Assessment report

30 STUDENTS · 50 QUESTIONS · SYNTHETIC SAMPLE

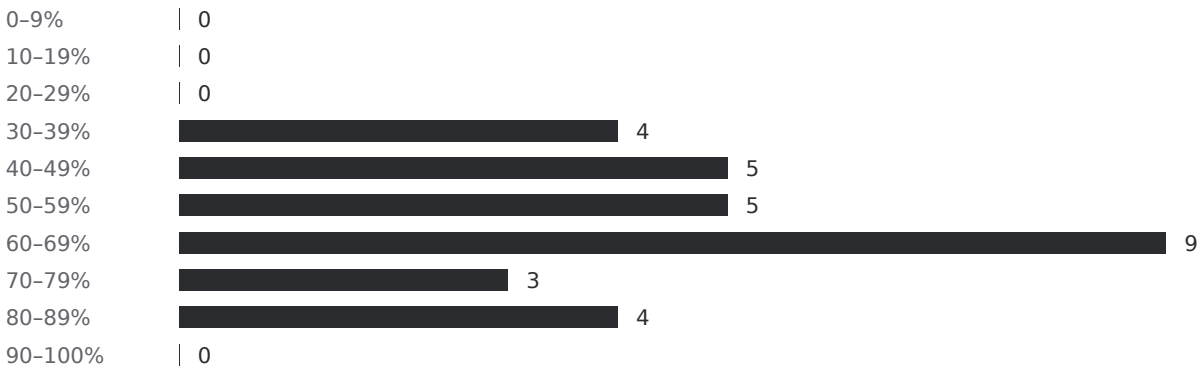
Fictional class example. All figures are calculated from the same 30 response rows in the companion CSV; no real student information is used. This sample adds the FAQ explanations after the analysis.

Class summary

Measure	Result	Measure	Result
Sheets graded	30	Sheets not graded	0
Mean	29 / 49 (59.2%)	Minimum	17 / 49 (34.7%)
Median	30 / 49 (61.2%)	Maximum	43 / 49 (87.8%)
Standard deviation	7.73 points	Reliability (KR-20)	0.84 · Good
Measurement error	3.14 points	Possible score	49 points

Q15 accepts B or D. Q50 has no key and is excluded; blanks and multiple marks on scored questions earn no credit.

Score distribution



Questions to review

Q1 — Everyone correct

Q2 — Possible wrong key: top scorers mostly chose C; Very hard

Q3 — Possible wrong key: top scorers mostly chose D

Q4 — Possible wrong key: top scorers mostly chose A

Q6 — Possible wrong key: top scorers mostly chose E; Very hard

Q14 — Low discrimination

Q20 — Low discrimination

Q22 — Low discrimination

Q32 — Negative discrimination: stronger students did worse

Q42 — Low discrimination

Warnings are prompts to review the answer key and question wording, not automatic decisions to discard a question.

Item analysis

Choice columns count responses; % correct is the rounded share receiving credit. Discrimination is corrected item–total correlation. U–L compares the upper and lower 27% (8 students per group).

Q	Key	% correct	Discrim.	U–L	A	B	C	D	E	Blank	Multi
1	A	100%	—	0.00	30	0	0	0	0	0	0
2	B	13%	-0.49	-0.38	0	4	26	0	0	0	0
3	C	50%	-0.85	-1.00	0	0	15	15	0	0	0
4	D	37%	-0.16	-0.13	19	0	0	11	0	0	0
5	E	53%	0.53	0.75	0	12	0	0	16	1	1
6	A	10%	-0.42	-0.25	3	8	2	6	11	0	0
7	B	67%	0.41	0.63	2	20	2	5	1	0	0
8	C	57%	0.26	0.38	6	0	17	3	3	1	0
9	D	60%	0.43	0.50	2	2	5	18	2	0	1
10	E	70%	0.47	0.63	0	1	1	5	21	2	0
11	A	60%	0.42	0.63	18	5	2	1	3	1	0
12	B	57%	0.15	0.25	4	17	2	4	2	0	1
13	C	50%	0.30	0.38	6	1	15	3	5	0	0
14	D	70%	0.04	0.00	1	1	6	21	1	0	0
15	B/D	83%	0.26	0.38	3	9	2	16	0	0	0
16	A	63%	0.61	0.88	19	4	0	3	4	0	0
17	B	67%	0.46	0.63	1	20	0	2	5	1	1
18	C	70%	0.28	0.38	0	4	21	3	2	0	0
19	D	60%	0.50	0.75	5	1	1	18	3	1	1
20	E	73%	0.07	0.13	3	1	2	1	22	0	1
21	A	60%	0.44	0.50	18	4	2	4	2	0	0
22	B	73%	0.01	0.00	0	22	2	1	4	1	0
23	C	63%	0.25	0.38	1	3	19	2	3	1	1
24	D	53%	0.35	0.63	2	4	3	16	3	1	1
25	E	57%	0.23	0.25	1	3	3	4	17	0	2
26	A	50%	0.45	0.75	15	6	0	5	2	1	1
27	B	47%	0.20	0.25	5	14	0	3	5	3	0
28	C	60%	0.32	0.63	1	2	18	2	4	2	1
29	D	70%	0.35	0.38	1	6	1	21	0	1	0
30	E	67%	0.49	0.75	5	2	2	1	20	0	0
31	A	57%	0.39	0.50	17	5	5	1	0	0	2
32	B	50%	-0.29	-0.38	2	15	2	5	4	2	0
33	C	60%	0.57	0.75	7	2	18	2	1	0	0
34	D	70%	0.54	0.63	6	1	2	21	0	0	0
35	E	57%	0.43	0.63	3	5	1	3	17	1	0
36	A	60%	0.33	0.50	18	3	5	1	3	0	0
37	B	63%	0.37	0.63	1	19	1	3	5	1	0
38	C	60%	0.47	0.63	4	0	18	5	2	1	0
39	D	67%	0.30	0.38	3	3	2	20	0	1	1
40	E	63%	0.52	0.75	2	1	5	3	19	0	0
41	A	60%	0.45	0.63	18	5	2	2	2	0	1
42	B	67%	-0.01	0.25	1	20	1	3	5	0	0
43	C	50%	0.38	0.50	7	4	15	3	1	0	0
44	D	53%	0.44	0.63	1	3	3	16	7	0	0
45	E	63%	0.40	0.63	2	1	4	2	19	0	2
46	A	63%	0.43	0.63	19	3	3	2	1	2	0
47	B	57%	0.31	0.38	1	17	5	3	2	1	1
48	C	47%	0.71	0.88	4	5	14	3	4	0	0
49	D	63%	0.18	0.25	3	1	4	19	3	0	0
50	excl.	—	—	—	—	—	—	—	—	—	—

Answer similarity & scores

Answer similarity

Sheet 001 & Sheet 002 — 30 identical wrong answers; 49 of 49 scored answers identical.

This deliberately matching fictional pair demonstrates the similarity check. Similarity is a prompt to look closer, not proof of copying.

Individual results

Sheet	Score	Percentage	Sheet	Score	Percentage
Sheet 001	17 / 49	34.7%	Sheet 016	30 / 49	61.2%
Sheet 002	17 / 49	34.7%	Sheet 017	33 / 49	67.3%
Sheet 003	26 / 49	53.1%	Sheet 018	31 / 49	63.3%
Sheet 004	20 / 49	40.8%	Sheet 019	33 / 49	67.3%
Sheet 005	20 / 49	40.8%	Sheet 020	35 / 49	71.4%
Sheet 006	19 / 49	38.8%	Sheet 021	30 / 49	61.2%
Sheet 007	23 / 49	46.9%	Sheet 022	32 / 49	65.3%
Sheet 008	19 / 49	38.8%	Sheet 023	37 / 49	75.5%
Sheet 009	20 / 49	40.8%	Sheet 024	30 / 49	61.2%
Sheet 010	25 / 49	51.0%	Sheet 025	43 / 49	87.8%
Sheet 011	27 / 49	55.1%	Sheet 026	34 / 49	69.4%
Sheet 012	22 / 49	44.9%	Sheet 027	38 / 49	77.6%
Sheet 013	26 / 49	53.1%	Sheet 028	42 / 49	85.7%
Sheet 014	27 / 49	55.1%	Sheet 029	41 / 49	83.7%
Sheet 015	31 / 49	63.3%	Sheet 030	42 / 49	85.7%

Companion CSV

The sample CSV contains all 30 students’ responses and scores, the accepted answer key, class summary, score distribution, full item analysis, warnings, and similarity pairs. Sheet labels are anonymous fictional identifiers.

Premium adds analysis pages after the marked answer sheets. This sample contains only the report, plus the explanatory guide below; it is not a scanned class set.

Analysis guide

Definitions and calculations from the Bubble Sheet Grader FAQ · bsgrader.com/faq

Mean / Median / Minimum / Maximum

Mean is the average score. Median is the middle score when scores are ranked; with an even class size it is the average of the two middle scores. Minimum and maximum are the lowest and highest scores. Percentages divide the score by the number of scored questions.

Standard deviation

How spread out scores are around the mean, in points. A small number means the class scored similarly; a large number means scores ranged widely. It is the square root of the average squared distance from the mean (population variance).

Reliability (KR-20)

An estimate of consistency for a test scored right/wrong. With k scored questions and total-score variance V, KR-20 = k / (k − 1) × [1 − Σ p(1 − p) / V], where p is each question's share correct. The app uses item proportions rounded to two decimals. Ratings: Good (0.80+), Acceptable (0.70+), Fair (0.50+), or Low. It is usually interpreted on a 0–1 scale but can be negative for an inconsistent test. No estimate is shown when there is insufficient data or no score variance.

Standard error of measurement

An estimated amount of measurement uncertainty in points, derived from score spread and reliability: SD × √max(0, 1 − KR-20), calculated from unrounded values, so plugging in the rounded figures shown can differ slightly. A score of 17 with SEM 1 can be viewed as roughly 17 ± 1 for one standard error; it is not a guaranteed interval for a student's true ability.

Difficulty (% correct)

The share of graded students who answered correctly. 0.85 means 85% got it right. Despite the name, a higher value means an easier question. Blank and multiple-mark responses count as incorrect on scored questions; excluded questions have no difficulty statistic.

Discrimination (point-biserial)

How well getting a question right tracks the overall score. It correlates each student's right/wrong result with their total score minus that question, so the item cannot inflate its own correlation. Positive means stronger students got it right more often; 0.30+ is strong. Below −0.05 is a red flag; from −0.05 up to 0.15 counts as low. The statistic is unavailable if there is no variation or too little data.

Analysis guide, continued

Upper–lower index (U–L)

Rank students by total score, take the top 27% and bottom 27% (rounded to the nearest whole student), then subtract the lower group's share correct from the upper group's. Values are rounded to two decimals, with halves rounded away from zero. Near zero or negative means the question does not separate strong from weak students. This sample uses 8 students in each group.

Chose A–E, Blank, Multiple

How many students picked each option, left the question blank, or shaded more than one bubble. These counts show whether a wrong option is genuinely tempting (a useful distractor) or rarely chosen. Multiple means more than one student mark, even when the key accepts two alternatives. Each response belongs to exactly one count category.

Questions to review (warnings)

Possible wrong key: the most popular wrong option was chosen at least as often as accepted answers overall, and more upper-group students chose it than an accepted answer. When a possible wrong key is flagged it replaces the discrimination warning, since a wrong key also produces negative discrimination. Negative discrimination: below −0.05, meaning stronger students did worse. Low discrimination: from −0.05 up to 0.15, when fewer than 95% answered correctly. Very hard: under 25% correct. Everyone correct: 100% correct, so the question provides no score separation. Warnings require at least five graded students and need professional review, not automatic re-keying.

Answer similarity

A pairing check for unusually many identical wrong answers, not just matching scores. A pair is flagged when shared wrong letters cover at least four questions and at least 10% of scored questions, and are at least 75% of the smaller wrong-answer total (each sheet's wrong answers, including blanks and multiple marks). Blank and multiple marks do not count as shared wrong letters. The report lists up to 20 pairs with the most shared wrong answers. Similarity is an indicator to investigate, not proof of copying.

Score distribution

Counts of students in 10-percentage-point bands: 0–9%, 10–19%, and so on through 90–100%. It provides a quick view of the class's score range and concentration; the highest band includes 100%.

The report page

With Premium, the analysis is appended to the marked PDF after the marked answer sheets. The same calculated results are shown on screen and included in the CSV. This public sample adds these FAQ explanations so you can interpret the results.