The second group were categorized as intermediates; specialty trainees from all levels, whom had some experience with endoscopy and has done or assisted in at least 250 procedures. The third group was the expert group, which consisted of consultants in Gastroenterology specialty and each one of them at least has done 2500 procedures in order to be eligible to be placed in the group.

Results We have analysed the data of the participants performance collected from the simulator and compared the results of the three groups together. It was clear that the expert group have done better with shorter time than other groups (264.4 sec); intermediates (321.14 sec), and novices (822.05 sec). The results were analysed further using the IBM SPSS® Software. The date generated showed a statistical significance between the groups having a p value of (P< 0.022). Then the experts’ results were isolated to define a set of benchmark ranges for the ERCP Module. The average of experts’ performance was collected, then we have calculated the Standard Deviation of each mean. Later, the mean of each task was trimmed by excluding any consultant performance beyond the standard deviation by 1 ± six out of twelve metrics were considered significant based on literature ad up do date practice which were included; total time of procedure (178.2 – 361.8 seconds), papilla contact before cannulation (2.25 – 3.25 times), number of cannulations to the PD (1 time), number of cannulations to the CBD (1 time), number of contrast injections to the PD (9.6 – 19.6), and number of contrast injections to the CBD (6.6 – 18.4). The new recalculated mean was used to set a reference criterion and a benchmark range for the performance of the ERCP module in the GI Mentor 2 Simulator.

Afterwards, we recruited these results as a reference in our syllabus. Through analysis of operators’ performance and psychological dynamics in practice, we created an evidence-based curriculum that we deemed to be suitable for training ERCP using virtual reality simulator and demonstrated that is possible to define and develop a virtual reality training curriculum for ERCP using structured scientific methodology.

Conclusion The ERCP module in the Simbionix GI Mentor 2 simulator demonstrate face and construct validity as they show statistically significant differences between novice, intermediate, and expert groups as proved in our results and has been done previously in other studies.

We have defined a reference criterion level to develop proficiency performance benchmark for all metrics obtained from our studies on the ERCP procedure based on 5 experts. And we have demonstrated and managed to set a proficiency performance benchmark range in the ERCP module to be used as a baseline when comparing any operator performance on the simulator. Also, our study further our understanding and knowledge of endoscopic expertise and provides trainees with predefined proficiency performance benchmarks that can be used to help and support in their learning of endoscopic skills.

Furthermore, this study has demonstrated that it is possible to define and develop a virtual reality training curriculum for ERCP using structured scientific methodology.