



## **AP<sup>®</sup> Music Theory: 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<sup>®</sup> 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 administrations<sup>1</sup>. 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.

---

<sup>1</sup> 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 exam<sup>2</sup>. 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+.

---

<sup>2</sup> 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 SWEP (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)<sup>3</sup>, and implemented in the

---

<sup>3</sup> 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<sup>®</sup> MUSIC THEORY**

**3XBP**

**Table 1 : Parts of the Exam**

Description	Time	Maximum Possible Score	Weight	Maximum Weighted Score	Contribution to Composite
I - Multiple choice	1 hour 23 min	78		90.0	45.0%
Ia - Aural	48 min	49	1.1538	56.5	28.3%
Ib - Non Aural	35 min	29	1.1538	33.5	16.7%
II - Free response	1 hour 11 min			110.0	55.0%
Melodic Dictation (IIA-1)		9	1.0000	9.0	4.5%
Melodic Dictation (IIA-2)		9	1.0000	9.0	4.5%
Harmonic Dictation (IIA-3)	63 min (IIA)	24	.3750	9.0	4.5%
Harmonic Dictation (IIA-4)		27	.3333	9.0	4.5%
Partwriting (IIA-5)		29	.9310	27.0	13.5%
Composition (IIA-6)		10	2.7000	27.0	13.5%
Sight-Singing (IIB-1)	8 min (IIB)	9	1.1111	10.0	5.0%
Sight-Singing (IIB-2)		9	1.1111	10.0	5.0%
Composite	2 hours 34 min	200		200.0	

Note:

Scoring formula for Section I (multiple choice) :  $R-0.333W$

**AP<sup>®</sup> MUSIC THEORY****3XBP****Table 2 : Sample Sizes**

	3TBP 1997	3UBP 1998	3VBP 1999	3WBP 2000	3XBP 2001
Number in total group	3,369	4,084	4,820	5,304	6,135
% change from previous year	20.1%	21.2%	18.0%	10.0%	10.0%
Number in on-time group	3,326	4,038	4,764	5,181	5,970
Number in test analysis sample	3,320	4,025	4,760	5,174	5,964

**Table 3 : AP Grade Information (Total Group)**

	3TBP 1997	3UBP 1998	3VBP 1999	3WBP 2000	3XBP 2001
Percent at AP5	19.3	19.5	16.7	20.3	19.8
Percent at AP4	19.4	18.8	28.7	20.8	23.9
Percent at AP3	29.0	28.8	26.5	27.0	29.7
Percent at AP2	23.5	23.4	21.4	23.2	19.1
Percent at AP1	8.8	9.5	6.7	8.7	7.5
Percent below AP5	80.7	80.5	83.3	79.7	80.2
Percent below AP4	61.3	61.7	54.6	58.9	56.3
Percent below AP3	32.3	32.9	28.1	31.9	26.6
Percent below AP2	8.8	9.5	6.7	8.7	7.5
Mean	3.17	3.15	3.27	3.21	3.30
Standard Deviation	1.23	1.25	1.17	1.25	1.20

AP<sup>®</sup> MUSIC THEORY

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 - 90	0 - 89	0 - 88	0 - 88	0 - 88
Free response	0 - 110	0 - 110	2 - 110	0 - 110	0 - 110
Composite	0 - 200	3 - 197	2 - 197	1 - 198	0 - 198
Mean:					
Multiple choice	41.36	38.03	42.02	42.49	45.54
Free response	55.83	52.55	55.83	56.46	55.39
Composite	97.19	90.57	97.85	98.93	100.91
Standard Deviation:					
Multiple choice	16.30	15.74	16.27	16.84	16.52
Free response	20.20	20.45	21.58	21.55	21.67
Composite	34.61	34.27	36.04	36.39	36.32
Mean as % of maximum:					
Multiple choice	46	42	47	47	51
Free response	51	48	51	51	50
Composite	49	45	49	49	50
St. Dev. as % of maximum:					
Multiple choice	18	17	18	19	18
Free response	18	19	20	20	20
Composite	17	17	18	18	18
Skewness:					
Multiple choice	N/A	N/A	N/A	N/A	-0.08
Free response	N/A	N/A	N/A	N/A	0.05
Composite	N/A	N/A	N/A	N/A	0.08

AP<sup>®</sup> MUSIC THEORY

3XBP

Table 5 : Information about part scores

	Mean	St. Dev.	Mean as % of max	St. Dev. as % of max	Skewness
Ia - Aural	23.51	9.47	48%	19%	0.09
Ib - Non Aural	16.00	6.37	55%	22%	-0.28
Melodic Dictation (IIA-1)	4.43	2.69	49%	30%	0.20
Melodic Dictation (IIA-2)	3.73	2.41	41%	27%	0.63
Harmonic Dictation (IIA-3)	13.75	6.72	57%	28%	-0.07
Harmonic Dictation (IIA-4)	13.33	6.60	49%	24%	-0.03
Partwriting (IIA-5)	13.75	6.98	47%	24%	0.15
Composition (IIA-6)	5.06	2.71	51%	27%	-0.20
Sight-Singing (IIB-1)	5.54	2.85	62%	32%	-0.26
Sight-Singing (IIB-2)	4.53	2.48	50%	28%	0.35

AP<sup>®</sup> MUSIC THEORY

3XBP

Table 6 : Information about multiple choice section

	3TBP 1997		3UBP 1998		3VBP 1999		3WBP 2000		3XBP 2001	
	<u>Ia</u>	<u>Ib</u>	<u>Ia</u>	<u>Ib</u>	<u>Ia</u>	<u>Ib</u>	<u>Ia</u>	<u>Ib</u>	<u>Ia</u>	<u>Ib</u>
Mean P+ / AIS+	54	65	53	59	56	64	60	63	61	66
Observed Delta:										
Mean	12.5	11.2	12.6	11.8	12.3	11.3	11.9	11.6	11.8	11.0
Standard Deviation	1.8	2.4	2.1	2.6	2.2	2.2	2.1	1.5	1.9	2.4
Equated Delta:										
Mean	12.5	11.2	12.6	11.8	12.8	11.8	11.9	12.0	12.3	11.4
Standard Deviation	1.8	2.4	2.2	2.7	2.3	2.3	2.3	1.9	2.1	2.7
Correlation with criterion:										
Mean	0.41	0.52	0.40	0.51	0.41	0.53	0.41	0.48	0.41	0.49
Standard Deviation	0.10	0.08	0.11	0.08	0.09	0.11	0.11	0.09	0.10	0.12
Items where corr < 0.20	2	0	1	0	1	0	2	0	1	1
Mean number of:										
Rights	25.31	19.96	27.72	17.06	28.58	17.33	28.55	17.48	29.23	18.88
Wrongs	19.32	8.83	21.61	9.87	19.77	8.68	17.92	9.66	17.18	8.68
Omits	2.35	1.90	2.67	1.85	2.66	.92	2.53	1.62	2.59	1.34
NR (not reached)	.00	.30	.00	.22	.00	.07	.00	.23	.00	.10
Speededness:										
% completing section	-	90.0	-	93.2	-	96.4	-	91.3	-	96.7
% completing 75% of section	-	99.7	-	99.4	-	99.9	-	99.7	-	99.8
Items reached by 80% of sample	-	31	-	29	-	27	-	29	-	29
NR variance / Score variance	-	0.03	-	0.03	-	0.01	-	0.03	-	0.02

AP<sup>®</sup> MUSIC THEORY

3XBP

Table 7 : Correlations among part scores

	IIA-1	IIA-2	IIA-3	IIA-4	IIA-5	IIA-6	IIB-1	IIB-2	SWOP
IIA-1	1.000								
IIA-2	0.718	1.000							
IIA-3	0.679	0.657	1.000						
IIA-4	0.646	0.627	0.707	1.000					
IIA-5	0.375	0.396	0.415	0.374	1.000				
IIA-6	0.380	0.405	0.425	0.377	0.504	1.000			
IIB-1	0.607	0.556	0.589	0.550	0.275	0.286	1.000		
IIB-2	0.645	0.624	0.584	0.571	0.279	0.314	0.647	1.000	
SWOP	0.701	0.706	0.665	0.622	0.559	0.512	0.563	0.608	1.000

	Aural	Non Aural	SWEP
Aural	1.000		
Non Aural	0.602	1.000	
SWEP	0.741	0.703	1.000

**AP<sup>®</sup> MUSIC THEORY**

**3XBP**

**Table 8 : Reliability Information**

	3TBP	3UBP	3VBP	3WBP	3XBP
	1997	1998	1999	2000	2001
<b>Reliability:</b>					
Multiple choice	0.872	0.869	0.874	0.885	0.882
Ia - Aural	0.797	0.813	0.814	0.819	0.824
Ib - Non Aural	0.802	0.777	0.787	0.802	0.795
Free response	0.858	0.859	0.870	0.881	0.869
Composite	0.923	0.922	0.928	0.934	0.929
<b>Correlation between MC and FR:</b>					
Observed	0.795	0.791	0.809	0.795	0.807
Estimated True-Score	0.919	0.916	0.928	0.900	0.922
<b>Std. Error of Measurement:</b>					
Multiple choice	5.05	5.12	5.00	4.95	5.66
Ia - Aural	4.08	4.19	4.12	3.92	4.58
Ib - Non Aural	2.97	2.94	2.83	3.02	3.33
Free response	7.62	7.67	7.76	7.44	7.82
Composite	9.59	9.55	9.67	9.38	9.65
<b>RELCLASS-COMP consistency:</b>					
Exact agreement	N/A	N/A	N/A	N/A	0.66
AP5/AP4 grade boundary	N/A	N/A	N/A	N/A	0.92
AP4/AP3 grade boundary	N/A	N/A	N/A	N/A	0.88
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	0.95
<b>RELCLASS-COMP accuracy:</b>					
Exact agreement	0.74	0.73	0.76	0.75	0.75
AP5/AP4 grade boundary	0.94	0.94	0.95	0.94	0.94
AP4/AP3 grade boundary	0.92	0.91	0.92	0.92	0.92
AP3/AP2 grade boundary	0.92	0.92	0.93	0.93	0.93
AP2/AP1 grade boundary	0.96	0.95	0.97	0.97	0.97