

Artificial Intelligence Takes Vascular Surgery Qualifying Examination: A Comparative Performance Study On The Licensing Exam

Sonal Kumar, BA, Ross University School of Medicine, Miramar, FL

Objective:

Generative artificial intelligence (AI) tools are transforming medical education by offering innovative methods to enhance teaching and learning. Despite their potential, research on their use in vascular surgery is limited. This study aims to evaluate and compare the effectiveness of AI in preparing for the Vascular Surgery Qualifying Exam (VSQE), exploring their potential as educational supplements.

Methods:

From the Vascular Education and Self-Assessment Program version 6 (VESAP6), we selected 269 text-only multiple-choice questions out of 642. We excluded 143 image-based questions. These questions were input into four AI tools: ChatGPT 3.5, Google Gemini, Microsoft Bing, and Claude 3.5. Each question with answer choices was submitted into an incognito window of each AI tool without additional context. A χ^2 test was used to assess if the percentage of correct answers varied by question difficulty and subject, with a significance level of $p < 0.05$. Data analysis was conducted using Stata 18.5.

Results:

Claude 3.5 achieved the highest accuracy with 65.7% percent correct, outperforming Google Gemini (55.3%), ChatGPT (55.0%), and Microsoft Bing (53.9%). While ChatGPT, Google Gemini, and Microsoft Bing did not show significant accuracy variations by topic ($p=0.548$, $p=0.145$, and $p=0.797$, respectively), Claude 3.5 demonstrated significant performance differences across topics ($p=0.001$), mastering lower extremity (86%), dialysis access (80%), cerebrovascular (77%), venous lymph (70%) and vascular medicine (68.9%).

Conclusion:

Claude 3.5 outperformed other AI tools in answering VSQE6 questions and shows promise as a supplementary tool in vascular surgery education. It is particularly effective in areas like lower extremity, dialysis access, and cerebrovascular. However, current AI capabilities may not fully meet the evolving needs of medical education, and future advancements may offer more substantial benefits, though traditional methods remain essential.

	ChatGPT 3.5	Google Gemini	Microsoft Bing	Claude 3.5
Abdominal	66.7%	50.0%	50.0%	66.7%
Aortoiliac disease	53.6%	50.0%	50.0%	64.3%
Cerebrovascular	60.0%	62.9%	60.0%	77.1%
Dialysis Access	43.33%	60%	70.0%	80.0%
Lower Extremity	65.5%	72.4%	51.7%	86.2%
Peripheral arterial	50.0%	50.0%	50.0%	16.7%
Radiation Safety	50.0%	50.0%	61.1%	64.7%
Renal and Mesenteric	39.3%	32.1%	39.3%	35.7%
Upper Extremity	56.7%	53.3%	53.3%	50.0%
Vascular Medicine	72.4%	65.5%	48.3%	68.9%
Venous and Lymphatic	55.6%	55.6%	59.3%	70.4%
Peripheral venous	16.7%	33.3%	20.0%	50.0%

Table 2. Comparison of accuracy on VESAP6 questions by discipline by LLM

Fig 1 Table 2. Comparison of accuracy on VESAP6 questions by discipline by LLM

Fig 1