



Improving Pediatric/Neonatology Residents' Newborn Resuscitation Skills With a Digital Serious Game: DIANA

Serena Bardelli^{1†}, Giulio Del Corso^{2†}, Massimiliano Ciantelli^{1,3}, Marta Del Pistoia^{1,3}, Francesca Lorenzoni^{1,3}, Nicoletta Fossati⁴, Rosa T. Scaramuzzo^{1,3*} and Armando Cuttano^{1,3}

¹ Centro di Formazione e Simulazione Neonatale "NINA," U.O. Neonatologia, Dipartimento Materno-Infantile, AOUP, Pisa, Italy, ² Department of Mathematics, Gran Sasso Science Institute (GSSI), L'Aquila, Italy, ³ U.O. Neonatologia, Dipartimento Materno-Infantile, AOUP, Pisa, Italy, ⁴ Institute of Medical and Biomedical Education, Faculty of Medicine, St. George's University of London, London, United Kingdom

OPEN ACCESS

Edited by:

Philipp Deindl,
University Medical Center
Hamburg-Eppendorf, Germany

Reviewed by:

Shashi Kant Dhir,
Guru Gobind Singh Medical College
and Hospital, India
Brenda Law,
University of Alberta, Canada

*Correspondence:

Rosa T. Scaramuzzo
rosa.scaramuzzo@gmail.com

†These authors share first authorship

Specialty section:

This article was submitted to
Neonatology,
a section of the journal
Frontiers in Pediatrics

Received: 23 December 2022

Accepted: 21 February 2022

Published: 01 April 2022

Citation:

Bardelli S, Del Corso G, Ciantelli M, Del Pistoia M, Lorenzoni F, Fossati N, Scaramuzzo RT and Cuttano A (2022) Improving Pediatric/Neonatology Residents' Newborn Resuscitation Skills With a Digital Serious Game: DIANA. *Front. Pediatr.* 10:842302. doi: 10.3389/fped.2022.842302

Background: Serious games, and especially digital game based learning (DGBL) methodologies, have the potential to strengthen classic learning methodology in all medical procedures characterized by a flowchart (e.g., neonatal resuscitation algorithm). However, few studies have compared short- and long-term knowledge retention in DGBL methodologies with a control group undergoing specialist training led by experienced operators. In particular, resident doctors' learning still has limited representation in simulation-based education literature.

Objective: A serious computer game DIANA (**D**igital **A**pplication in **N**ewborn **A**ssessment) was developed, according to newborn resuscitation algorithm, to train pediatric/neonatology residents in neonatal resuscitation algorithm knowledge and implementation (from procedure knowledge to ventilation/chest compressions rate). We analyzed user learning curves after each session and compared knowledge retention against a classic theoretical teaching session.

Methods: Pediatric/neonatology residents of the Azienda Ospedaliera Universitaria Pisana (AOUP) were invited to take part in the study and were split into a game group or a control group; both groups were homogeneous in terms of previous training and baseline scores. The control group attended a classic 80 min teaching session with a neonatal trainer, while game group participants played four 20 min sessions over four different days. Three written tests (pre/immediately post-training and at 28 days) were used to evaluate and compare the two groups' performances.

Results: Forty-eight pediatric/neonatology residents participated in the study. While classic training by a neonatal trainer demonstrated an excellent effectiveness in short/long-term knowledge retention, DGBL methodology proved to be equivalent or better. Furthermore, after each game session, DGBL score improved for both procedure knowledge and ventilation/chest compressions rate.

Conclusions: In this study, DGBL was as effective as classic specialist training for neonatal resuscitation in terms of both algorithm memorization and knowledge retention. User appreciation for the methodology and ease of administration, including remotely, support the use of DGBL methodologies for pediatric/neonatology residents education.

Keywords: DGBL, digital games, technology-enhanced training or learning, neonatal resuscitation, memory and retention, newborn infants, healthcare education, serious game

INTRODUCTION

Globally, an estimated 2.5 million newborns die each year worldwide from childbirth asphyxia (defined as a failure to initiate or sustain spontaneous breathing at birth) (1) as ~15% of full term births require effective resuscitation (2). Correctly performed neonatal resuscitation can save around 700,000 lives worldwide every year (SIN [Società Italiana di Neonatologia, Italian Neonatology Society] Survey on the organization of care in the delivery room, 2020). However, resuscitation guidelines are not adhered in more than 90% of cases (3).

Digital game based learning (DGBL) methodologies have proved effective in multiple medical contexts (4–6) by integrating the advantages of the classic teaching process with the possibilities offered by the use of simulations (replicability, standardized teaching environment, user adaptability of the procedure). They can be applied to most flowchart-based medical procedures and, crucially, their high repeatability and the possibility of dividing each session into several parts can stimulate procedural memory (7, 8). Further advantages of DGBL methodologies include the provision of an optimal context for user result analysis (every action performed by the learner is stored) and a higher attention/appreciation rate by users.

While it is questioned whether DGBL approach can fully replace classic teaching methodologies (9–12), DGBL methods are known to be effective in checking what was learned and reinforcing motivation to enhance adult learning in medical education (13) and, more in general, in higher education (14). With particular regard to medical practice (14–16), and especially neonatal resuscitation (9, 17, 18), numerous existing studies demonstrate the effectiveness of DGBL/simulation methods in stimulating better learning. However, many of these studies lack a scoring baseline (pre-test), a subsequent follow up to evaluate knowledge retention, and/or a homogeneous and independent control group.

DGBL methodologies can be applied to most flowchart-based medical procedures. In this study, we implemented a new ad hoc digital serious game **DIANA** (**D**igital **A**pplication in **N**ewborn **A**ssessment) and we developed it for neonatal resuscitation teaching. Rather than focusing on a single skill (e.g., endotracheal intubation) this computer game aims to teach the entire neonatal resuscitation algorithm. Unlike most published studies, which involved medical students (9, 19, 20) and expert neonatal professionals (17, 21) as learners, we tested it on a group of resident students of varying experience, using a randomized control study design with the primary goal of testing short- and medium-/long-term knowledge retention [primary

endpoint: compare knowledge retention of DGBL and classical training]. The analysis is done by comparing the DGBL group with an independent group undergoing classic training (e.g., 80 min theoretical teaching session provided by an expert neonatal trainer). Indeed, despite an autonomous training using didactic material (9), the choice of a guided approach provides a more controlled training path (10). In addition, several other secondary endpoints were tested to evaluate the performance obtain from DGBL recording scores: knowledge scores, time decision, ventilation/chest compression rate, and user acceptance of this new training methodology.

MATERIALS AND METHODS

Software Description

The DIANA software was developed according to newborn resuscitation flowchart to verify DGBL methodology for training. The DIANA software code was implemented with the real-time development platform Unity (<https://unity.com/>). The video game was divided into four sessions (i.e., distributed study) with an inter-study interval (ISI) of 48 h to consolidate information memory through repetition (4). Each game session consisted of a theoretical and an interactive part. The interactive part started with 1 min of equipment check. The interactive part simulated a clinical case, where the user would choose how to proceed from one of four options provided. A virtual assistant would intervene in case of errors, and provide detailed instructions to enhance learning without diminishing the gaming experience (22). In the theoretical part, the same virtual assistant, with a human voice, would give a theoretical tutorial using videos to demonstrate technical skills. The first session included an interactive game and complete theoretical teaching about the whole neonatal resuscitation procedure. In the second session, the theoretical part addressed equipment check, neonatal care, and assisted ventilation. The interactive part of the video game followed on from the first session, with successful resuscitation after correctly assisted ventilation. In the third session, the theoretical part dealt with endotracheal intubation skills, chest compressions, and drug delivery, with the interactive part of the video game ending after the execution of chest compressions. Lastly, the fourth session consisted of three activities: a tutorial on venous umbilical catheter insertion, a mini game related to the procedure, and the full execution of resuscitation simulation as in the first session (**Figure 1**). To the aim of the present study, residents did not have free access to the software except for sessions scheduled on the basis of the time intervals described in the study.

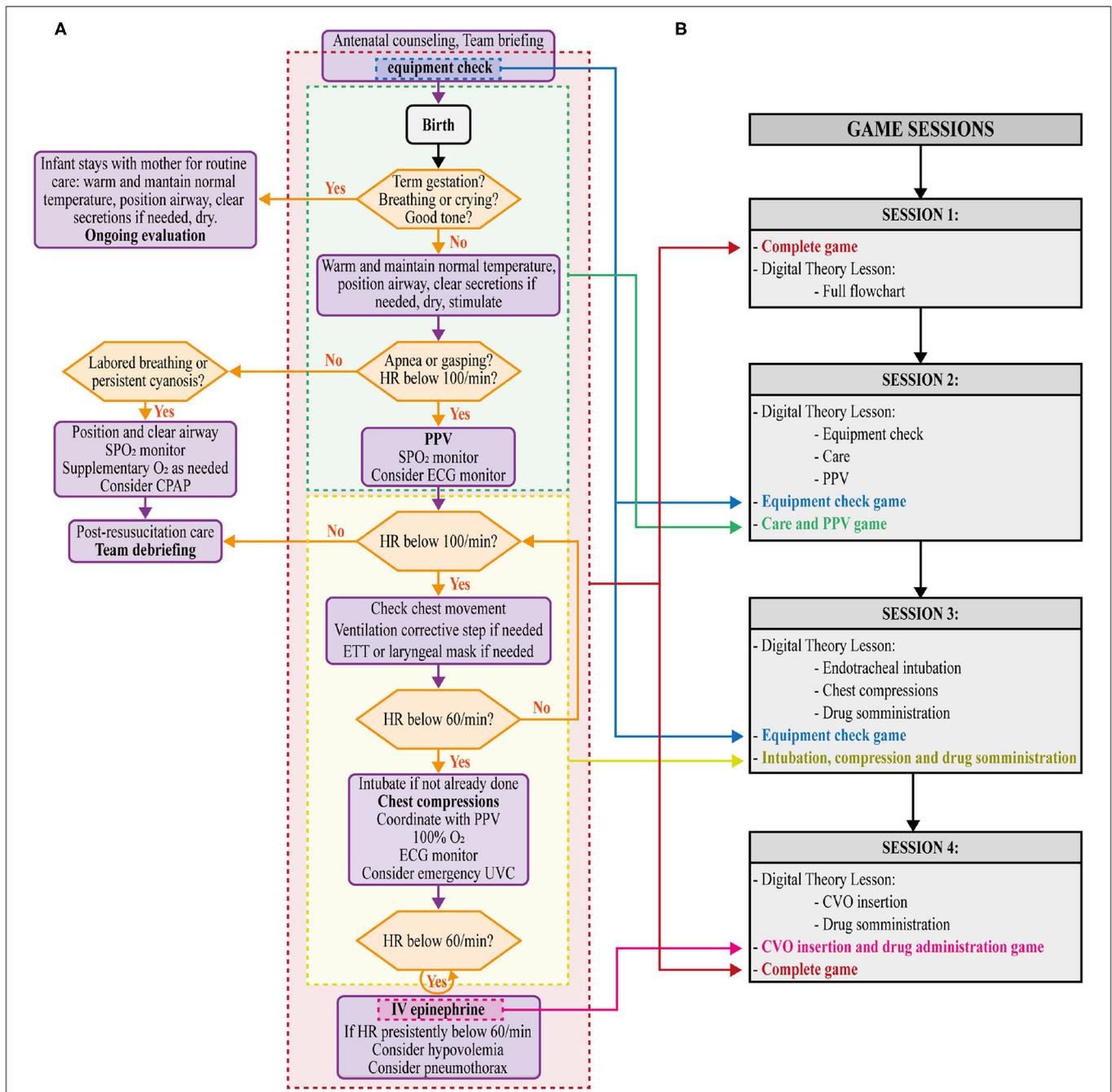


FIGURE 1 | (A) Newborn resuscitation flow chart showing corresponding sections in the Digital Application in Newborn Assessment (DIANA) game (equipment check, neonatal care and PPV, intubation, chest compression and drug administration, umbilical vein catheterization [UVC, CVO in Italian and consequently in this game version] and drug administration, and complete). **(B)** Details of the game sessions (1,2,3,4).

In this work, we scheduled the DIANA sessions to ensure the same time practice between residents. However, for future practical uses of DIANA to support classical training, this fixed schedule is not imposed by the software. Indeed, DIANA does not impose on the user the sequential use of the game levels (e.g., a practitioner can freely select one of the four sessions). This allows the end user to freely practice on a single flowchart

topic or to assess their knowledge of the entire algorithm. The only limitation is that the user within the session will be guided to follow the theoretical part first and then the practical part.

Within the interactive video game, the user had 1 min to select the essential tools (Figure 2A), categorized as totally correct, partially correct, and incorrect. Depending on the tool, size and setting selection would be required. After 1 min, the chosen

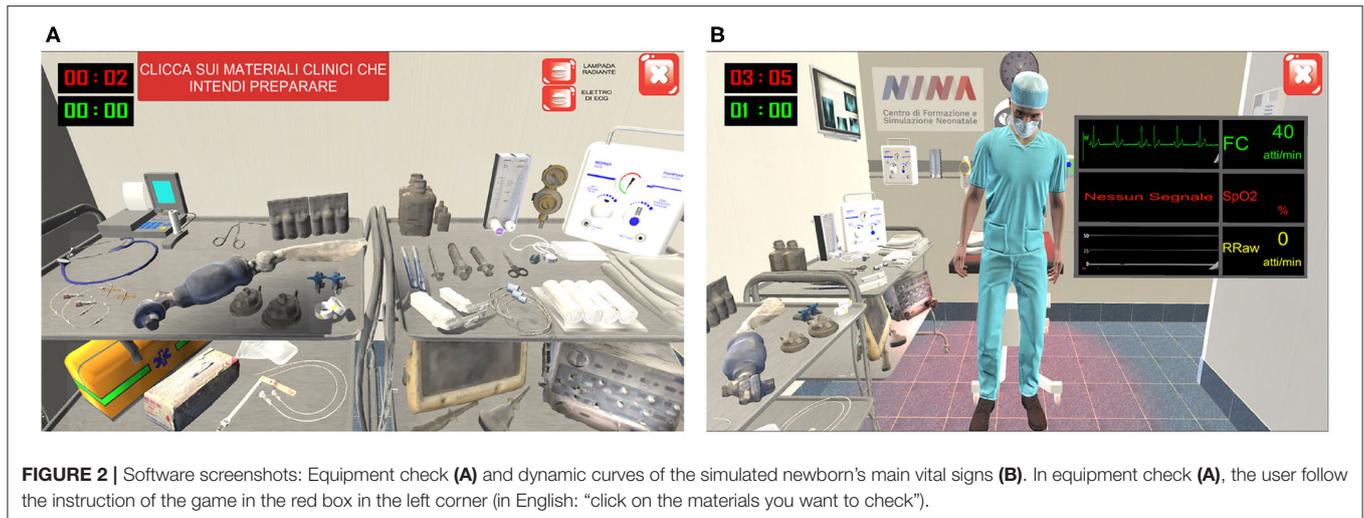


FIGURE 2 | Software screenshots: Equipment check (A) and dynamic curves of the simulated newborn's main vital signs (B). In equipment check (A), the user follows the instruction of the game in the red box in the left corner (in English: "click on the materials you want to check").

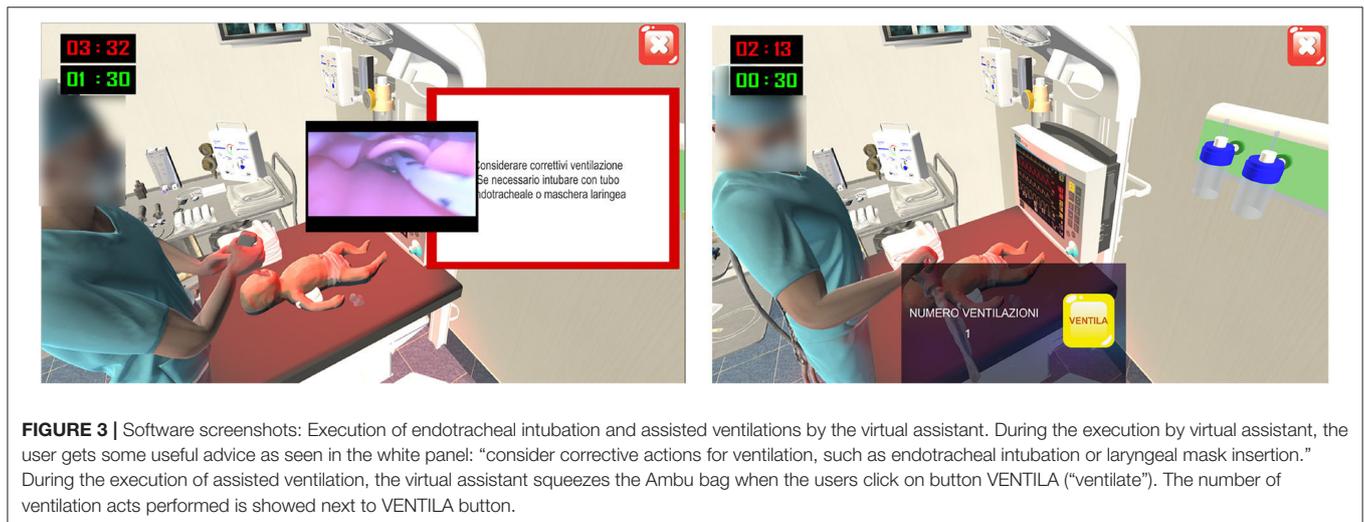


FIGURE 3 | Software screenshots: Execution of endotracheal intubation and assisted ventilations by the virtual assistant. During the execution by virtual assistant, the user gets some useful advice as seen in the white panel: "consider corrective actions for ventilation, such as endotracheal intubation or laryngeal mask insertion." During the execution of assisted ventilation, the virtual assistant squeezes the Ambu bag when the users click on button VENTILA ("ventilate"). The number of ventilation acts performed is shown next to VENTILA button.

tools would appear in a box, checked in green ("selection made") or red ("missing" equipment). When assessing the clinical state of the patient, a monitor would show dynamic curves and heart rate, respiratory rate, and oxygen saturation (Figure 2B). Practical procedures were performed by the virtual assistant (Figure 3). During ventilation execution, the user defined the timing of the ventilation by selecting a "Ventilation" button. The game was designed to last 30 s, during which, every 10 s, the assistant's voice would reassuringly provide feedback to the user, e.g., advising them to increase or reduce the rhythm or complimenting him/her for maintaining an optimal respiratory rate in assisted ventilation. Importantly, chest compressions execution would imply cooperation between user and virtual assistant: the former would perform the required three chest compressions, following one assisted breath by the latter.

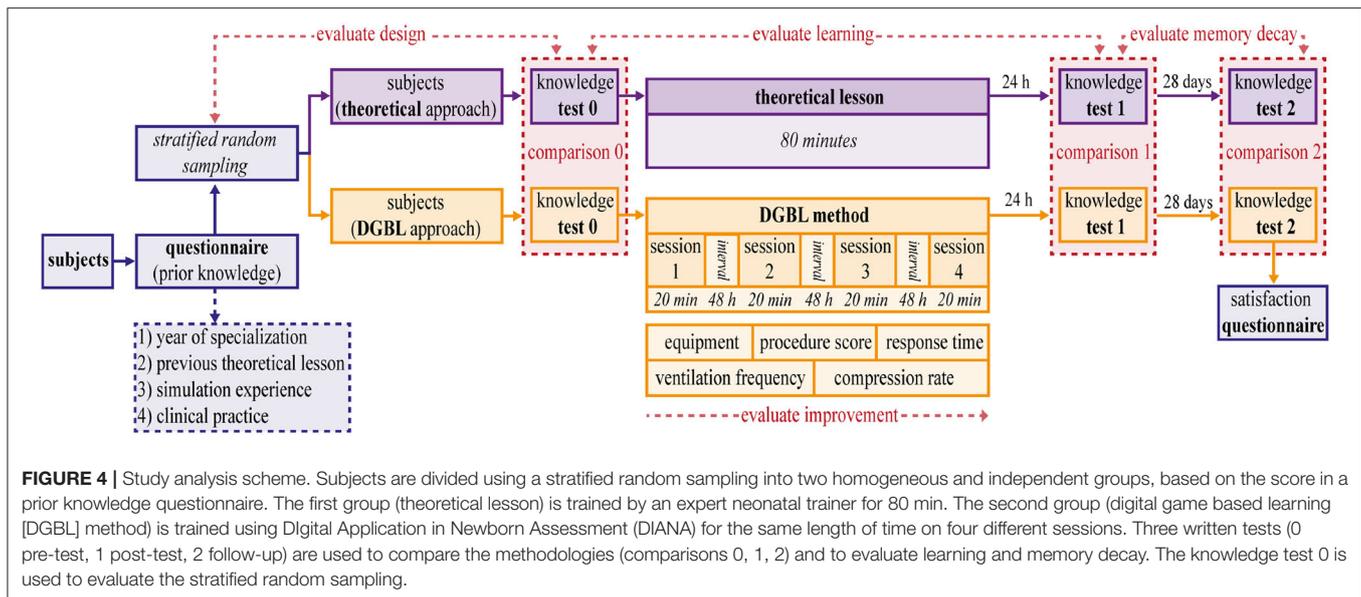
Study Design and Procedure

Study participants filled a questionnaire to assess their previous knowledge and experience (Figure 4). Based on questionnaire

results, two homogeneous groups [Stratified random sampling, similar to other DGBL studies, (12, 23)] were randomized to either the classic teaching process (frontal teaching session) or the one based on digital simulations (DGBL), respectively.

The theoretical teaching session (Figure 4, in purple) was given in person by an expert neonatal trainer, with no more than 10 medical residents for each group, which allowed them to take a very interactive lesson. After finishing the theoretical part, residents practiced the technical skills of PPV, chest compression, and endotracheal intubation on a medium fidelity mannequin (Newborn ANNE, <https://laerdal.com/it/doc/222/Newborn-Anne>). Neither in the theory lesson nor in the software a specific (limited) clinical case was presented and discussed. On the contrary, in both training residents were asked to perform the whole resuscitation algorithm.

The DGBL group training methodology is based on the use of DIANA software. The software guided the user through the entire resuscitation flowchart divided into four phases. Indeed, starting



from the promising results obtained even with a single session of a serious game approach (9, 17, 21), DGBL group (**Figure 4**, in orange) training was based on the natural subdivision allowed by a digital game: four sessions of 20 min each, separated by a 48 h break; knowledge tests began 24 h after the last session, with the same evaluation process as for the classic training group.

Both the groups (**Figure 4**, in purple) underwent three knowledge tests about neonatal resuscitation algorithm and equipment check. The test was administered at three different times: immediately before the tutorial (pre-test 0), at 24 h (post-test 1) and at 28 days (follow-up test 2) after training ending; the questions and answers remained the same, while their order was randomly altered. Specific time intervals between assessments were chosen to capture actual knowledge retention. A 24 h post-training time interval was specifically chosen to filter out the positive effects of short-term memory on scores (24). The 28-day interval to evaluate of memory decay has been widely used in DGBL (25). Unlike a much longer interval adopted by other authors (9, 17), it minimizes the high risk of study drop out within a medical resident population, or the confounding effect of further training. Similarly, candidates were not made aware of our study's assessment methods and timings, including the 28-day delayed test, in order to prevent skewed outcomes. The three scores for either learning method were compared to evaluate the two methodologies, their strengths and limitations (comparisons 0, 1, 2 in red in **Figure 4**). Knowledge test 0 was also used to evaluate the design.

Furthermore, in DGBL group, user improvement was evaluated as the sessions progressed by recording any change in individual tests' numerical values (equipment score, procedure score, response time, ventilation frequency, compression rate) as common in DGBL methodologies (26). At the end of data collection, a user satisfaction questionnaire was administered to DGBL group, to integrate subsequent versions of the software with user suggestions.

Measures

The primary endpoint of this study is to compare the effectiveness between DGBL (DIANA) and classic learning methodology on knowledge retention based on knowledge questionnaire performance. The several secondary endpoints regarding the evaluation of the effectiveness of the DGBL methodologies on the user's performance during the gaming sessions and the satisfactions evaluation of this new methodology are summarized in **Table 1** and described in sections "Knowledge test scores" and "DGBL scores."

Knowledge Test Scores

Knowledge tests are used in DGBL analysis to evaluate performance (20). The test used in this work was written by neonatal resuscitation trainers accredited by SIN, and consisted of 21 questions (each with 1 correct and 5 incorrect answers) related to the correct resuscitation procedure and a list of 40 items (21 correct, 6 partially correct, and 13 incorrect) to check. The knowledge test score was calculated by allocating 1 point for each correct answer, 0 for null, and -0.2 for incorrect ones, so that the average score could be assumed to be zero in case of randomly selected answers. The result was then normalized by the number of questions. The equipment score, on the other hand, consisted of the number of correct instruments (21) selected from the list of 40 items.

DGBL Scores

During the execution of DIANA game, the following parameters were recorded: decision-making/response time, answer correctness from the multiple options included in the simulation, choice of equipment before each simulation, uniformity, and correctness of ventilations/compressions timing. A positive score was assigned for a correct answer, a negative value for an incorrect selection, and a neutral (null) score for selecting the "Get help" option, available for every

TABLE 1 | Description of the variables observed during the study divided between primary endpoints (evaluate the effectiveness of DGBL and classic learning methodology on knowledge retention) and secondary endpoints (evaluate the effectiveness on user's performance during the gaming sessions).

Comparison	Feature observed	Comparison Tool	Question to answer
DGBL (DIANA) and classic learning methodology [primary endpoint]	Knowledge retention and equipment checklist	Knowledge tests (pre-training, 1 day post, and 28 days post-training)	Did the DGBL training methodology prove as effective as theoretical teaching session in knowledge retention?
DGBL(DIANA) games performance [secondary endpoints]	Knowledge retention	Performance of different session game scores	Was the DGBL training methodology effective to learn a flowchart reducing decision time and increasing scores results?
	Equipment checklist		Was the DGBL training methodology effective to learn the equipment checklist?
	Ventilation rate		Was the DIANA ventilation game effective to learn the correct ventilation rate to perform during a PPV procedure?
	Chest compression rate		Was the DIANA chest compression game effective to learn the correct rate to perform during a newborn resuscitation?
	Satisfaction of new methodology	Satisfaction questionnaire	Has the DGBL methodology been considered useful and effective by users?

question to cover the operator's inability to make a decision. Choosing this option was followed by a detailed explanation of the correct decision by the virtual assistant to stimulate learning and improve subsequent sessions' performance. Knowledge score was calculated as the number of correct answers normalized by the number of questions for each session. The equipment score consisted of the number of correct instruments selected from a list of 40 items (21 correct, 6 partially correct, and 13 incorrect). As some game sessions covered only part of the resuscitation procedure (**Figure 1**), the knowledge score was calculated on three question subsets: on care and assisted ventilation (PPV) (sessions 1-2-3-4), on intubation and compressions (sessions 1-3-4), and those on drug administration (session 1-4), respectively. For each answer, the response time (i.e., the time elapsed between the question administration and the execution of the action) was also calculated.

Compression and Ventilation Scores

In the games involving compressions and ventilations, choosing a score that rewarded maintenance of a correct frequency and penalized frequency fluctuations was essential. The number

of acts per minute is not necessarily a reliable parameter to tell an excellent performance (i.e., correct and uniform rate) from a sub-optimal one, such as correct but non-uniform rate with marked variations in frequency. With reference to **Figure 5**, we defined the sequence of acts $1, \dots, n$ and the corresponding $\Delta_i := t_i - t_{i-1}$ as the difference between the time of act i and the time of the previous act $i - 1$. The correct timing intervals are then defined $[min_{freq}, max_{freq}]$ (40–60 ventilations per minute and 80–100 [+30] compressions per minute, where +30 represents the ventilations performed alternately by the virtual assistant). These ranges represent the reference values that the user must maintain and correspond to an interval $[min_{timing}, max_{timing}] = [1/max_{freq}, 1/min_{freq}]$ between the minimum and maximum of the time interval allowed to perform a correct number of acts per minute. Therefore, the correctness value of the i th act is defined as follows:

$$d_i := \begin{cases} 0 & \text{if } \Delta_i \in [min_{timing}, max_{timing}] \\ \max(|\Delta_i - min_{timing}|, |\Delta_i - max_{timing}|) & \text{otherwise} \end{cases}$$

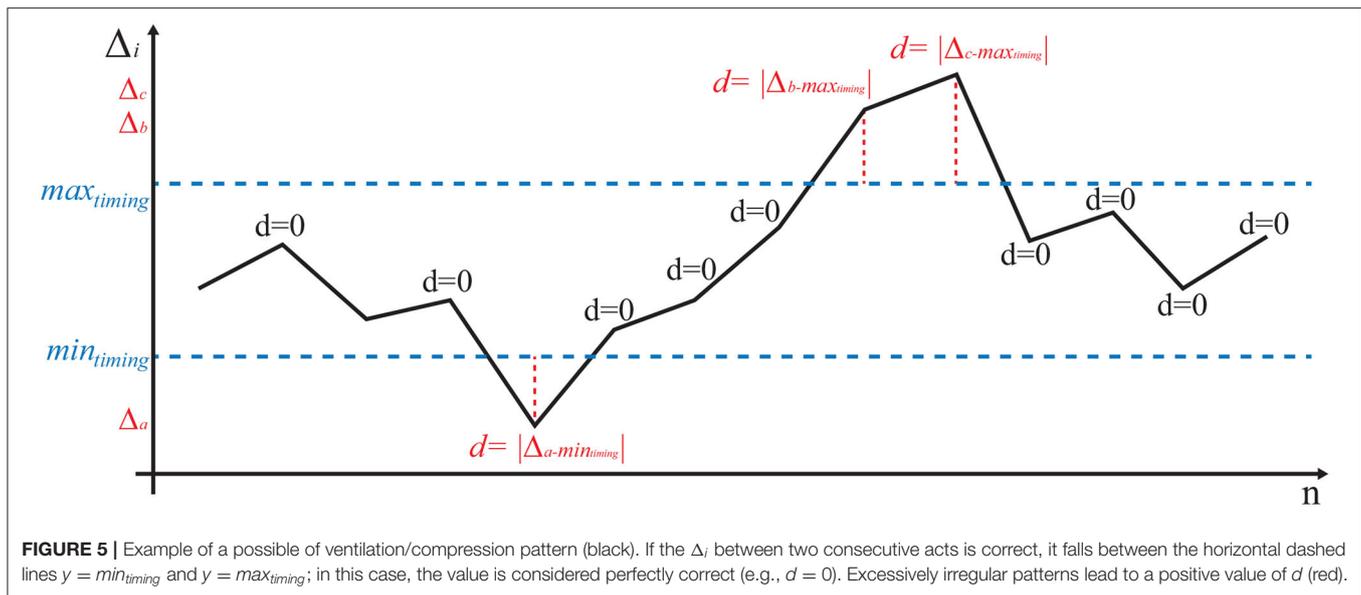
With reference to **Figure 5**, every act falling within the correct ranges is rated as zero, while any variation outside the range (in red in the figure) increases the score in proportion to how much it deviates from the reference values. The first score is defined as the average of the $\{d_i\}_{i=1}^n$ [e.g., $score_{mean} = mean_{i=1}^n(d_i)$]. A null score represents a candidate who has always maintained an optimal frequency of acts while a higher score identifies any deviation from the correct execution. The second score is based on the standard deviation of the $\{d_i\}_{i=1}^n$ [e.g., $score_{std} = std_{i=1}^n(d_i)$]. This score characterizes the irregularity of the values and is, therefore, indicative of maintaining a non-homogeneous timing during the test.

Ethical Approval

Users were pediatric/neonatology residents of the Azienda Ospedaliera Universitaria Pisana (AOUP) who consented to the acquisition, processing, and dissemination of data in anonymized form. The study was approved by the local Institutional Review Board for Ethic Issues. All analyzed data were anonymized and the entire analysis was blinded.

Statistical Analysis

The study design is based on a stratified random sampling to control the nuisance factors. The strata are designed on the basis of a score extrapolated from a questionnaire of previous theoretical/clinical/practical experience. This score was used to create four levels of competence (0 no experience, 1: one of the three experiences, 2: two experiences, up to 3 for those who participated in all simulation, theory, and practice experiences), then used in the study design to divide the candidates of the two groups. The uniformity of the knowledge test 0 score distributions of the two groups' clinical experience was tested using a Kolmogorov–Smirnov (KS) two-sided test. A further indicator of uniformity is the amount of times a random sampling could have produced a better subdivision



than the chosen design. This estimate was achieved by using a Monte Carlo method for probability estimation: 100,000 times the group of all candidates (associated with their respective knowledge test score 0) is randomly divided into two groups (27 and 21, respectively). This (artificial) subdivision represents a possible result of a random fully experimental design. Then, the Kolmogorov–Smirnov distance D between the two sets is calculated and compared with that obtained in the stratified random sampling. The knowledge test scores calculated before learning, at the end of learning and 28 days later, were evaluated by comparing the means, variances and distributions (KS test). The normality of the scores obtained was tested by Shapiro–Wilk test. Variances were compared by F -test for (independent) groups comparison and by Pitman–Morgan test of variance for paired sample for internal group comparisons. Under the assumptions of normality and homogeneity of variances, the independent t -test was used to compare means. In the absence of these hypotheses, the non-parametric (conservative) Wilcoxon signed-rank test and the Mann–Whitney U -test were used. Considering that the scores calculated in the knowledge tests 0, 1, and 2 are repeated measures of the same group and the frequent absence of the hypothesis of normality, the values are preliminary compared using a Friedman test. Post hoc pairwise analysis through the previously described paired tests are then applied to detect variations of the score. Bonferroni correction is presented to counter the problem of multiple post hoc analysis. The comparison between independent groups (i.e., DGBL vs. theory) pre-training, at 1 day and at 28 days is instead carried out with non-paired tests. To analyze the performance of the individual game sessions, the same tests were applied to learning score procedure, the response times of the questions, and the uniformity of the ventilation/compression timing. One-sided versions of the tests were applied to test the monotony of the scores. Statistical analysis was carried out using the software R [4.1.1] (27).

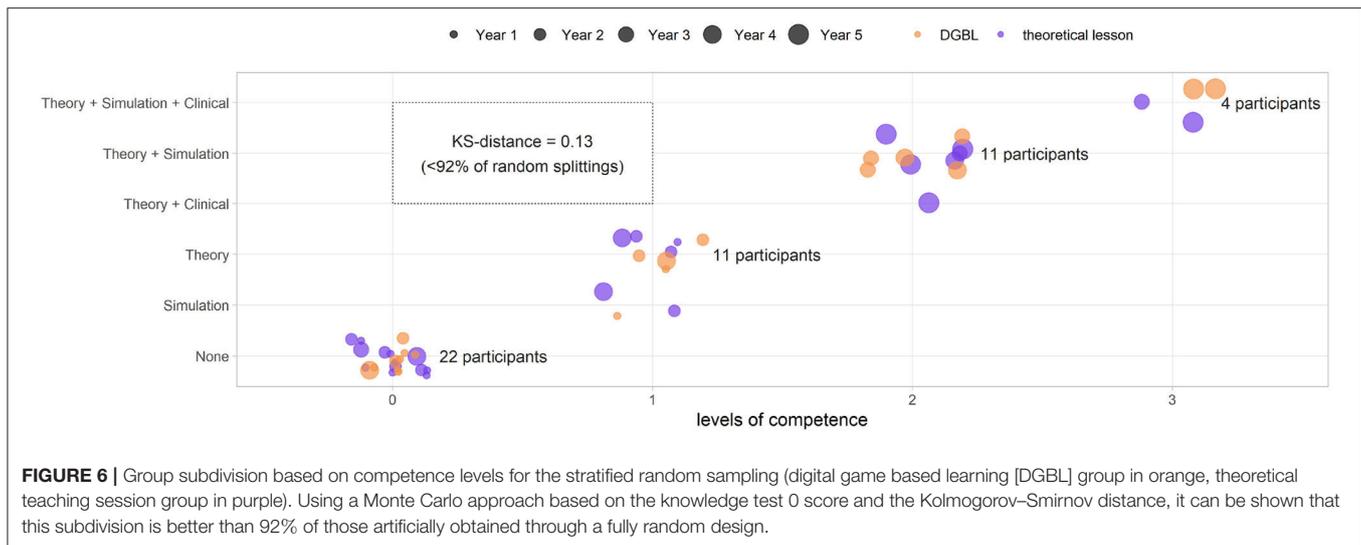
RESULTS

Participant Characteristics and Stratified Random Sampling

Sixty-three pediatric/neonatology residents from the Azienda Ospedaliera Universitaria Pisana (AOUP) were recruited for the study, ranging from the first to the fifth specialty year with a high variability in previous training. The level of competence of each resident depends on the experience acquired before the start of the analysis (year of specialty, practice using a simulator, having attended theoretical training, and also real clinical practice with newborns). These nuisance variables (i.e., a variable that may alter the outcome of the study but is of limited interest in the chosen design) were of no interest to the study and had to be controlled to ensure homogeneity of the two groups using the stratified random sampling. By applying the Monte Carlo approach against the Kolmogorov–Smirnov distance calculated with the chosen design ($d = 0.13$), only 8% of the random subdivisions thus generated show a distance $D < d = 0.13$, confirming the validity of the design used.

Furthermore, the validity of the study design was tested also by comparing the knowledge test 0 and the check equipment scores between the two groups: no statistically significant differences were found (two-sided Mann–Whitney U -test $p = 0.21 \gg 0.05$ and two-sided independent t -test $p = 0.51 \gg 0.05$ for equipment score). Furthermore, the distributions of both values were also not dissimilar (two-sided Kolmogorov–Smirnov test, $p \gg 0.5$). The experiment design, and the corresponding subdivision of the population in strata, allowed to obtain a homogeneous level of past experience (as shown by the level of competence in **Figure 6**). The two groups were therefore considered uniform in the baseline scores (knowledge test 0) and homogeneously subdivided according to the confounding variables.

The design led to two groups uniform in terms of previous experiences (**Figure 7**). Candidates who dropped out for personal



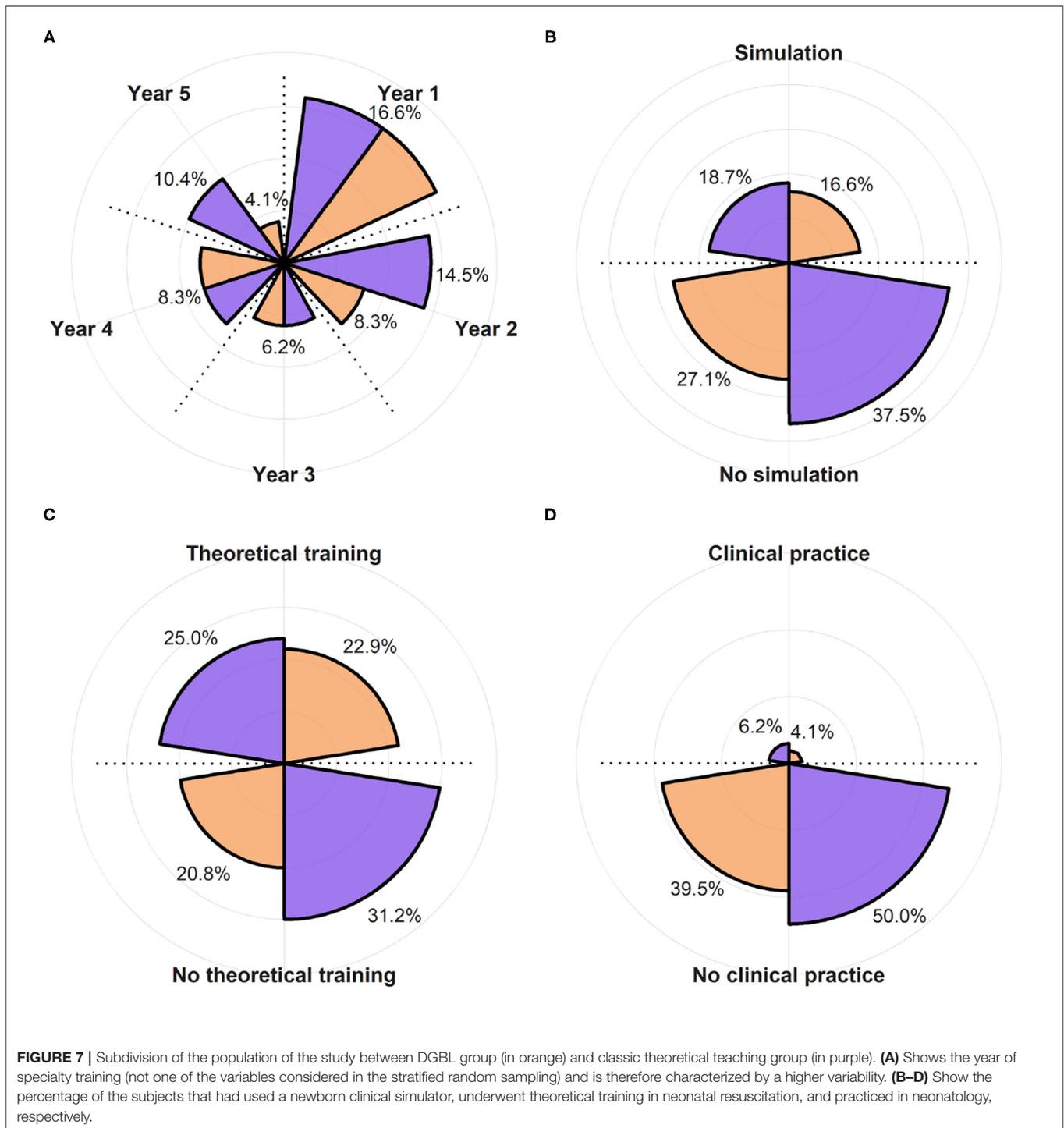
reasons, or those failed to meet learning and testing sessions deadlines, were excluded from the study: of a total of 15, the majority affected DGBL group, yielding 27 candidates for the classic learning group and 21 for DGBL group. In the breakdown of the study sample by specialty year, 56% of the residents clustered around first and second year (**Figure 7A**), only 35.3% of the trainees had practiced at the simulator before this study (**Figure 7B**), whereas 47.9% had already received theoretical training in neonatal resuscitation (**Figure 7C**). User characteristics that could significantly impact results (e.g., neonatal clinical experience, as shown in **Figure 7D**) were uncommon in this cohort (only 10.3% of candidates); this setting required a proper design in order to prevent concentrating the few candidates with any particular characteristic in only one of the two groups. Consequently, the reference sample can be described as having a dominant component of students of the first years, mostly with no previous experience (45.8%). The older residents were the ones with greater medical experience (clinical/simulation/theoretical), with all fifth-years students having received at least one theoretical teaching session and one practical tutorial at the simulator.

Comparison Between DGBL and Classic Learning Knowledge Retention

The first analysis was based on the scores obtained in the knowledge tests 0, 1, 2 (respectively, pre-training, 1 day post-test, and 28 days later follow-up). None of the observed test score distributions could be assumed to be normal except pre-training scores (Shapiro–Wilk test, $\alpha = 0.05$) as shown in **Figure 8** (purple for classic learning and orange for DGBL approach).

After a preliminary Friedman test ($\alpha = 0.05$) that found differences in scores between the knowledge tests 0, 1, 2 for both the DGBL ($p \ll 0.001$) and the theoretical training ($p \ll 0.001$), we moved on to the post-hoc pairwise analysis. The effectiveness of the theoretical teaching session was proved by

an increase in pre-training and post-training tested scores at 1 day, with an increase in median scores from 42.8 to 71.4% (paired one-sided Wilcoxon signed-rank test, $p \ll 0.001$). An even greater increase in scores was found for DGBL training, with median scores ranging from 42.8% pre-training to 83.8% post-training (paired one-sided Wilcoxon signed-rank test, $p \ll 0.001$). There was no statistically significant reduction in scores following the 28-day wait ($\alpha = 0.05$). Even considering a conservative Bonferroni correction factor ($m = 2$) to control the family-wise error rate, the reported results have much lower p -values than the corrected $\tilde{\alpha} = \alpha/m = 0.025$. The initial pre-training scores could be considered coincident both as medians (two-sided Mann–Whitney U -test, $p = 0.21 \gg \alpha = 0.05$) and as distributions (two-sided Kolmogorov–Smirnov test, $p = 0.97 \gg \alpha = 0.05$). This allowed to compare the score increases for the two methodologies. Therefore, considering the post-/pre-training score differences, DGBL method was statistically not inferior to the classic teaching session (one-sided Mann–Whitney U -test, $p = 0.005$). As represented graphically in **Figure 8** (28 days), score variance decreased between pre-training and post-training (1 day) for both methodologies (one-sided paired Pitman–Morgan test, $p \ll \alpha = 0.001$). There was no statistically significant variance increase 28 days post-learning for DGBL group ($p = 0.07 > \alpha = 0.05$), while variance increased significantly for the classic methodology group ($p = 0.02 < \alpha = 0.05$). Furthermore, the variance at 28 days for the classic learning group was greater than that of DGBL group, with values more distributed over the score range (one-sided F -test, $p = 0.03 < \alpha = 0.05$). The variance of the analyzed scores makes it possible to distinguish between a population with a homogeneous knowledge (low variance) compared to one with marked differences between the scores of the individuals (high variance). For this reason, we want to investigate whether following learning there is a simple increase in scores, which is an indication of an effective transmission of knowledge, or even a consequent reduction in the variance of scores, that is representative of uniformity of skills following



learning (e.g., we were able to teach them what we wanted to teach them).

Equipment Game

Equipment scores were divided into three categories: totally correct, partially correct, and incorrect. Learning was considered to be effective if users selected a greater number of correct

options and fewer of the incorrect/partially correct ones. The scores evaluated at steps 0, 1, 2 (respectively, pre-training, 1 day post-training, 28 days follow-up) of classic learning (in purple, **Figure 9A**) and DGBL methodology (in orange, **Figure 9B**) are shown in **Figure 9**. All the score distributions were non-normal, except scores for the correct tools at the 0/pre-training evaluation (Shapiro–Wilk, $\alpha = 0.05$). A preliminary Friedman

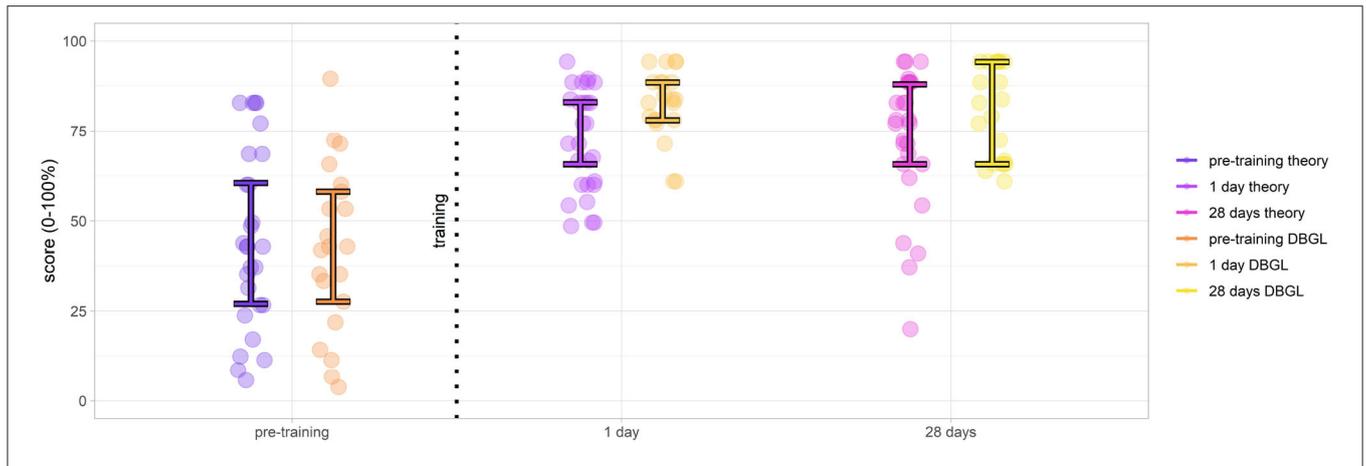


FIGURE 8 | Results of knowledge tests evaluated pre-training, 1 day post-training, and at 28 days follow-up (score medians and middle 50% interquartile ([0.25, 0.75]); theoretical teaching session group scores in purple, digital game based learning (DGBL) method group scores in orange). Although pre-training groups are comparable, post-training scores demonstrate the effectiveness of both methodologies and DGBL, in particular.

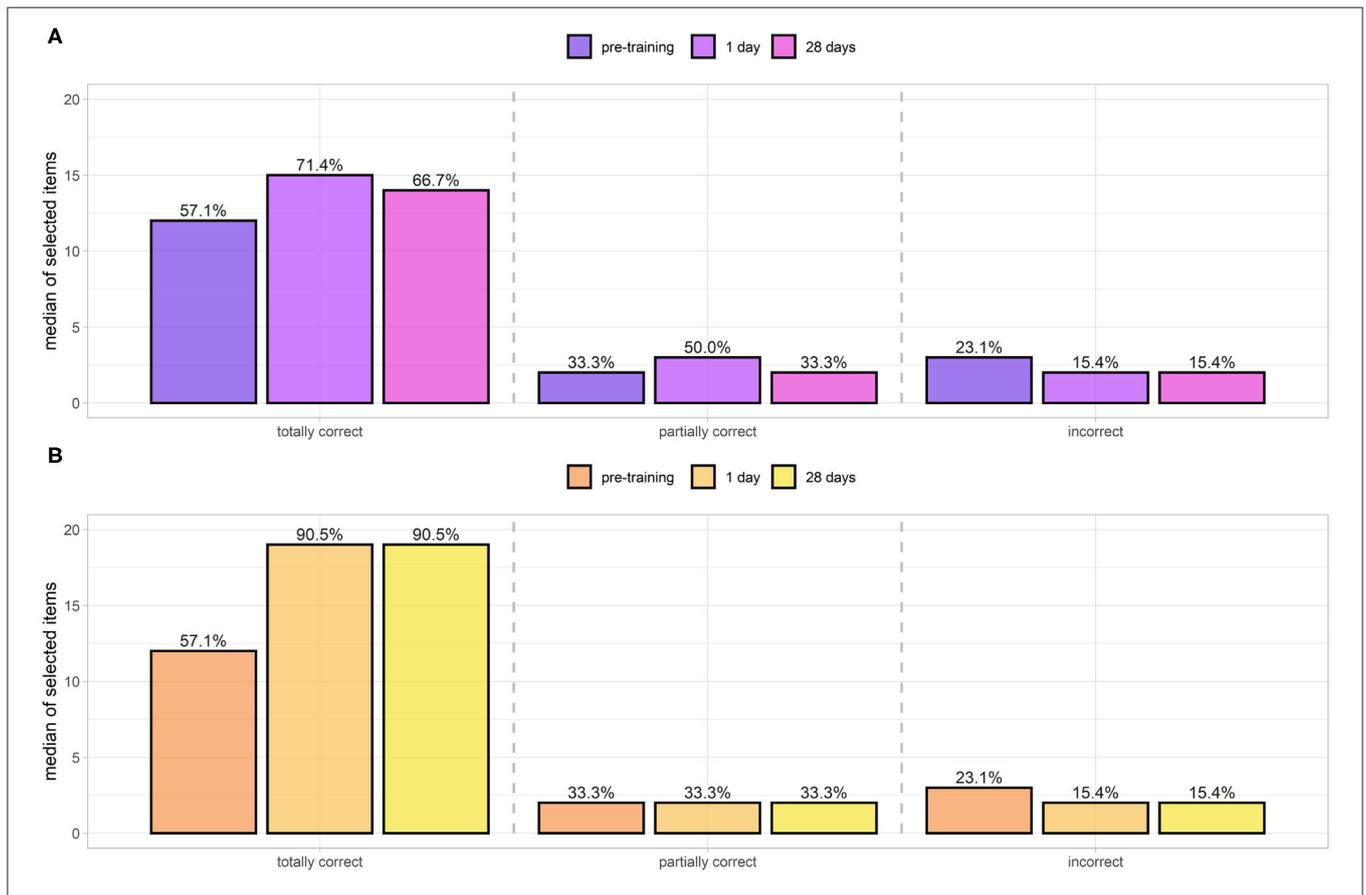


FIGURE 9 | Equipment scores of totally correct/partially correct and incorrect items selected by the control group (standard teaching session, **A**) and digital game based learning (DGBL) group (**B**). Initial equipment scores for the two methodologies are not statistically different. DGBL methodology leads to a greater increase in correct item selection. It also reduces selection of incorrect/partially correct items, whereas after theoretical teaching no reduction is observed.

test ($\alpha = 0.05$) is performed to detect if there is a difference among the three assessments (knowledge test 0, 1, 2) for both

DGBL/classic learning and for the totally/partially correct and incorrect items. A statistical significance of the learning effect

is only found for the totally correct items (classical theoretical learning, $p = 0.02$) and for both totally correct ($p \ll 0.001$) and incorrect ($p = 0.007$) items score for the DGBL training. Classic learning (Figure 9A) was effective in achieving memorization of totally correct objects (57.1 to 71.4%, paired one-sided Wilcoxon signed-rank test $\text{con } p \ll 0.001$). No other statistically significant improvement ($\alpha = 0.05$) was noted in any of the other scores, either in relation to the 1 or 28-day assessment. Conversely, there was an increase in the partially correct objects chosen in Test 1 (33.3 to 50.0%, $p = 0.01$). The DGBL methodology proved more effective (Figure 9B), with not only a statistically significant improvement in pre-training/ 1-day scores for correct items (57.1 to 90.5%, paired one-sided Wilcoxon signed-rank test $\text{con } p \ll 0.001$), but also with a moderate a reduction of incorrect items (23.1 to 15.4%, $p = 0.03$), which is not statistically relevant for the classical learning method. The initial scores for the correct objects coincide for the two groups for both the median (57.1%, paired two-sided Wilcoxon signed-rank test $\text{con } p = 0.72 \gg 0.05$) and the mean values (59.2% theoretical teaching session, 56.7% DGBL, two-sided independent t -test, $p = 0.51 \gg \alpha = 0.05$). As the coinciding baselines allow an analysis of the pre-/post-training differences of the two groups, the DGBL methodology led to a significantly greater improvement than the classic learning one (one-sided Mann-Whitney U -test, $p = 0.009 < \alpha = 0.05$). We did not carry out the same analysis for partially correct and incorrect objects, as uniformity between the two strategies cannot be assumed at the $\alpha = 0.05$ level. In this analysis, considering a Bonferroni correction factor ($m = 2$) did not change the result of the effect of training on score of correct items. However, the effect of the reduction of incorrect items for the DGLB group can no longer be considered statistically significant.

Item Choice in Equipment Check

With regard to the total number of times that each tool was selected during the knowledge tests (regardless of the learning mode), training methodology can be improved. Indeed, Figure 10 highlights the elements for which the methodology worked well (increasing values for totally correct from the center outwards, and decreasing for partially correct/incorrect ones) and those for which it does not (stable scores among the sessions). For almost all totally correct options, learning proved effective with both methodologies; however, some options were too obviously correct, e.g., adrenaline administration (45–47/47) or pulse oximeter use (41–40/44). The best training effect was seen on discouraging the selection of endotracheal tubes (ET) (0, 1) (30–19/17). For other incorrect items (intensive care ventilator, E.R. Bag, E.T. tube size) and partially correct ones [laryngeal mask airway (31–33/33), check neonatal incubator (37–43/43), E.R. bag (20–18/15), and intensive care ventilator (19–25/27)], the learning was not effective enough, as users continued to rate them as necessary despite training indicating otherwise. We are planning the implementation of software changes, which will allow to investigate communication effectiveness for these learning objectives. It should be emphasized that some incorrect tools [ultrasound machine, ultrasound probe, LISAcath(R)] proved poor distractors, as users hardly ever

selected them. Therefore, future versions of the game will not include those items.

DGBL Game Performance Knowledge Retention

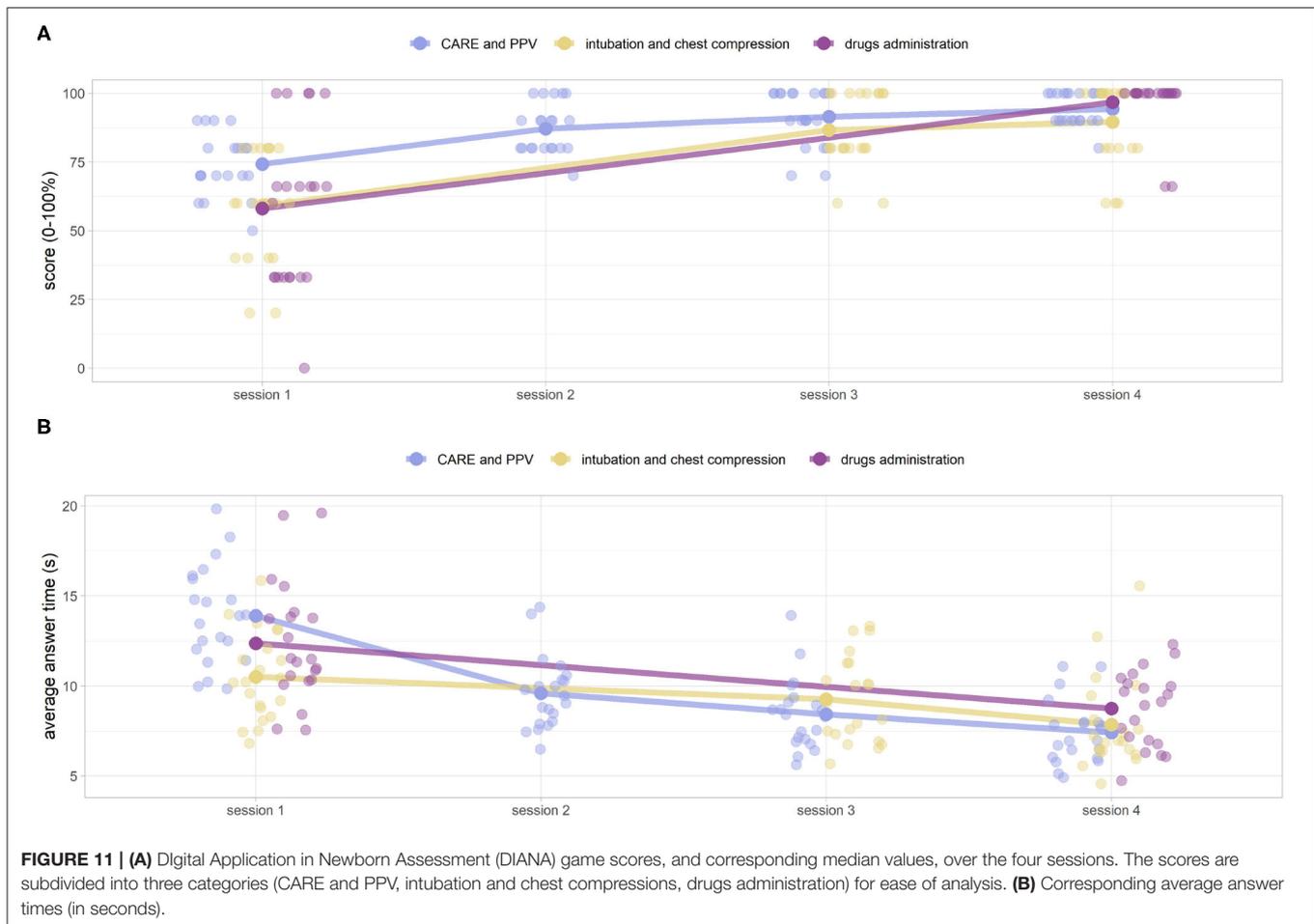
Figure 11A shows the scores and the respective averages of the game sessions (e.g., the number of correct answers) by category and session number (1, 2, 3, 4). The medians response times (seconds) for the entire corresponding series of questions are shown in Figure 11B. After three preliminary Friedman tests for scoring (CARE and PPV, intubation and chest compression, and drug administration) and three more for answer times ($\alpha = 0.05$), all identifying a statistical difference, we moved on to post hoc pairwise analysis. Both panels show a strong monotonicity in the functions, with increasing scores (~ 65.3 to 96.7) and decreasing times (~ 11.9 to 7.7 s) as the sessions progress (one sided paired Wilcoxon signed-rank test, $\alpha = 0.05$), except for equipment CARE and PPV/intubation and chest compression scores between session 3 and session 4 ($\alpha = 0.05$), which did not show a statistically significant increase (knowledge plateau). After a Bonferroni correction by a factor ($m = 5$) for the CARE and PPV and a factor 3 for intubation and chest compression, the same results remained valid at a level of $\tilde{\alpha} = \alpha/m = 0.01$ except for the CARE and PPV scores between session 2 and 3, whose increase was no longer statistically significant. Test values at sessions 1 and 4 highly correlate ($\rho = 0.84$, Pearson test for linear correlation $p \ll 0.001$) with knowledge test scores 0 and 4. The game scores are therefore predictive of success in the following knowledge test.

Equipment Game

As anticipated by knowledge tests (Figure 8), the sessions improve users' ability to choose the correct objects. Indeed, the number of totally correct objects selected (Figure 12A) and partially correct/incorrect ones (Figure 12B), respectively, increased and decreased after each session. Specifically, since all the scores are not normal (Shapiro-Wilk test at the alpha level = 0.05), we proceeded to test the monotony of the score with a non-parametric test (preliminary Friedman test at a level $\alpha = 0.05$ that revealed a statistical difference between the totally correct scores, and *post-hoc* one-sided paired Wilcoxon signed-rank test, respectively, $p \ll 0.001$, $p = 0.01$, $p = 0.003$ between sessions 1–2, 2–3, and 3–4). Unlike procedure memorization highlighted by the scores (Figure 11), there is still a statistically significant improvement for this game between sessions 3 and 4 ($p = 0.003$). A Bonferroni correction factor $m = 5$, the number of pairwise analysis carried out, can be applied ($\tilde{\alpha} = \alpha/m = 0.01$). Despite the correction, the results presented remain unchanged. Furthermore, candidates made fewer mistakes when the tool name was paired with its picture (medians of 0 for partially correct items and incorrect ones vs. 33.3 and 15.4% for the same students during the knowledge test), as shown in Figures 12B, 9B, respectively.

Ventilation and Compression Game

Figure 13 shows the deviation from the correct ventilation (40–60 breaths per minute = intervals between 100 and 150



statistical point of view, the values of the scores and the STDs of the rates all follow non-normal distributions (Shapiro–Wilk test, $p < 0.001$ for scores and $p < 0.01$ for STDs). Regarding the ventilations there is an improvement of both parameters. Score values are decreasing with monotony (one tailed Wilcoxon signed-rank test $p < 0.01$). With regard to STDs, on the other hand, there is a statistically significant reduction between the first/second session ($p = 0.03$) and the third/fourth session ($p = 0.001$) but not between the second/third. The same tests, applied to compressions, were all statistically inconclusive.

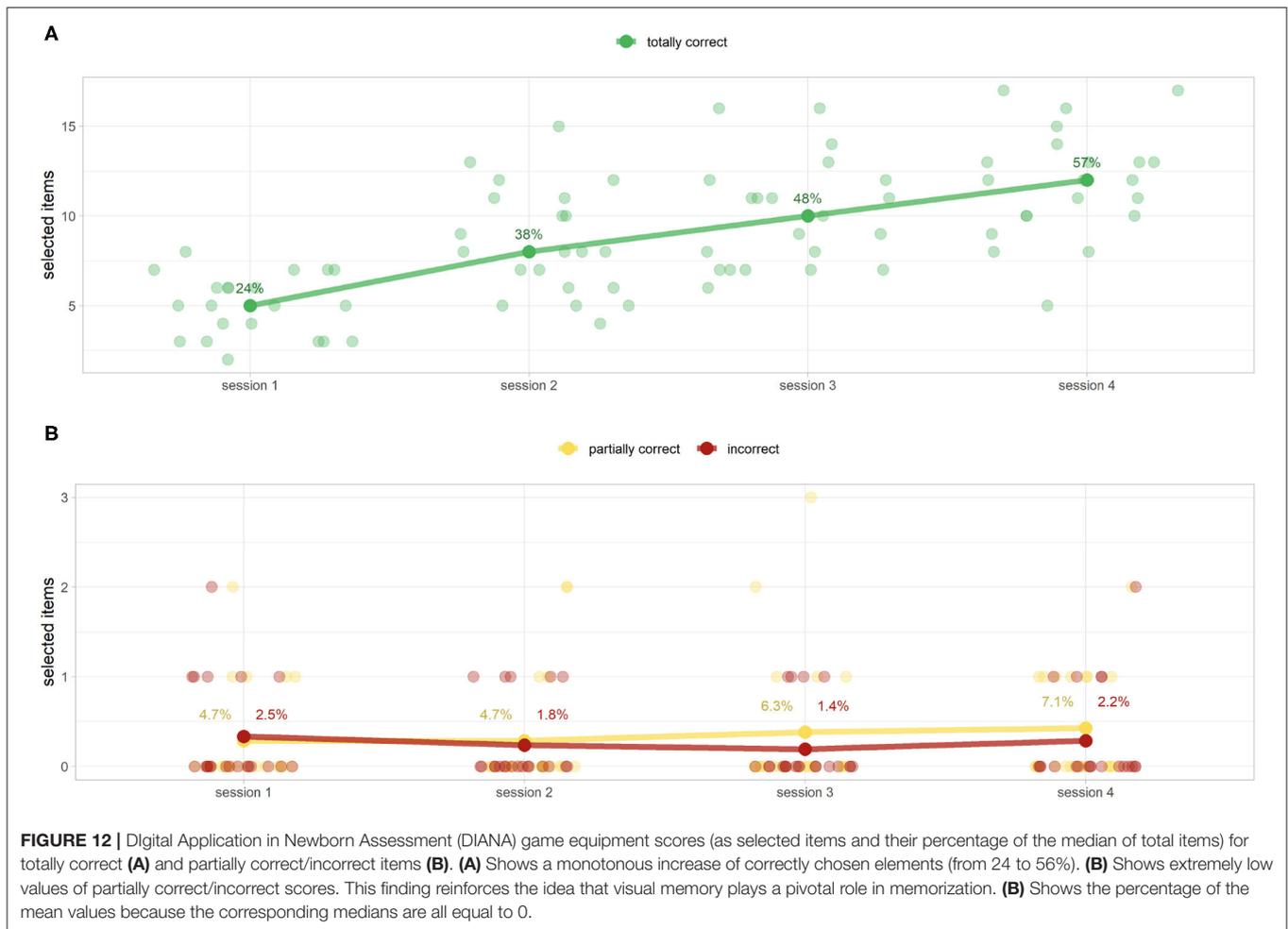
Satisfaction Questionnaires

Of the data collected from DGBL user satisfaction questionnaires (20/21), we evaluated perceived utility and enjoyment of the procedure (both using a five-level Likert scale). Ratings were generally positive in terms of perceived utility [40% (5/5) and 60% (4/5)] and procedure agreeableness [40% (5/5) and 60% (4/5)]. Suggestions mainly concerned the need to increase available equipment game time, perceived as too short. Positive feedback was given on spreading the game sessions over different days.

DISCUSSION

This study successfully applied a DGBL-based approach to neonatal resuscitation teaching through the use of a

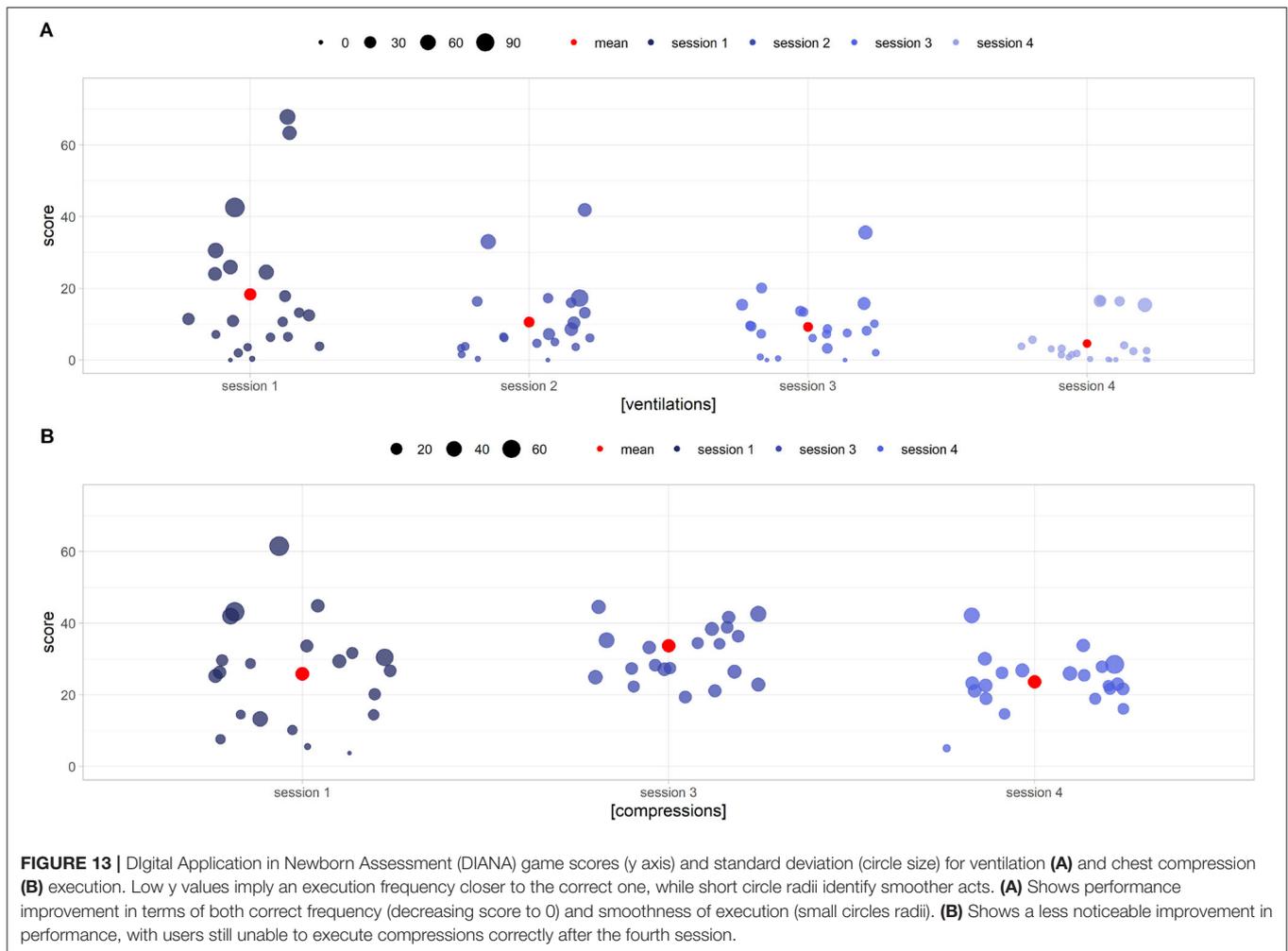
newly developed software (DIANA). DIANA game focused on the entire neonatal resuscitation algorithm (including equipment check, neonatal care, drug administration, assisted ventilation, and chest compressions). The study was aimed at pediatric/neonatology residents [a learners' category considered in few studies as in the mixed study group by (12)], while the majority of previous findings in this field have focused on undergraduate medical students (9, 19, 20), healthcare professionals (11, 12, 21), and experienced neonatal providers (17, 21). This study's sample size is similar to other DGBL studies in the medical/neonatal field (9, 10, 17, 28, 29). Learner allocation (Stratified random sampling) to two experience-based groups (year of specialty training, previous theoretical teaching session, simulation experience, clinical practice experience) proved effective in obtaining homogeneous baseline scores (Section 3.1). Furthermore, the subdivision obtained was better than 92% of those eventually obtained by applying a completely random method. This study is among the few that (1) fully exploit the ability of a game to extract user data (e.g., ventilation/compression game scores, response time, etc.), (2) define a treatment group and an independent control group through a baseline score (pre-test), and (3) evaluate two follow-ups (short- and long-term knowledge retention). In addition, compared with the majority of published studies, which tested learner months apart (9, 10, 17), we preferred to keep the testing



interval shorter [and yet longer than 2 weeks, in line with best practice in assessing DGBL learning (30)]; as the studied cohort was recruited among pediatric/neonatology residents, specialty training would invariably continue to provide reinforcement of the skills assessed. It should be noted that 28 days are considered a sufficient timeframe to evaluate memorization of a procedure in the medium to long term (25).

The DGBL methodology proved to be useful and appreciated by users to teach both neonatal resuscitation algorithm and ventilation execution. Furthermore, it proved to be even more effective than the classic frontal teaching session for both short-term procedure memorization and equipment game score. In particular, the scores related to short-term knowledge retention proved to be higher than those obtained by the theoretical frontal teaching session, in line with the existing literature (10, 19, 20). Also in line with the limited number of studies with a follow-up at more than 28 days (9, 10, 17), long-term knowledge retention for DGBL group was as good as the control group one. Furthermore, candidates who had received classic training demonstrated a regression to lower scores, unlike DGBL methodology learners. DGBL methodology was particularly effective in the learning of clinical equipment checking. Although the classic theory teaching

session led to a statistically significant, but moderate increase in the number of correct objects chosen (57.1–71.4%), DGBL-based approach led to a much greater improvement (57.1–90.5%). Furthermore, while the classic teaching session had almost no effect on changing the scores for partially correct/incorrect items (33.1–50.0 and 23.1–15.4%), DIANA game reduced or leave unchanged scores for both partially correct (33.1–33.1%) and incorrect tools (23.1–15.4%). This score discrepancy is likely due to the difference in the way the learning objective is conveyed: DGBL approach breaks learning into sub-games (one of which is the explicit teaching of which tools should be used), while during a theory teaching session the tools are named progressively at the time of their use. Overall, the DGBL methodology subdivision of learning into multiple sessions was confirmed to be effective for the learning of neonatal resuscitation in line with further previous simulation-based studies (28, 29), especially for the maintenance of the acquired competencies (31). The information collected in the DGBL sessions allows performance analysis (learning curve) related to flowchart learning, response times, equipment check, and timing of assisted ventilations/chest compressions. Procedure learning was effectively achieved, in line with the existing literature (9, 17, 21): the first three sessions



showed significant improvements in learning, while the fourth highlighted a learning plateau. Of note, there was a constant improvement in response times along the four sessions, with a total reduction of more than 30% of the initial one. Similarly, there was a steady improvement in the correct equipment check score (from 24 to 57%). In the assisted ventilation game, DGBL methodology proved to be effective, as residents responded to the feedback from the game and learned to keep the correct rate independently. However, in the chest compressions game, similar in execution to the assisted ventilation one, we did not observe the same effectiveness; candidates did not improve in either the frequency (remaining outside the required clinical range) or the regularity of compressions. This pattern persisted across all four sessions. The discrepancy between these two results could be induced by the differences between the two games. Indeed, during the compression game, the user must interact with the virtual assistant which performs ventilation. To complete the task before next the ventilation, users tend to perform excessively clustered and irregular compressions. To facilitate the reading of the discussion presented above, the results and consequences of the study are shown in **Table 2**.

The administration of a user satisfaction questionnaire confirmed a greater appreciation for DGBL as a training methodology than the classic frontal theory teaching session, in line with the existing literature (13, 19, 20). DGBL methodology usability is crucial for future developments, as learners positively disposed to digital tools tend to respond more effectively (26). Based on the satisfaction questionnaire results, appreciation was lower for the check equipment game compared to the others, despite its effectiveness on improving user scoring.

One of the limitations of this study is the inability of digital software to teach the execution of technical skills. Particularly for complex tasks (also to be combined with another operator), such as chest compressions, this methodology proved ineffective: users' acts remained too frequent, inappropriately clustered and not coordinated with the virtual assistant. Furthermore, the knowledge test does not guarantee that users will apply those skills effectively in a clinical context. Future versions of this software will be developed from the analysis presented in this study and the suggestions collected through the satisfaction questionnaire. Specifically, we aim to reduce the number of sessions to 3 (learning plateau detected at the fourth one),

TABLE 2 | Summary of study results.

Comparison	Results
Primary endpoint	The DGBL methodology proved to be even more effective than the classic frontal teaching session for both short-term procedure memorization and equipment game score. Long-term knowledge retention for the DGBL group was as good as that of the control group (classical training).
Secondary endpoints	The answer time (seconds) decreased after each session. The score of the games increased for the first three sessions and then reached a plateau.
	The sessions improved users' ability to choose the correct equipment.
	There was a statistically significant improvement in the execution (rate) of ventilation after each session.
	There was no statistically significant improvement in the execution (rate) of chest compression
	Ratings were positive in terms of perceived utility and procedure agreeableness.

allow no time limitation for the equipment check game, and exclude from the tool list the obviously incorrect options (poor distractors). To overcome the limitation of learning technical skills in DGBL methodology, future developments may require integration with a physical support structure to allow the candidate to practice clinical tasks. To improve the application of these training methodologies, we are developing the online implementation of DIANA (both in Italian and English) to allow the autonomous use of DIANA in further medical realities, as a free tool for training and re-training. Because of the online platform, we are already extending the same analysis on a wider population. In this way, we can use our findings (on both population characteristics and expected scores) to estimate the required sample size to improve future studies. The future development of a hardware device for the execution of practical skills will also allow to overcome a known limitation in simulation field (i.e., by lack of a report on the technical performance of the user with high-fidelity mannequins). A high-fidelity simulator could offer a report on the correct execution of the flowchart based on human external observation of simulation. However, with a hardware device designed to record the events performed, both in terms of decision-making and practical performance, it will be possible to conduct a more detailed and precise study of the effectiveness of these two training methodologies.

We will also seek to modify the software with/without hardware integration to widen the potential user base, including other clinical specialties and varying levels of experience. In particular, we aim to extend this learning tool to users less accustomed to digital technology to further assess the impact of user mindset on the effectiveness of DGBL methodology (26). Moreover, as DGBL is unlikely to be adopted as a stand-alone teaching method (11) [especially in higher education (14)], future research may involve using the two methods in sequence, e.g., reinforcing the classic theory teaching session by DGBL, or a simulator-based introduction to a classic teaching session. This

blended approach has been already validated for simulations outside neonatology (32). Considering the positive feedback obtained by remotely testing DGBL in other healthcare education contexts (17), deployment of DGBL to support healthcare education in low-income countries could represent another future development in the use of this learning technology.

CONCLUSION

In this study, DGBL methodology for pediatric/neonatology resident training proved to be superior to theoretical teaching session (led by a neonatal expert trainer) on short- and long-term knowledge retention of memorization of the correct equipment to assemble. In addition, DGBL proved to be at least as effective as the teaching lesson for memorization and retention of neonatal resuscitation algorithm. DIANA game allows individual user session analysis, with an improvement in "session-after-session" scores and a reduction in decision-making times. We propose that DGBL could be a valuable addition to classic learning methodology for all medical procedures involving a procedural algorithm.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Dipartimento Materno-Infantile, Azienda Ospedaliero Universitaria Pisana. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SB fully implemented the software DIANA, collected the data, drafted the initial manuscript, reviewed, and revised the manuscript. GD designed the study and the data collection tools, analyzed the data, carried out statistical analysis, drafted the initial manuscript, reviewed, and revised the manuscript. MD and FL collected the data, were involved in the classic training sessions, and revised the manuscript. RTS, MC, and AC wrote the knowledge test, taught during the classic training sessions, integrated the analysis with the corresponding medical discussion, and reviewed and revised the manuscript. NF reviewed, critically analyzed, and revised the manuscript including revision of English language. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fped.2022.842302/full#supplementary-material>

REFERENCES

- Moshiro R, Mdoe P, Perlman JM. A global view of neonatal asphyxia and resuscitation. *Front Pediatr.* (2019) 7:489. doi: 10.3389/fped.2019.00489
- Wyckoff MH, Wyllie J, Aziz K, de Almeida MF, Fabres J, Fawke J, et al. Neonatal life support: 2020 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. *Circulation.* (2020) 142(16 Suppl. 1):S185–221. doi: 10.1542/peds.2020-038505C
- McCarthy LK, Morley CJ, Davis PG, Kamlin COF, O'Donnell CP. Timing of interventions in the delivery room: does reality compare with neonatal resuscitation guidelines? *J Pediatr.* (2013) 163:1553–7. doi: 10.1016/j.jpeds.2013.06.007
- Larsen DP, Butler AC, Roediger HL III. Test-enhanced learning in medical education. *Med Educ.* (2008) 42:959–66. doi: 10.1111/j.1365-2923.2008.03124.x
- Pesare E, Roselli T, Corriero N, Rossano V. Game-based learning and gamification to promote engagement and motivation in medical learning contexts. *Smart Learn Environ.* (2016) 3:1–21. doi: 10.1186/s40561-016-0028-0
- Sliney A, Murphy D. JDoc: A serious game for medical learning. In: *First International Conference on Advances in Computer-Human Interaction*. Sainte Luce (2008). p. 131–6. doi: 10.1109/ACHI.2008.50
- Eichenbaum H. Memory Systems, Vol. 3. In: Byrne JH, editor. *Learning and Memory: A Comprehensive Reference*. 2nd ed. Oxford: Academic Press. p. 295–312. doi: 10.1016/B978-0-12-809324-5.21085-7
- Healy AF, Bourne LE. *Learning and Memory of Knowledge and Skills*. Thousand Oaks, CA: Sage Publications, Inc (1995). doi: 10.4135/9781483326887
- Hu L, Zhang L, Yin R, Li Z, Shen J, Tan H, et al. NEOGAMES: a serious computer game that improves long-term knowledge retention of neonatal resuscitation in undergraduate medical students. *Front Pediatr.* (2021) 9:645776. doi: 10.3389/fped.2021.645776
- Rondon S, Sassi FC, de Andrade CRF. Computer game-based and traditional learning method: a comparison regarding students-knowledge retention. *BMC Med Educ.* (2013) 13:30. doi: 10.1186/1472-6920-13-30
- Yeo CL, Ho SKY, Tagamolila VC, Arunachalam S, Bharadwaj SS, Poon WB, et al. Use of web-based game in neonatal resuscitation-is it effective? *BMC Med Educ.* (2020) 20:1–11. doi: 10.1186/s12909-020-02078-5
- Akl EA, Sackett KM, Pretorius R, Bhoopathi PSS, Mustafa R, Schünemann H, et al. Educational games for health professionals. *Cochrane Database Syst Rev.* (2008) 1:CD006411. doi: 10.1002/14651858.CD006411.pub2
- Rutledge C, Walsh CM, Swinger N, Auerbach M, Castro D, Dewan M, et al. Gamification in action: theoretical and practical considerations for medical educators. *Acad Med.* (2018) 93:1014–20. doi: 10.1097/ACM.0000000000002183
- Brown CL, Comunale MA, Wigdahl B, Urdaneta-Hartmann S. Current climate for digital game-based learning of science in further and higher education. *FEMS Microbiol Lett.* (2018) 365:fny237. doi: 10.1093/femsle/fny237
- McGaghie WC, Issenberg SB, Cohen MER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. *Acad Med.* (2011) 86:706. doi: 10.1097/ACM.0b013e318217e119
- Harrington CM, Chaitanya V, Dicker P, Traynor O, Kavanagh DO. Playing to your skills: a randomised controlled trial evaluating a dedicated video game for minimally invasive surgery. *Surg Endosc.* (2018) 32:3813–21. doi: 10.1007/s00464-018-6107-2
- Ghoman SK, Schmölzer GM. The RETAIN simulation-based serious game—a review of the literature. In: *Healthcare (Basel)*. Vol. 8. Multidisciplinary Digital Publishing Institute (2020). p. 3. doi: 10.3390/healthcare8010003
- Ghoman SK, Patel SD, Cutumisu M, von Hauff P, Jeffery T, Brown MR, et al. Serious games, a game changer in teaching neonatal resuscitation? A review. *Arch Dis Childhood-Fetal Neonatal Edn.* (2020) 105:98–107. doi: 10.1136/archdischild-2019-317011
- Swiderska N, Thomason E, Hart A, Shaw B. Randomised controlled trial of the use of an educational board game in neonatology. *Med Teach.* (2013) 35:413–5. doi: 10.3109/0142159X.2013.769679
- Boeker M, Andel P, Vach W, Frankschmidt A. Game-based e-learning is more effective than a conventional instructional method: a randomized controlled trial with third-year medical students. *PLoS ONE.* (2013) 8:e82328. doi: 10.1371/journal.pone.0082328
- Cutumisu M, Patel SD, Brown MR, Fray C, von Hauff P, Jeffery T, et al. RETAIN: a board game that improves neonatal resuscitation knowledge retention. *Front Pediatr.* (2019) 7:13. doi: 10.3389/fped.2019.00013
- Erhel S, Jamet E. Digital game-based learning: impact of instructions and feedback on motivation and learning effectiveness. *Comput Educ.* (2013) 67:156–67. doi: 10.1016/j.compedu.2013.02.019
- All A, Castellar EPN, Van Looy J. Assessing the effectiveness of digital game-based learning: best practices. *Comput Educ.* (2016) 92:90–103. doi: 10.1016/j.compedu.2015.10.007
- Donovan JJ, Radosevich DJ. A meta-analytic review of the distribution of practice effect: now you see it, now you don't. *J Appl Psychol.* (1999) 84:795. doi: 10.1037/0021-9010.84.5.795
- Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychol Bull.* (2006) 132:354. doi: 10.1037/0033-2909.132.3.354
- Cutumisu M, Ghoman SK, Lu C, Patel SD, Garcia-Hidalgo C, Fray C, et al. Health care providers' performance, mindset, and attitudes toward a neonatal resuscitation computer-based simulator: empirical study. *JMIR Ser Games.* (2020) 8:e21855. doi: 10.2196/21855
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team (2020). Available online at: <https://www.R-project.org/>
- Cordero L, Hart BJ, Hardin R, Mahan JD, Nankervis Ca. Deliberate practice improves pediatric residents' skills and team behaviors during simulated neonatal resuscitation. *Clin Pediatr.* (2013) 52:747–52. doi: 10.1177/0009922813488646
- Sawyer T, Sierocka-Castaneda A, Chan D, Berg B, Lustik M, Thompson M. Deliberate practice using simulation improves neonatal resuscitation performance. *Simul Healthcare.* (2011) 6:327–36. doi: 10.1097/SIH.0b013e31822b1307
- All A, Castellar EPN, Van Looy J. Measuring effectiveness in digital game-based learning: a methodological review. *Int J Ser Games.* (2014) 1:3–20. doi: 10.17083/ijsg.v1i2.18
- Matterson HH, Szyld D, Green BR, Howell HB, Pusic MV, Mally PV, et al. Neonatal resuscitation experience curves: simulation based mastery learning booster sessions and skill decay patterns among pediatric residents. *J Perinat Med.* (2018) 46:934–41. doi: 10.1515/jpm-2017-0330
- Bonde MT, Makransky G, Wandall J, Larsen MV, Morsing M, Jarmer H, et al. Improving biotech education through gamified laboratory simulations. *Nat Biotechnol.* (2014) 32:694–7. doi: 10.1038/nbt.2955

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Bardelli, Del Corso, Ciantelli, Del Pistoia, Lorenzoni, Fossati, Scaramuzza and Cuttano. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.