

Appendix E — Scoring tables

The following tables present the complete item-level response data for all participants. Table 1 shows responses and total scores for LLM participants under full conditions (phase 1). Table 2 shows responses and total scores for LLM participants under answer-only conditions (phase 2). Table 3 shows responses and total scores for human participants under full conditions. The correct answer for each item is indicated in the header row of each table. All responses are reproduced exactly as recorded, without correction or exclusion.

Table 1

Full conditions, LLMs

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	T
CA	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	
D1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
D2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
D3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
D4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
D5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
Q1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
Q2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	B	C	A	B	B	A	19
Q3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	B	C	A	B	B	A	19
Q4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	B	C	A	B	B	A	19
Q5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
M1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
M2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
M3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
M4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
M5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
G1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
G2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	B	C	A	B	B	A	19
G3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
G4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	B	C	A	B	B	A	19
G5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
X1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
X2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
X3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
X4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
X5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
A1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
A2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
A3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
A4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
A5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20

P1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
P2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
P3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
P4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
P5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
C1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
C2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
C3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
C4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
C5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20

T: Total score

CA: Correct Answers

D: DeepSeek

Q: Qwen

M: Mistral

G: Gemini

X: Grok

A: Claude

P: Copilot

C: ChatGPT

Summary:

795/800 Correct Answers = 99.3%

Table 2

Partial conditions, LLMs

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	T
CA	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	
D1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
D2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
D3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
D4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
D5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
Q1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	B	C	A	B	B	A	19
Q2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	B	C	A	B	B	A	19
Q3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	B	C	A	B	B	A	19
Q4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	B	C	A	B	B	A	19
Q5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	B	C	A	B	B	A	19
M1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
M2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
M3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
M4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
M5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
G1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
G2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
G3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
G4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
G5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
X1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
X2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
X3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
X4	C	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	19
X5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
A1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
A2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
A3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
A4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
A5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20

P1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
P2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
P3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
P4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
P5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
C1	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
C2	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
C3	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
C4	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	20
C5	A	C	A	B	C	B	B	C	A	B	A	C	C	C	B	C	A	B	B	A	19

T: Total score

CA: Correct Answers

D: DeepSeek

Q: Qwen

M: Mistral

G: Gemini

X: Grok

A: Claude

P: Copilot

C: ChatGPT

Summary: 793/800 Correct Answers = 99.1%

Table 3

Full conditions, humans

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	T
CA	A	C	A	B	C	B	B	C	A	B	A	C	C	C	A	C	A	B	B	A	
1	B	B	B	C	A	B	C	A	B	A	C	C	B	C	A	B	A	C	B	A	7
2	A	A	B	B	A	C	C	B	A	A	C	C	B	A	A	A	C	A	C	C	4
3	C	A	B	B	B	B	A	A	C	A	B	A	A	A	C	B	B	C	A	C	2
4	A	A	A	A	B	C	C	A	A	C	C	B	A	C	A	C	C	A	C	A	6
5	B	B	B	A	A	A	C	B	A	A	B	C	A	A	B	B	C	A	A	B	2
6	B	A	A	C	A	A	A	B	C	B	C	A	B	C	A	C	B	B	B	C	7
7	A	B	C	C	A	A	B	B	C	A	C	A	B	A	C	C	A	A	C	C	4
8	B	A	C	A	C	B	B	C	B	B	C	A	C	C	B	B	C	B	A	A	9
9	B	B	C	A	C	A	A	A	C	B	A	B	A	C	C	B	B	C	B	B	5
10	C	B	A	A	C	A	A	B	B	A	A	C	C	A	A	C	B	A	C	A	9
11	A	B	C	A	B	A	C	C	B	C	A	B	B	C	B	A	B	C	A	A	5
12	B	B	C	C	A	A	B	B	C	B	B	A	B	B	C	B	C	C	A	A	3
13	A	B	A	C	C	A	C	A	A	B	C	C	B	C	A	B	B	B	B	A	11
14	B	A	C	C	A	C	B	A	C	C	B	A	B	B	C	C	A	B	A	A	4
15	B	B	C	C	A	B	B	C	B	C	C	B	C	B	B	A	C	A	B	A	6
16	C	C	B	A	C	A	C	B	B	A	C	C	B	B	C	B	C	B	C	C	3

T: Total score

CA: Correct Answers

1-16: Human participants

Summary: $87/320 = 27.2\%$

Mean: 5.4