

Supplemental Table 1: ChatGPT's performance per exam section of each of the BCSC and OphthoQuestions testing sets

	BCSC (n=260)	OphthoQuestions (n=260)	P value*
General Medicine Correct Incorrect	15 (75%) 5 (25%)	15 (75%) 5 (25%)	1.000
Fundamentals Correct Incorrect	11 (55%) 9 (45%)	13 (65%) 7 (35%)	0.748
Clinical Optics Correct Incorrect	12 (60%) 8 (40%)	5 (25%) 15 (75%)	0.054
Pathology and Intraocular Tumors Correct Incorrect	8 (40%) 12 (60%)	9 (45%) 11 (55%)	1.000
Neuro-ophthalmology Correct Incorrect	8 (40%) 12 (60%)	2 (10%) 18 (90%)	0.065
Pediatrics and Strabismus Correct Incorrect	13 (65%) 7 (35%)	4 (20%) 16 (80%)	0.010**
Oculoplastics Correct Incorrect	10 (50%) 10 (50%)	8 (40%) 12 (60%)	0.751
Cornea Correct Incorrect	12 (60%) 8 (40%)	12 (60%) 8 (40%)	1.000
Uveitis Correct Incorrect	7 (35%) 13 (65%)	13 (65%) 7 (35%)	0.113

Glaucoma Correct Incorrect	10 (50%) 10 (50%)	5 (25%) 15 (75%)	0.191
Lens and Cataract Correct Incorrect	13 (65%) 7 (35%)	9 (45%) 11 (55%)	0.341
Retina and Vitreous Correct Incorrect	15 (75%) 5 (25%)	8 (40%) 12 (60%)	0.054
Refractive Surgery Correct Incorrect	11 (55%) 9 (45%)	8 (40%) 12 (60%)	0.527

*Fisher's Exact test

**Statistically significant at the 0.05 level