	MMMU-Pro			MMMU (Val)	Δ_1	Δ_2
	Standard (4 Opts)	Standard (10 Opts)	Vision			
Random Choice	24.9	12.8	12.4	22.1	-9.3	-9.7
Frequent Choice	27.8	12.1	12.1	26.8	-14.7	-14.7
Human Expert (Low)	75.4	73.0	73.0	76.2	-3.2	-3.2
Human Expert (Medium)	82.1	80.8	80.8	82.6	-1.8	-1.8
Human Expert (High)	88.6	85.4	85.4	88.6	-3.2	-3.2
GPT-4o (0513) (OpenAI, 2024a)	**64.7**	54.0	**49.7**	**69.1**	-15.1 (↑ 1)	**-19.4** (-)
Claude 3.5 Sonnet (Anthropic, 2024)	63.7	**55.0**	48.0	68.3	**-13.3** (↓ 1)	-20.3 (-)
Gemini 1.5 Pro (0801) (Reid et al., 2024)	60.6	49.4	44.4	65.8	-16.4 (-)	-21.4 (-)
Gemini 1.5 Pro (0523) (Reid et al., 2024)	57.6	46.5	40.5	62.2	-15.7 (-)	-21.7 (-)
GPT-4o mini (OpenAI, 2024b)	55.3	39.9	35.2	59.4	-19.5 (↑ 1)	-24.2 (↑ 1)
Qwen2-VL-72B (Qwen, 2024)	**59.3**	**49.2**	**43.3**	**64.5**	**-15.3** (-)	-21.2 (-)
InternVL2-Llama3-76B (Chen et al., 2024)	55.0	41.9	38.0	58.3	-16.4 (↓ 1)	**-20.3** (↓ 1)
InternVL2-40B (Chen et al., 2024)	47.4	36.3	32.1	55.2	-18.9 (-)	-23.1 (↓ 1)
LLaVA-OneVision-72B (Li et al., 2024a)	52.3	38.0	24.0	56.8	-18.8 (-)	-32.8 (↑ 5)
Qwen2-VL-7B (Qwen, 2024)	46.6	34.1	27.0	54.1	-20.0 (↑ 1)	-27.1 (↓ 1)
Pixtral-12B (Mistral, 2024)	47.5	33.4	25.0	52.5	-19.1 (↑ 1)	-27.5 (-)
InternVL2-8B (Chen et al., 2024)	42.6	32.5	25.4	51.2	-18.7 (-)	-25.8 (↓ 3)
MiniCPM-V2.6 (Yao et al., 2024)	40.6	30.2	24.2	49.8	-19.6 (↑ 1)	-25.6 (↓ 3)
VILA-1.5-40B (Lin et al., 2024)	46.8	35.9	14.1	51.9	-16.0 (↓ 2)	-37.8 (↑ 9)
LLaVA-NEXT-72B (Liu et al., 2024a)	43.0	31.0	19.2	49.9	-18.9 (-)	-30.7 (-)
LLaVA-OneVision-7B (Li et al., 2024a)	42.8	29.5	18.7	48.8	-19.3 (↑ 2)	-30.1 (↓ 1)
LLaVA-NeXT-34B (Liu et al., 2024a)	44.5	30.3	17.2	48.1	-17.8 (↓ 2)	-30.9 (↓ 1)
Idefics3-8B-Llama3 (Laurençon et al., 2024)	40.8	30.1	15.6	46.6	-16.5 (↓ 1)	-31.0 (-)
Qwen2-VL-2B (Qwen, 2024)	34.8	25.3	17.2	41.1	-15.8 (-)	-23.9 (↓ 3)
Phi-3.5-Vision (Abdin et al., 2024)	37.8	26.3	13.1	43.0	-16.7 (-)	-29.9 (↑ 3)
LLaVA-NeXT-7B (Liu et al., 2024a)	33.7	19.4	14.6	35.3	-15.9 (-)	-20.7 (↓ 3)
LLaVA-NeXT-13B (Liu et al., 2024a)	33.9	19.8	14.5	36.2	-16.4 (-)	-21.7 (↓ 1)

Table 1: Results of models on MMMU-Pro and MMMU (Val). Δ_1: Standard (10 options) - MMMU (Val); Δ_2: Vision - MMMU (Val). (↓) represents a decrease in ranking, while (↑) indicates an increase. The best-performing model in each category is **in-bold**, and the second best is underlined.