



AP[®] Chemistry: Summary Tables

These tables show year-to-year fluctuation and trends over time in a number of areas.
They include:

- average scores for both the total exam and for the exam sections;
- percentage of students receiving each AP grade; and
- reliability coefficients and standard errors of measurement.

COLLEGE BOARD
AP[®] EXAMINATIONS
FORMS 3XBP

Parts of the Exam

Table 1 provides information about the different parts of the exam, including details about section timing, the maximum possible score on the section, the associated weight, if any, and the maximum possible weighted score. Weights are assigned so that each section of the exam contributes a predetermined percentage to the composite score.

The scoring formula for the multiple-choice section is shown below the table. Each correct response (Right) is scored as 1 point and each incorrect response (Wrong) is scored as $-1/(k-1)$ points, where k is the number of alternatives. Non-responses are scored 0 points. If the overall score on any multiple-choice section is negative, that score is converted to a zero.

Information about any multiple-choice items that were not scored is also shown below the table. In case there are multiple-choice items that are not scored, the associated weights are changed so that the percentage contribution of the section is unchanged.

Sample Sizes

Table 2 gives information about sample sizes across the last five administrations¹. The number in the total group of candidates for whom scores were reported, and the percentage increase from the previous year are shown. This table also shows the number of candidates in the “on-time” group and the test analysis sample. The “on-time” group consists of all cases available as of June 28.

The test analysis sample is based on the “on-time” group, but candidates who did not respond to the first six multiple-choice items as well as candidates who are missing scores on any part of the exam are excluded from the test analysis sample. If the “on-time” group consists of more than 100,000 candidates, the test analysis sample consists of a representative sample of that group.

For the French, German, and Spanish Language exams, the numbers reported for the “on-time” group and test analysis sample are based on candidates who belong to the “standard” group. This group consists of candidates who reported not having any extended exposure to the language of study outside the classroom.

¹ For all tables which present comparative data across the past five administrations, information is shown for fewer years if the exam has been administered for less than five years.

AP Grade Information

Table 3 provides information on grades across the last five administrations. For the reader's convenience, both the percentages at each grade, and percents below each grade are shown. The mean and standard deviation of the grades are also shown.

Grade information is based on the total group of candidates. For the French, German, and Spanish Language exams, grade information is also provided for the "standard" group in *Table 3A*.

Raw Score Information

Table 4 provides information about the raw scores on the multiple choice, free response and composite score scales across the last five administrations. Means and standard deviations are shown, both in raw score scale units, and as percentages of the maximum possible scores. The information in this table is obtained from the score distributions for the "on-time" group. For French, German, and Spanish Language exams, information is based on the standard group of candidates.

Information About Part Scores

Table 5 provides information about part or section scores for the current administration of the exam. The information shown is for unweighted scores. The mean and standard deviation are also shown as percentages of the maximum possible scores.

For exams which allow the candidates choice among a set of free response questions, *Table 5A* shows the percent of the group choosing each question, the mean and standard deviation of the scores for that question, and the deviation of the mean from the pooled mean of all questions in the set. This information is also shown for the scores on the other, mandatory parts of the exam.

The information in these tables is based on the test analysis sample. For the French, German, and Spanish Language exams, information is based on the standard group.

Information About Multiple Choice Section

Table 6 shows information from the Item Analysis of the multiple-choice section of the exam². This information is based on the test analysis sample. For exams with more than one multiple-choice part, information is shown separately for each part. For the French, German, and Spanish Language exams, information is based on the standard group.

For administrations prior to and including 1999, the mean percentage of correct responses (P+) across all items is shown. For administrations in 2000 and 2001, P+ is replaced by the average item score (AIS+). P+ and AIS+ should not be compared directly to each other; incorrect responses and non-responses are treated in the same way in the calculation of P+, but not in the calculation of AIS+.

² Table 6 is skipped for Studio Art as these subjects do not have any multiple choice items.

For each item, the correlation of the scores on the item with the scores on the overall multiple-choice section (or part, for exams with more than one multiple choice part) is computed. The mean and standard deviation of these correlations as well as the number of items where the correlation is less than 0.2 are reported in the table.

The mean number of correct responses (Rights), incorrect responses (Wrongs), and non-responses are reported in the table. Non-responses are classified as omits if the candidate responds to at least one item following the non-response. If the candidate does not respond to any of the following items, then the non-responses are classified as “not reached (NR).”

A large number of omits and NRs may result from the section being speeded, that is, from candidates having insufficient time to complete the section. The table also shows some additional information that is related to speededness.

Correlations

Correlations among scores on different parts of the exam are shown in *Table 7*. The type of correlation coefficient reported is a Pearson correlation. The labels that appear in the table correspond to those in *Table 1*. The weighted score on the multiple-choice section (weighted sum of scores if there is more than one multiple choice part) is labeled SWOP (Sum of Weighted Objective Parts). The sum of weighted scores on the free response section is labeled SWEF (Sum of Weighted Essay Parts).

The information in this table is based on the test analysis sample. For the French, German, and Spanish Language exams, information is based on the standard group.

For exams which allow choice among a set of free response questions, correlations are shown separately for each group of candidates.

Reliability

Table 8 provides information about the reliability and standard error of measurement (SEM) for the multiple choice, free response, and composite scores. Coefficient Alpha is used as the estimate of reliability.

The Pearson correlation between the weighted scores on the multiple choice and free response sections and estimated true-score correlation is shown in the table. The estimated true-score correlation should be interpreted with caution because Coefficient Alpha is an underestimate of reliability, particularly for the free response section.

Also shown is the reliability of classification for the reported AP grades. The methodology is described in Livingston and Lewis (1995)³, and implemented in the

³ Livingston, S. A., and Lewis, C. (1995). Estimating the consistency and accuracy of classification based on test scores. *Journal of Educational Measurement*, 32, 179–197.

computer program RELCLASS-COMP. The version of the program used for the analyses in this report is 4.12.

RELCLASS-COMP consistency is based on an estimated joint distribution of reported grades on the current form of the exam and the grades on an alternate (parallel) form of the exam. RELCLASS-COMP accuracy is based on an estimated joint distribution of reported grades on the current form of the exam and the grades based on an all-forms average (true score). In each case, the proportion of grades with exact agreement is the sum of the entries in the leading diagonal of the contingency table representing the joint distribution; this quantity as well as the reliability of classification at each grade boundary is shown in the table.

Reliability of classification at each grade boundary is estimated by collapsing the joint distribution into a 2x2 table at the grade boundary, and summing the two entries in the leading diagonal.

The information in this table is based on the test analysis sample. For the French, German, and Spanish Language exams, information is based on the standard group.

Selected formulae used to compute statistics in AP test analyses

(1) Coefficient Alpha:

$$\mathbf{a} = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n \mathbf{s}_i^2}{\mathbf{s}_x^2} \right]$$

where n is the number of items, \mathbf{s}_i^2 is the variance of scores on the i -th item, and \mathbf{s}_x^2 is the variance of the total score (sum of scores on the individual items).

(2) Standard Error of Measurement (SEM):

$$\mathbf{s}_e = \mathbf{s}_x \sqrt{1 - \mathbf{a}}$$

where \mathbf{a} is the reliability estimated using (1) above, and \mathbf{s}_x is the standard deviation of the total score.

(3) Reliability of composite scores:

$$\mathbf{a}_c = 1 - \frac{\sum_{j=1}^k (w_j \mathbf{s}_{e_j})^2}{\mathbf{s}_c^2}$$

where k is the number of part scores in the composite, w_j is the weight associated

with the j -th part score, \mathbf{s}_{e_j} is the SEM of the j -th part score, and \mathbf{s}_c^2 is the variance of the composite score.

- (4) True score correlation (correlation corrected for attenuation):

$$r_{t_x, t_y} = \frac{r_{a,b}}{\sqrt{\mathbf{a}_a \mathbf{a}_b}}$$

where $r_{a,b}$ is the correlation between the observed scores, and \mathbf{a}_a and \mathbf{a}_b are the reliabilities of the two scores.

- (5) AIS+ (average item score) for the i -th item:

$$\frac{R_i + \frac{1}{k} O_i}{R_i + O_i + W_i}$$

where k is the number of alternatives, R_i is the number of correct responses, O_i is the number of omits, and W_i is the number of incorrect responses. Candidates with “not reached” scores on the item are excluded from the calculation of AIS+. Note the way in which omits and incorrect responses are treated in the calculation of AIS+ compared to the calculation of candidates’ multiple choice scores. With AIS+, negative values are not possible.

AP[®] CHEMISTRY**3XBP****Table 1 : Parts of the Exam**

Description	Time	Maximum Possible Score	Weight	Maximum Weighted Score	Contribution to Composite
I - Multiple choice	1 hour 30 min	75	.9600	72.0	45.0%
II - Free response	1 hour 30 min			88.0	55.0%
Equilibrium-Solubility (Q1)	40 min (Q1-Q3)	10	1.7600	17.6	11.0%
Choice Problem (Q2 or Q3)		10	1.7600	17.6	11.0%
Chemical Reactions (Q4)		15	.8800	13.2	8.3%
Mixed Topic (Q5)	50 min (Q4-Q8)	10	1.3200	13.2	8.3%
Lab Kinetics (Q6)		10	1.3200	13.2	8.3%
Choice Essay (Q7 or Q8)		8	1.6500	13.2	8.3%
Composite	3 hours			160.0	

Note:

Scoring formula for Section I (multiple choice) : $R-0.250W$

AP[®] CHEMISTRY**3XBP****Table 2 : Sample Sizes**

	3TBP 1997	3UBP 1998	3VBP 1999	3WBP 2000	3XBP 2001
Number in total group	41,874	44,937	48,899	52,786	55,406
% change from previous year	9.0%	7.3%	8.8%	7.9%	5.0%
Number in on-time group	41,296	44,447	48,457	52,547	55,013
Number in test analysis sample	9,130	8,885	9,680	52,490	54,939

Table 3 : AP Grade Information (Total Group)

	3TBP 1997	3UBP 1998	3VBP 1999	3WBP 2000	3XBP 2001
Percent at AP5	16.0	16.3	17.2	16.4	15.8
Percent at AP4	14.5	13.9	14.7	14.4	14.3
Percent at AP3	27.9	28.1	25.3	27.5	27.8
Percent at AP2	22.5	21.3	22.0	21.0	20.8
Percent at AP1	19.0	20.4	20.8	20.7	21.3
Percent below AP5	83.9	83.7	82.8	83.6	84.2
Percent below AP4	69.4	69.8	68.1	69.2	69.9
Percent below AP3	41.5	41.7	42.8	41.7	42.1
Percent below AP2	19.0	20.4	20.8	20.7	21.3
Mean	2.86	2.84	2.85	2.85	2.82
Standard Deviation	1.32	1.34	1.37	1.35	1.34

AP[®] CHEMISTRY

3XBP

Table 4 : Information about raw scores (On-time Group)

	3TBP 1997	3UBP 1998	3VBP 1999	3WBP 2000	3XBP 2001
Range of obtained scores:					
Multiple choice	0 - 72	0 - 72	0 - 72	0 - 72	0 - 72
Free response	0 - 88	0 - 88	0 - 88	0 - 88	0 - 88
Composite	0 - 160	0 - 160	0 - 158	0 - 160	0 - 158
Mean:					
Multiple choice	23.65	28.59	27.58	28.52	29.02
Free response	38.89	34.35	37.24	38.85	35.33
Composite	62.53	62.93	64.82	67.37	64.36
Standard Deviation:					
Multiple choice	14.06	15.90	15.38	16.20	15.68
Free response	19.27	17.62	19.39	19.71	19.25
Composite	31.97	32.36	33.41	34.68	33.69
Mean as % of maximum:					
Multiple choice	33	40	38	40	40
Free response	44	39	42	44	40
Composite	39	39	41	42	40
St. Dev. as % of maximum:					
Multiple choice	20	22	21	23	22
Free response	22	20	22	22	22
Composite	20	20	21	22	21
Skewness:					
Multiple choice	N/A	N/A	N/A	N/A	0.20
Free response	N/A	N/A	N/A	N/A	0.21
Composite	N/A	N/A	N/A	N/A	0.21

AP[®] CHEMISTRY

3XBP

Table 5 : Information about part scores

	Mean	St. Dev.	Mean as % of max	St. Dev. as % of max	Skewness
Equilibrium -Solubility (Q1)	3.92	2.60	39%	26%	0.36
Choice Problem (Q2 or Q3)	3.64	2.57	36%	26%	0.42
Chemical Reactions (Q4)	7.39	4.41	49%	29%	0.13
Mixed Topic (Q5)	3.87	2.66	39%	27%	0.53
Lab Kinetics (Q6)	4.32	2.51	43%	25%	0.30
Choice Essay (Q7 or Q8)	2.87	2.49	36%	31%	0.46

Table 5A : Information about choice questions

Question Selected	Q2	Q3	Q7	Q8
Percent choosing	52.0%	48.0%	51.5%	48.5%
Mean	4.18	3.06	3.67	2.02
St. Dev.	2.70	2.28	2.46	2.24
Deviation from pooled mean*	0.21	-0.23	0.32	-0.34
Section I (multiple choice):				
Mean	33.01	27.29	29.83	30.74
St. Dev.	16.65	15.39	16.51	16.08
Deviation from pooled mean*	0.17	-0.18	-0.03	0.03
Equilibrium-Solubility (Q1):				
Mean	4.38	3.42	4.00	3.83
St. Dev.	2.66	2.44	2.67	2.53
Deviation from pooled mean*	0.18	-0.19	0.03	-0.03
Chemical Reactions (Q4):				
Mean	8.09	6.62	7.51	7.25
St. Dev.	4.48	4.21	4.49	4.33
Deviation from pooled mean*	0.16	-0.17	0.03	-0.03
Mixed Topic (Q5):				
Mean	4.28	3.42	3.86	3.87
St. Dev.	2.76	2.46	2.70	2.61
Deviation from pooled mean*	0.16	-0.17	0.00	0.00
Lab Kinetics (Q6):				
Mean	4.65	3.97	4.27	4.39
St. Dev.	2.59	2.36	2.57	2.44
Deviation from pooled mean*	0.13	-0.14	-0.02	0.03

* in pooled standard deviation units.

AP[®] CHEMISTRY

3XBP

Table 6 : Information about multiple choice section

	3TBP 1997	3UBP 1998	3VBP 1999	3WBP 2000	3XBP 2001
Mean P+ / AIS+	43	49	48	52	53
Observed Delta:					
Mean	13.7	13.1	13.2	12.8	12.7
Standard Deviation	2.0	2.0	1.8	1.6	1.7
Equated Delta:					
Mean	13.1	12.3	12.4	12.2	12.1
Standard Deviation	2.1	1.9	2.1	1.7	1.8
Correlation with criterion:					
Mean	0.47	0.52	0.51	0.46	0.45
Standard Deviation	0.12	0.11	0.11	0.12	0.12
Items where corr < 0.20	0	0	0	2	1
Mean number of:					
Rights	31.93	36.00	35.59	35.87	36.78
Wrongs	29.43	26.99	27.47	26.32	26.19
Omits	12.59	10.14	10.68	10.71	11.18
NR (not reached)	1.05	.91	1.26	1.10	.84
Speededness:					
% completing section	70.5	71.2	41.6	68.4	69.6
% completing 75% of section	99.6	99.4	99.6	99.5	99.5
Items reached by 80% of sample	74	74	74	73	74
NR variance / Score variance	0.04	0.04	0.03	0.04	0.04

AP[®] CHEMISTRY

3XBP

Table 7 : Correlations among part scores

Group choosing Q2 and Q7 (N=16,081)							
	Q1	Q2	Q4	Q5	Q6	Q7	SWOP
Q1	1.000						
Q2	0.658	1.000					
Q4	0.632	0.606	1.000				
Q5	0.648	0.598	0.653	1.000			
Q6	0.607	0.604	0.548	0.585	1.000		
Q7	0.631	0.633	0.592	0.588	0.563	1.000	
SWOP	0.753	0.723	0.689	0.733	0.708	0.686	1.000

Group choosing Q2 and Q8 (N=12,498)							
	Q1	Q2	Q4	Q5	Q6	Q8	SWOP
Q1	1.000						
Q2	0.606	1.000					
Q4	0.586	0.546	1.000				
Q5	0.611	0.556	0.628	1.000			
Q6	0.564	0.551	0.488	0.545	1.000		
Q8	0.555	0.521	0.521	0.574	0.481	1.000	
SWOP	0.730	0.689	0.650	0.716	0.680	0.661	1.000

Group choosing Q3 and Q7 (N=12,233)							
	Q1	Q3	Q4	Q5	Q6	Q7	SWOP
Q1	1.000						
Q3	0.601	1.000					
Q4	0.587	0.521	1.000				
Q5	0.606	0.555	0.613	1.000			
Q6	0.572	0.567	0.497	0.558	1.000		
Q7	0.596	0.534	0.548	0.555	0.537	1.000	
SWOP	0.723	0.683	0.647	0.695	0.676	0.653	1.000

3XBP

Table 7 continued : Correlations among part scores

Group choosing Q3 and Q8 (N=14,127)

	Q1	Q3	Q4	Q5	Q6	Q8	SWOP
Q1	1.000						
Q3	0.567	1.000					
Q4	0.548	0.488	1.000				
Q5	0.587	0.531	0.586	1.000			
Q6	0.538	0.535	0.471	0.537	1.000		
Q8	0.526	0.494	0.488	0.534	0.472	1.000	
SWOP	0.699	0.665	0.628	0.687	0.655	0.637	1.000

	Q1	Q2/Q3	Q4	Q5	Q6	Q7/Q8	SWOP
Q1	1.000						
Q2/Q3	0.629	1.000					
Q4	0.605	0.563	1.000				
Q5	0.628	0.579	0.634	1.000			
Q6	0.583	0.579	0.516	0.569	1.000		
Q7/Q8	0.574	0.547	0.534	0.546	0.492	1.000	
SWOP	0.736	0.704	0.665	0.719	0.690	0.626	1.000

AP[®] CHEMISTRY

3XBP

Table 8 : Reliability Information

	3TBP	3UBP	3VBP	3WBP	3XBP
	1997	1998	1999	2000	2001
Reliability:					
Multiple choice	0.908	0.927	0.923	0.933	0.929
Free response	0.836	0.840	0.862	0.864	0.886
Composite	0.923	0.935	0.937	0.942	0.947
Correlation between MC and FR:					
Observed	0.839	0.868	0.845	0.864	0.859
Estimated True-Score	0.963	0.984	0.947	0.962	0.947
Std. Error of Measurement:					
Multiple choice	4.46	4.42	4.47	4.32	4.18
Free response	7.81	7.08	7.25	7.26	6.49
Composite	8.91	8.28	8.43	8.38	7.72
RELCLASS-COMP consistency:					
Exact agreement	N/A	N/A	N/A	N/A	0.69
AP5/AP4 grade boundary	N/A	N/A	N/A	N/A	0.94
AP4/AP3 grade boundary	N/A	N/A	N/A	N/A	0.91
AP3/AP2 grade boundary	N/A	N/A	N/A	N/A	0.90
AP2/AP1 grade boundary	N/A	N/A	N/A	N/A	0.92
RELCLASS-COMP accuracy:					
Exact agreement	0.73	0.75	0.75	0.76	0.77
AP5/AP4 grade boundary	0.95	0.95	0.95	0.95	0.96
AP4/AP3 grade boundary	0.93	0.93	0.93	0.93	0.94
AP3/AP2 grade boundary	0.92	0.92	0.93	0.93	0.93
AP2/AP1 grade boundary	0.94	0.94	0.94	0.95	0.95