainees in the experimental group will demonstrate superior performance in venipuncture blood collection on both the first and seventh day after training, indicating improved skill acquisition and transfer compared to the control group.

Methods

The study was conducted at the Medical Simulation Center of Shandong University. The research methods and experimental details received approval from the Ethics Committee of the School of Clinical Medicine, Shandong University (SDULCLL2022-20).

Participants

We posted a recruitment advertisement on the WeChat group of medical students at Shandong University. It was written to recruit second- and third-year medical students with no experience in peripheral venipuncture. Students interested in participating contacted us to arrange a time for training.

A total of 19 medical students were enrolled in this study and randomly assigned to either the control group ($n = 10$) or the experimental group ($n = 9$) using a random number table. One participant in the experimental group was terminated due to personal reasons, resulting in a final sample of 18 students who completed the entire training and assessment protocol (control group: $n = 10$; experimental group: $n = 8$). The randomization process was conducted by an independent researcher not involved in the training or assessment.

The study employed a double-blind design. Participants were unaware of their group allocation throughout the study; they were informed that they would receive one of two different training protocols but were not told which protocol was

hypothesized to be superior. The evaluators who assessed procedural performance during the two on-site assessments were also blinded to group assignment. Students were identified only by code numbers during all assessments, and the evaluators had no access to any information linking the code numbers to group allocation. Therefore, both participants and evaluators remained blinded throughout all assessment phases.

Our study adhered to the principles outlined in the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of the School of Clinical Medicine, Shandong University (SDULCLL2022-20), and all participants provided informed consent prior to their involvement in the study.

Task

The venipuncture procedure used in this study is a multi-step procedure. In accordance with the standards for step-by-step venipuncture procedures, the trainees were required to perform a series of preoperative preparations, self-induction, checking and explaining with the patient, the puncture procedure (identifying the puncture site, applying a tourniquet, disinfecting the skin, checking the puncture needle, puncturing, and collecting a blood sample), and post-puncture handling (pressing on the puncture site, disposing of the specimen, documenting the procedure, and arranging the items).

Training Model

In this study, the control and experimental groups used the same brand of venipuncture arm models (Shandong Yifutai Manufacturer, Model SY/H008, Jinan, China). The portable venipuncture skin module used by trainees in the experimental group to learn the operation on the SP arm was also from this manufacturer (Shandong Yifutai Manufacturer, Model SY/W033, Jinan, China). All models had the same skin texture and vessel orientation, but the vein diameter of the portable venipuncture device was slightly thinner and more difficult to puncture.

Standardized Patients

This study involved three SPs, all of whom had contractual agreements with the Medical Simulation Center of Shandong University. All completed required training and were experienced in working with SPs. All SPs involved in the study were appropriately paid.

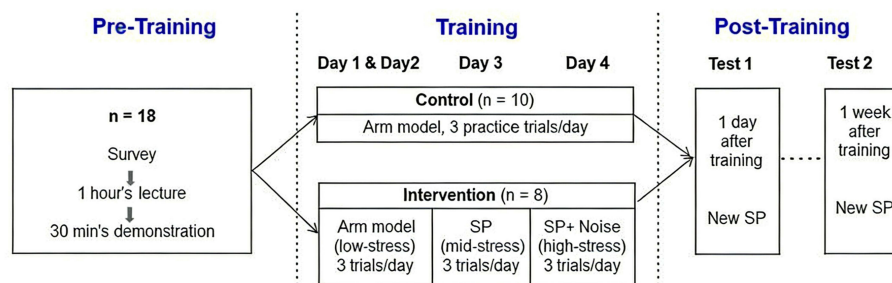
Training Procedures

Before the official start of the experiment, we organized a 90-minute course for all trainees in the conference room of the Medical Simulation Center, which consisted of a 60-minute theoretical lecture and a 30-minute on-site operational demonstration. The instructors were faculty members of the Medical Simulation Center with extensive clinical and training experience in venipuncture blood collection.

The training phase consisted of four days in a week. Each trainee was expected to perform three venipunctures per training day. The control group was provided with the same training model for all four days and practiced in a quiet room without environmental distractions. Trainees did not receive any instructional feedback over the training phase.

On training Day 1 and Day 2, the experimental group completed three maneuvers in the same arm vein puncture model as the control group, without any environmental distractions and receiving zero instructional feedback. On Day 3, the wearable venipuncture training device was attached to the arm of SP 1. The trainees were required to perform three maneuvers on the model without any intervention from SP 1. On Day 4, trainees in the experimental group repeated the same three maneuvers on the venipuncture model attached to the arm of SP 1; however, SP 1 on this training day questioned or blamed the trainee's performance twice. The script for accusations and verbal challenges was pre-determined (as shown in [Supplementary Box 1](#)). In addition, we intentionally disrupted the operation by playing various background sounds, including a baby cry and alarms of ambulance.

All trainees underwent two tests, one day and one week after the training period. Each trainee performed the operation only once on each testing day. The tests were done on the model attached to the arms of SP 2 (Test 1) and SP 3 (Test 2), respectively, without any verbal intervention. This created a testing environment that differed from the training scenarios. [Figure 1](#) displays the training procedure of this study.



豆包AI生成

Figure 1 Flow chart and timeline of training procedures for experimental and control groups. **n** stands for number of participant; **Control** and **Intervention** are two different training groups; **Day 1–3** refers to training day, **Test 1** and **Test 2** refers to two post-training tests on skill retention.

Scoring of Performance

Two instructors graded the procedures based on videotaped. The two instructors scored jointly until they achieved agreement on all points. Their scoring on Test 1 and Test 2 was under the guidance of the venipuncture scoring standard developed by the Medical Simulation Center at Shandong University. The scoring standard for venipuncture performance consists of five sub-tasks with 38 items, totaling 100 points. Subtasks include preoperative preparation, pre-puncture operation, puncture process, post-puncture treatment, and overall professionalism behaviors. The standard was jointly formulated by five clinical experts and has been used in national operation skill competitions. The scoring sheets can be found in the [Supplementary Box 2](#).

To assess inter-rater reliability, intraclass correlation coefficients (ICC) were calculated using a two-way random-effects model for absolute agreement. ICC values were interpreted as: <0.40 = poor, 0.40 – 0.59 = fair, 0.60 – 0.74 = good, and 0.75 – 1.00 = excellent reliability.

Video Recording

Trainees' performance was recorded by a state-of-the-art camera system in the training rooms (2DE22041W-D3, Hikvision, Hangzhou, China). Specifically, three high-quality cameras, strategically positioned on the top, the left and right side of the training room, respectively, about 2.2 meters above the floor. The performance desk was meticulously arranged to ensure all the movements of the trainees be captured clearly with trainee's voice ([Figure 2](#)).

Video Analysis

Video data were analyzed using BORIS (Behavioral Observation Research Interactive Software), a video annotation software (<https://boris.readthedocs.io/en/latest/>). BORIS is commonly used for behavioral research, and it allowed researchers to calibrate meaningful points (events) and time periods in performance videos.

Video analysis reported these measures: total operational score (TS), total task time (TTT), banding tourniquet time (BTT), sterilization of the skin time (SST), puncture and blood collection time (PBT). The definitions of points and events for video annotation were determined by experts and educators of the study and are shown in [Box 1](#).

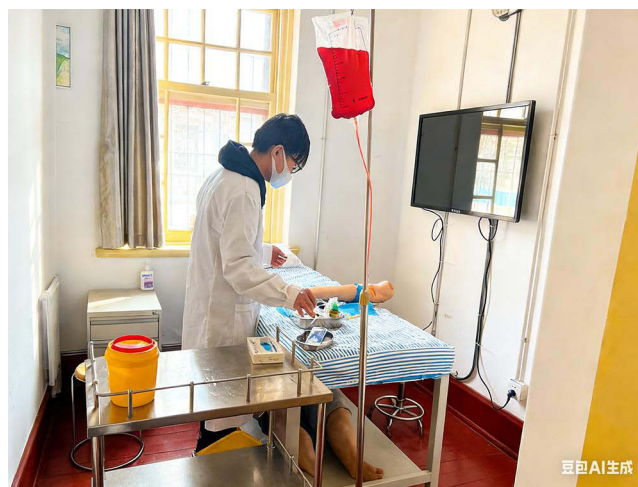


Figure 2 Simulation training setting. Written informed consent has been obtained from all individuals depicted in the image for publication purposes.

We also searched for the behaviors affecting patient safety done by the participants in each trial which include the number of doctor-patient communications (DPC), violations of aseptic technique (VPA), and incorrect sequence of performance (ISP). DPC includes verifying patient information, explaining the purpose of the venipuncture, asking about allergy history, determining patient readiness, and explaining postoperative precautions; VPA includes sterile swabs touching nonsterile areas, crossing over open bottles of antiseptic solution, touching sterilized skin, and blood collection needles touching nonsterile items or areas; ISP includes disinfecting before applying a tourniquet, removing a needle before loosening a tourniquet, and withdrawing a needle before loosening a fist.

Statistics

All data were statistically analyzed using SPSS version 27.0 (IBM Corp., Armonk, NY, USA). Normality of data distribution was assessed using the Shapiro–Wilk test.

For between-group comparisons (experimental vs. control), independent-sample t-tests were used for normally distributed continuous variables: TTT, BTT, SST, PBT and TS. Mann–Whitney *U*-tests were applied to non-normally distributed variables: DPC, VPA, and ISP.

Box 1 Events and Dependent Measures

Events and event definitions
<ol style="list-style-type: none"> 1. Take the tourniquet: the first frame when the hand reaches for the tourniquet and touches the tourniquet 2. Finish tying the tourniquet: first frame that the hand leaves the tourniquet after tying it. 3. Dipping the swab into the vial: the first frame the swab enters the vial. 4. End of disinfection: the first frame when the swab leaves the skin after the second disinfection of the skin is completed. 5. Start of puncture: the first frame when the hand reaches the handle of the blood collection needle. 6. End of blood collection: the first frame when the needle is removed from the blood collection tube.
Dependent measures
<ol style="list-style-type: none"> 1. Bundle the tourniquet time (BTT = event 2 - event 1): from the time the hand touches the tourniquet to the time it is tied and removed from the tourniquet. 2. Sterilization of the skin time (SST = event 4 - event 3): from the time the swab enters the sterilised bottle to the time it leaves the skin after sterilisation is complete. 3. Puncture and blood collection time (PBT = event 6 - event 5): from the time the hand touches the blood collection needle handle to the time the needle is removed from the blood collection tube. 4. Total task time (TTT = BTT + SST + PBT)

For within-group comparisons (Test 1 vs. Test 2), paired t-tests were used for the same normally distributed variables and Wilcoxon signed-rank tests for non-normal variables.

Pearson correlation analysis was performed to examine relationships between selected variables.

To provide a more comprehensive understanding of the findings, effect sizes and 95% confidence intervals were calculated for all primary outcomes. Cohen’s d was reported for t-tests (interpreted as small = 0.2, medium = 0.5, large = 0.8), and rank-biserial correlation was reported for Mann–Whitney U and Wilcoxon signed-rank tests. Means and standard deviations were reported in this paper. A p-value < 0.05 was considered statistically significant.

Results

There were 10 participants in the control group (mean age 20.50 ± 0.85, 50% female) and 8 participants in the experimental group (mean age 20.63 ± 0.74, 50% female). The participants were all second- and third-year medical students. None of them had experience in practicing peripheral venipuncture maneuvers before the study.

Inter-rater reliability was good to excellent across all four assessment sessions. The ICC were 0.778 (Test 1 live scoring), 0.697 (Test 1 video scoring), 0.878 (Test 2 live scoring), and 0.685 (Test 2 video scoring), indicating acceptable to excellent consistency between the two evaluators.

Compared to the control group, the experimental group did not show a significant improvement in their total operative scores on Test 1 and Test 2 after training (Table 1A). However, within-group comparisons revealed that the control group’s live scoring significantly increased from Test 1 to Test 2 (P = 0.000, d = 1.90, 95% CI: -11.21 to -5.09), while no significant changes were observed in the other score comparisons (Table 1B).

Total task time was defined as the time from the beginning of the trainee’s positioning of the puncture site to the end of the puncture site pressure after needle removal. At Test 1, the experimental group showed significantly shorter sterilization time (P = 0.042, d = 1.28, 95% CI: 0.35 to 16.53); the difference in total task time was not significant. At Test 2, the experimental group demonstrated significantly shorter total task time (P = 0.039, d = 1.07, 95% CI: 1.61 to 55.14), sterilization time (P = 0.023, d = 1.20, 95% CI: 1.19 to 13.64), puncture time (P = 0.035, d = 1.09, 95% CI: 0.41

Table 1 Comparison of Total Operational Scores (TS)

(A) Between Two Training Groups				
Group	Test 1		Test 2	
	Live scoring	Video scoring	Live scoring	Video scoring
Control (n = 10)	67.80 ± 10.30	66.40 ± 9.71	75.95 ± 8.62	69.40 ± 11.09
Intervention (n = 8)	65.38 ± 11.69	74.00 ± 8.23	69.19 ± 10.88	70.50 ± 6.19
P	0.689 ^a	0.097 ^b	0.160 ^b	0.755 ^a
Cohen’s d	0.22 ^c	0.84	0.69	0.12 ^c
95% CI	[-16.57, 1.37] ^d	[-16.74, 1.54]	[-2.97, 16.49]	[-10.05, 7.85] ^d
(B) From the two post-training tests				
Test	Control (n = 10)		Intervention (n = 8)	
	Live scoring	Video scoring	Live scoring	Video scoring
Test 1	67.80 ± 10.30	66.40 ± 9.71	67.00 ± 10.25	74.00 ± 8.23
Test 2	75.95 ± 8.62	69.40 ± 11.09	68.62 ± 9.52	70.50 ± 6.19
P	0.000 ^e	0.368 ^e	0.382 ^e	0.326 ^e
Cohen’s d	1.90	0.30	0.33	0.37
95% CI	[-11.21, -5.09]	[-10.15, 4.15]	[-13.48, 5.86]	[-4.33, 11.33]

Notes: ^a Mann–Whitney U-test (Z = -0.400 for Test 1 Live; Z = -0.312 for Test 2 Video); ^b Independent-sample t-test; ^c For non-parametric tests, effect size is reported as rank-biserial correlation (r): Test 1 Live r = 0.09; Test 2 Video r = 0.07; ^d 95% confidence interval for the mean difference (Control minus Intervention); ^e Paired t-test.

to 10.07), and also significantly shorter tourniquet tying time ($P < 0.001$, $r = 0.84$) compared to the control group (Table 2A).

Within-group comparisons revealed that the control group’s total task time increased significantly from Test 1 to Test 2 ($P = 0.032$, $d = 0.80$, 95% CI: -61.05 to -3.35), while the experimental group showed no significant change. No significant differences were found in any sub-task times for either group (Table 2B).

For the three behavioral indicators that affect the quality of patient safety (DPC, VPA, ISP) Mann-Whitney *U*-tests revealed a significant difference between groups for DPC at Test 1 ($P = 0.000$, $r = 0.87$), indicating that the experimental group had fewer communications than the control group. No other between-group differences were significant (Table 3A). Within-group comparisons using Wilcoxon signed-rank tests showed a significant decrease in DPC from Test 1 to Test 2 in the control group ($P = 0.004$, $r = 0.90$); no other within-group changes were significant (Table 3B).

To explore the relationship between some of these key variables of operational performance, a correlation study was conducted between the data from the two post-training assessments.

The correlation for Test 1 showed a moderate correlation between the performance score and violation of aseptic rule ($r = -0.40$, $P = 0.099$). The r values with all other variables were very weak and not significant (Table 4A). The total time only showed a weak correlation with the variables of clarity of verbal communication ($r = 0.21$) and incorrect sequence of performance ($r = -0.39$). None of them reached the a priori significance level. Therefore, it was not necessary to perform further regression analyses on the influences affecting the total score.

The correlation statistics of Test 2 showed the weak correlation between the total performance score and the three performance variables (sterilization time, $r = 0.24$; clarity of verbal communication, $r = 0.35$; and violation of aseptic rule, $r = -0.27$). However, none of these correlations reached the a priori significance level. Therefore, it was not necessary to perform further regression analysis on the influences affecting the total score.

The total task time in Test 2 was weakly correlated with elastic band time ($r = 0.21$, $P = 0.400$) and puncture and blood collection ($r = 0.29$, $P = 0.240$), but strongly correlated with sterilization of the skin time ($r = 0.74$, $P < 0.001$). Correlations for all other variables were weak and not significant (Table 4B). Therefore, it was not necessary to perform further regression analysis on the factors affecting total operating time.

Table 2 Comparison of Operating Time

(A) Between the Two Training Groups and Over Two Post-Training Tests								
Group	Test 1				Test 2			
	TTT	BTT	SST	PBT	TTT	BTT	SST	PBT
Control (n = 10)	107.30 ± 26.06	11.60 ± 2.12	23.52 ± 6.86	23.53 ± 8.44	139.50 ± 32.77	12.00 ± 4.44	23.71 ± 6.90	22.12 ± 6.27
Intervention (n = 8)	92.00 ± 23.35	14.63 ± 4.32	17.37 ± 5.43	15.09 ± 2.77	111.13 ± 15.43	12.25 ± 5.34	16.30 ± 5.14	16.88 ± 1.45
P	0.214 ^a	0.419 ^b	0.055 ^a	0.042 ^a	0.039 ^a	0.000 ^b	0.023 ^a	0.035 ^a
Cohen's d	0.61	0.19	0.99	1.28	1.11	0.84	1.22	1.15
95% CI	[-9.75, 40.35]	[-6.32, 0.26] ^c	[-0.16, 12.47]	[0.35, 16.53]	[1.61, 55.14]	[-5.13, 4.63] ^c	[1.19, 13.64]	[0.41, 10.07]
(B) From the Two Tests								
Test	Control (n = 10)				Intervention (n = 8)			
	TTT	BTT	SST	PBT	TTT	BTT	SST	PBT
Test 1	107.30 ± 26.06	11.60 ± 2.12	23.52 ± 6.86	23.52 ± 10.44	92.00 ± 23.35	14.62 ± 4.10	17.37 ± 5.43	15.08 ± 2.77
Test 2	139.50 ± 32.77	12.00 ± 4.45	23.70 ± 6.70	22.12 ± 6.27	111.13 ± 15.43	12.25 ± 5.34	16.29 ± 5.14	16.88 ± 1.45
P	0.032 ^d	0.800 ^e	0.924 ^d	0.659 ^d	0.084 ^d	0.663 ^e	0.273 ^d	0.209 ^d
Cohen's d	0.80	0.12 ^f	0.03	0.14	0.71	0.62 ^f	0.42	0.49
95% CI	[-61.05, -3.35]	[-1.98, 2.78] ^g	[-4.37, 4.00]	[-5.57, 8.37]	[-41.55, 3.30]	[-5.58, 0.84] ^g	[-1.06, 3.21]	[-4.86, 1.27]

Notes: ^aIndependent-sample *t*-test; ^b Mann-Whitney *U*-test ($Z = -0.808$ for Test 1 BTT; $Z = -3.558$ for Test 2 BTT); ^c Rank-biserial correlation (r): Test 1 BTT $r = 0.19$; Test 2 BTT $r = 0.84$; ^d95% confidence interval for the mean difference (Control minus Intervention). ^e Paired *t*-test; ^f Mann-Whitney *U*-test (for within-group comparisons, Wilcoxon signed-rank test would be more appropriate; these *P* values are retained from original analysis); ^g Effect size for non-parametric comparison is rank-biserial correlation (r): Control BTT $r = 0.12$; Intervention BTT $r = 0.62$; ^h 95% confidence interval for the mean difference (Test 2 minus Test 1).

Abbreviations: TTT, total task time (s); BTT, banding tourniquet time (s); SST, sterilization of the skin time (s); PBT, puncture and blood collection time (s).

Table 3 Comparison of Indicators Affecting the Quality of Operations

(A) In Two Training Groups						
Group	Test 1			Test 2		
	DPC	VPA	ISP	DPC	VPA	ISP
Control (n = 10)	4.30 ± 1.06	0.30 ± 0.48	0.20 ± 0.42	4.20 ± 1.40	0.20 ± 0.42	0.10 ± 0.32
Intervention (n = 8)	4.13 ± 1.25	0.25 ± 0.46	0.50 ± 0.53	3.63 ± 1.60	0.50 ± 0.53	0.50 ± 0.54
P	0.000 ^a	0.819 ^a	0.192 ^a	0.407 ^a	0.192 ^a	0.067 ^a
r	0.87 ^b	0.05 ^b	0.31 ^b	0.20 ^b	0.31 ^b	0.43 ^b

(B) From the two tests						
Test	Control (n = 10)			Intervention (n = 8)		
	DPC	VPA	ISP	DPC	VPA	ISP
Test 1	4.30 ± 1.06	0.30 ± 0.48	0.20 ± 0.42	4.13 ± 1.25	0.20 ± 0.42	0.50 ± 0.54
Test 2	4.20 ± 1.40	0.20 ± 0.42	0.10 ± 0.32	3.63 ± 1.60	0.50 ± 0.53	0.50 ± 0.54
P	0.938 ^a	0.615 ^a	0.542 ^a	0.515 ^a	0.317 ^a	1.000 ^a
r	0.90 ^b	0.32 ^b	0.18 ^b	0.50 ^b	0.50 ^b	0.00 ^b

Notes: ^aP values are from Mann-Whitney U-tests (Table 3A) or Wilcoxon signed-rank tests (Table 3B). ^bRank-biserial correlation (r) = |Z|/√n for Table 3A (n = 18) and = |Z|/√n for Table 3B (n = 10 for control, n = 8 for intervention).

Table 4 Correlation Coefficients of Key Variables Affecting the Total Score

(A) Test 1								
	TS	TTT	BTT	SST	PBT	CPP	VAR	ISP
TS	I							
TTT	r = -0.00 P = 0.989	I						
BTT	r = 0.23 P = 0.370	r = 0.05 P = 0.847	I					
SST	r = -0.23 P = 0.365	r = 0.09 P = 0.724	r = -0.22 P = 0.376	I				
PBT	r = -0.14 P = 0.593	r = -0.12 P = 0.628	r = 0.04 P = 0.878	r = 0.20 P = 0.423	I			
CPP	r = 0.21 P = 0.409	r = 0.21 P = 0.398	r = -0.04 P = 0.880	r = 0.07 P = 0.795	r = 0.23 P = 0.367	I		
VAR	r = 0.43 P = 0.079	r = 0.12 P = 0.641	r = -0.08 P = 0.757	r = 0.18 P = 0.487	r = -0.16 P = 0.523	r = -0.36 P = 0.147	I	
ISP	r = 0.19 P = 0.443	r = -0.39 P = 0.113	r = 0.34 P = 0.163	r = -0.13 P = 0.614	r = 0.19 P = 0.455	r = -0.36 P = 0.139	r = 0.09 P = 0.729	I

(B) Test 2								
	TS	TTT	BTT	SST	PBT	CPP	VAR	ISP
TS	I							
TTT	r = -0.32 P = 0.190	I						
BTT	r = 0.19 P = 0.453	r = 0.21 P = 0.400	I					
SST	r = -0.26 P = 0.300	r = 0.74 P = 0.000	r = -0.05 P = 0.854	I				

(Continued)

Table 4 (Continued).

(B) Test 2								
	TS	TTT	BTT	SST	PBT	CPP	VAR	ISP
PBT	r = 0.29 P = 0.239	r = 0.29 P = 0.240	r = -0.30 P = 0.229	r = 0.39 P = 0.108				
CPP	r = 0.06 P = 0.808	r = -0.08 P = 0.741	r = -0.33 P = 0.182	r = 0.05 P = 0.856	r = 0.302 P = 0.223			
VAR	r = 0.41 P = 0.089	r = -0.01 P = 0.983	r = 0.45 P = 0.064	r = -0.12 P = 0.637	r = -0.29 P = 0.241	r = -0.47 P = 0.051		
ISP	r = -0.23 P = 0.352	r = -0.14 P = 0.580	r = 0.07 P = 0.794	r = -0.10 P = 0.692	r = -0.10 P = 0.705	r = -0.24 P = 0.346	r = 0.35 P = 0.153	

Abbreviations: TS, total score; TTT, total task time (s); BTT: banding tourniquet time (s); SST, sterilization of the skin time (s); PBT, puncture and blood collection time (s); DPC, number of doctor-patient communications; VPA, number of violations of aseptic technique; ISP, number of incorrect sequence of performance.

Discussion

We examined the impact of modifying the training difficulty on skill transfer in learning venipuncture. The results of the video data analysis partially support our research hypothesis. At Test 1, completed one day after training, the experimental group demonstrated significantly shorter puncture time ($P = 0.042$) and a near-significant trend for shorter skin disinfection time ($P = 0.055$) compared to the control group, although the difference in total task time did not reach statistical significance ($P = 0.214$). These findings indicate that trainees in the experimental group acquired the critical skills for doing venipuncture and applied these skills more efficiently in certain steps, even at an early stage.

At Test 2, conducted seven days post-training, the experimental group showed significantly shorter total task time ($P = 0.039$), skin disinfection time ($P = 0.023$), puncture time ($P = 0.035$), and tourniquet tying time ($P < 0.001$) compared to the control group (Table 2A). This broader advantage suggests that the benefits of staged difficulty training become more pronounced after a consolidation period.

Both groups showed slightly prolonged task times at Test 2 compared to Test 1. However, the prolongation was notably smaller in the experimental group. Within-group comparisons revealed that the control group's total task time increased significantly from Test 1 to Test 2 ($P = 0.032$), while the experimental group showed no significant change. Trainees in control group performed well when practising in a quiet and predictable environment for multiple times. However, they were less stable when transferred to a new context, such as when a different SP was included in the post-training tests. These results are consistent with previous studies on skill acquisition.²¹ In contrast, the trainees in the experimental group demonstrated greater resistance to interference and better adaptability of performance environment. They were able to apply the skills stably more consistently in challenging contexts and showed less skill decay over time.

Interestingly, the control group's live scoring significantly improved from Test 1 to Test 2 ($P = 0.000$, $d = 1.90$), while no such improvement was observed in the experimental group (Table 1B). This finding may reflect a practice effect or increased familiarity with the assessment environment in the control group, but the absence of corresponding changes in video scoring or behavioral indicators suggests that this improvement was context-specific and did not generalize to more objective measures.

From video recordings, we were able to extract unsatisfied points that affect patient safety and quality, including DPC, VPA and ISP. Contrary to our initial expectation, we found significant differences in DPC: at Test 1, the experimental group had fewer communications than the control group ($P = 0.000$, $r = 0.87$), and the control group's communications decreased significantly by Test 2 ($P = 0.004$, $r = 0.90$) (Table 3A and B). These findings suggest that both training approaches influenced communication patterns, possibly reflecting increased efficiency or different communication strategies. No significant differences were observed in VPA or ISP, indicating that these safety-critical behaviors may be more resistant to change through difficulty manipulation alone and might require explicit feedback or instruction, as shown in previous studies.^{22,23}

It should be noted that no significant differences were observed between the two groups in overall performance scores at either Test 1 or Test 2 (Table 1A), nor in within-group comparisons across the two tests (Table 1B). This suggests that the primary benefit of staged difficulty training may lie in enhancing procedural efficiency and adaptability to contextual challenges, rather than improving global competency scores.

This study has limitations. First, the sample size was relatively small ($N = 18$), and no a priori power analysis was conducted. This limits the statistical power to detect smaller effects and may affect the generalizability of the findings. Second, the follow-up assessments were limited to 1 day and 7 days post-training, capturing only short-term skill transfer. Third, the simulation model for venipuncture training was a standard device, which may not fully replicate clinical variability. Although inter-rater reliability was good, the rating scale itself remains inherently subjective.

Future studies should include extended follow-up periods (eg., 1–3 months) to evaluate long-term skill retention and investigate whether adjusting training difficulty yields similar results in more complex procedures.

Conclusion

In summary, adjusting training difficulty in the middle and late stages of training can enhance the skill transfer in performing the venipuncture blood collection procedure, with improvements primarily reflected in reduced task times rather than overall performance scores. Results also supported training with increasing challenge allow participants to build skill more robust with resistance to skill decay. Overall results inspire us to adjust the complexity of training scenarios or add interferences appropriately of health trainees in order to improve the training outcomes.

Abbreviations

SP, Standardized Patient; ICC, Intraclass Correlation Coefficients; BORIS, Behavioral Observation Research Interactive Software; TS, Total Score; TTT, Total Task Time; BTT, Banding Tourniquet Time; SST, Sterilization of the Skin Time; PBT, Puncture and Blood Collection Time; DPC, Doctor-Patient Communications; VPA, Violations of Aseptic Technique; ISP, Incorrect Sequence Of Performance.

Data Sharing Statement

The dataset supporting the conclusions of this article is not available because the participants of this study did not give written consent for sharing their data to the public. The data set is not available in concern of confidentiality of participants.

Ethics Approval and Consent to Participate

Our study adhered to the principles outlined in the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of the School of Clinical Medicine, Shandong University (SDULCLL2022-20), and all participants provided informed consent prior to their involvement in the study.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This research was supported by grant M2023357 from the Undergraduate Teaching Reform Research Project in Shandong Province, China.

Disclosure

The authors declare that they have no competing interests in this work.

References

- Vogel D, Harendza S. Basic practical skills teaching and learning in undergraduate medical education - a review on methodological evidence. *GMS J Med Educ.* 2016;33(4):Doc64. doi:10.3205/zma001063
- Elnicki DM, van Londen J, Hemmer PA, Fagan M, Wong R. U.S. and Canadian internal medicine clerkship directors' opinions about teaching procedural and interpretive skills to medical students. *Acad Med.* 2004;79(11):1108–1113. doi:10.1097/00001888-200411000-00022
- Bugaj TJ, Nikendei C. Practical clinical training in skills labs: theory and practice. *GMS J Med Educ.* 2016;33(4):Doc63. doi:10.3205/zma001062
- Bosse HM, Mohr J, Buss B, et al. The benefit of repetitive skills training and frequency of expert feedback in the early acquisition of procedural skills. *BMC Med Educ.* 2015;15:22. doi:10.1186/s12909-015-0286-5
- Herrmann-Werner A, Nikendei C, Keifenheim K, et al. “Best practice” skills lab training vs. a “see one, do one” approach in undergraduate medical education: an RCT on students' long-term ability to perform procedural clinical skills. *PLoS One.* 2013;8(9):e76354. doi:10.1371/journal.pone.0076354
- Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. *Simul Healthc.* 2006;1(4):252–256. doi:10.1097/01.SIH.0000242724.08501.63
- Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med.* 2004;79(10 Suppl):S70–81. doi:10.1097/00001888-200410001-00022
- Lopreiato JO, Sawyer T. Simulation-based medical education in pediatrics. *Acad Pediatr.* 2015;15(2):134–142. doi:10.1016/j.acap.2014.10.010
- Sakawi Y, Vetter TR. Airway management and vascular access simulation during a medical student rotation. *Clin Teach.* 2011;8(1):48–51. doi:10.1111/j.1743-498X.2010.00406.x
- Liu Y, Li J, Chang J, Xiao S, Pei W, Wang L. A new inexpensive ultrasound-guided central venous catheterization simulation model. *BMC Med Educ.* 2023;23(1):106. doi:10.1186/s12909-023-04080-z
- Ali KQ, Soofi SB, Hussain AS, et al. Simulator-based ultrasound training for identification of endotracheal tube placement in a neonatal intensive care unit using point of care ultrasound. *BMC Med Educ.* 2020;20(1):409. doi:10.1186/s12909-020-02338-4
- Promes SB, Chudgar SM, Grochowski CO, et al. Gaps in procedural experience and competency in medical school graduates. *Acad Emerg Med.* 2009;16 Suppl 2:S58-62. doi:10.1111/j.1553-2712.2009.00600.x
- Bruce AN, Kumar A, Malekzadeh S. Procedural skills of the entrustable professional activities: are graduating us medical students prepared to perform procedures in residency? *J Surg Educ.* 2017;74(4):589–595. doi:10.1016/j.jsurg.2017.01.002
- Burdet E, Osu R, Franklin DW, Milner TE, Kawato M. The central nervous system stabilizes unstable dynamics by learning optimal impedance. *Nature.* 2001;414(6862):446–449. doi:10.1038/35106566
- Christiansen L, Larsen MN, Madsen MJ, Grey MJ, Nielsen JB, Lundbye-Jensen J. Long-term motor skill training with individually adjusted progressive difficulty enhances learning and promotes corticospinal plasticity. *Sci Rep.* 2020;10(1):15588. doi:10.1038/s41598-020-72139-8
- Kerr R, Booth B. Specific and varied practice of motor skill. *Percept Mot Skills.* 1978;46(2):395–401. doi:10.1177/003151257804600201
- Poolton JM, Masters RS, Maxwell JP. The relationship between initial errorless learning conditions and subsequent performance. *Hum Mov Sci.* 2005;24(3):362–378. doi:10.1016/j.humov.2005.06.006
- Capio CM, Poolton JM, Sit CH, Holmstrom M, Masters RS. Reducing errors benefits the field-based learning of a fundamental movement skill in children. *Scand J Med Sci Sports.* 2013;23(2):181–188. doi:10.1111/j.1600-0838.2011.01368.x
- Bugdadi A, Sawaya R, Olwi D, et al. Automaticity of force application during simulated brain tumor resection: testing the Fitts and Posner model. *J Surg Educ.* 2018;75(1):104–115. doi:10.1016/j.jsurg.2017.06.018
- Sadideen H, Plonczak A, Saadeddin M, Kneebone R. How educational theory can inform the training and practice of plastic surgeons. *Plast Reconstr Surg Glob Open.* 2018;6(12):e2042. doi:10.1097/GOX.0000000000002042
- Christiansen L, Madsen MJ, Bojsen-Moller E, Thomas R, Nielsen JB, Lundbye-Jensen J. Progressive practice promotes motor learning and repeated transient increases in corticospinal excitability across multiple days. *Brain Stimul.* 2018;11(2):346–357. doi:10.1016/j.brs.2017.11.005
- Briggs A, Raja AS, Joyce MF, et al. The role of nontechnical skills in simulated trauma resuscitation. *J Surg Educ.* 2015;72(4):732–739. doi:10.1016/j.jsurg.2015.01.020
- Ashour A, Phipps DL, Ashcroft DM. The role of non-technical skills in community pharmacy practice: an exploratory review of the literature. *Int J Pharm Pract.* 2021;29(3):203–209. doi:10.1093/ijpp/riaa014

Advances in Medical Education and Practice

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>

Dovepress
Taylor & Francis Group