ED 166 237 TM 008 226

AUTHOR Echternacht, Gary

TITLE Alternate Methods of Equating GRE Advanced Tests. Project Report PR 71-17 (October 1971). GRE Board Professional Report GREB No. 69-2P.

INSTITUTION Educational Testing Service, Princeton, N.J.

REPORT NO GREB-69-2P

PUB DATE Jun 74

NOTE 90p.

AVAILABLE FROM Graduate Record Examinations, Educational Testing Service, Princeton, New Jersey 08541 (free while supplies last)

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.

DESCRIPTORS Achievement Tests; Aptitude Tests; *College Entrance Examinations; *Comparative Statistics; *Equated Scores; *Goodness of Fit; Graduate Study; Higher Education

IDENTIFIERS *Graduate Record Examinations

ABSTRACT

When two different forms of a particular test are given to different groups of candidates, it is often necessary to make the test results for the two tests or groups as comparable as possible; the statistical process used for this purpose is termed equating. Two different methods of equating Graduate Record Examinations (GRE) Advanced Tests were compared. One method used data from a group of items that were common to both tests, while the other method used data from the GRE verbal aptitude test and quantitative aptitude test, also taken by both groups of candidates. The results of 158 equatings for the 17 GRE Advanced Tests were tabulated and presented graphically. Out of the 17 different test series, nine different series had equating differences at one end or the other of the score distribution equal to about one half a standard deviation, and about three percent of the 158 equatings had differences of over one standard deviation. (Author/CTM)

GRE

ALTERNATE METHODS OF EQUATING GRE ADVANCED TESTS

Gary Echternacht

GRE Board Professional Report GREB No. 69-2P

Project Report PR 71-17
(October 1971)

June 1974

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

NG SERVICE, PRINCETON, NEW JERSEY □ BERKELEY, CALIFORNIA □ EVANST

ALTERNATE METHODS OF EQUATING GRE ADVANCED TESTS

Gary Echternacht

GRE Board Professional Report GREB No. 69-2P

Project Report PR 71-17
(October 1971)

June 1974

ALTERNATE METHODS OF EQUATING GRE ADVANCED TESTS[1]

When different forms of a particular test are given either concurrently or at different administrations, it is often required that the test results be made comparable. For instance, at ETS, scores from the GRE Advanced Biology Test used in the October 1967 administration were made comparable to the form used in the January 1965 administration. Thus, the integrity of the test can be protected by using different forms while, at the same time, test scores are comparable and on the same score scale though they come from different forms of the same test. This process of making tests comparable is termed equating and is carried out whenever a new test form is introduced in a testing program.

One basic requirement for the type of equating traditionally used at ETS, appropriately termed common-item equating, is the existence of a number of test items common to both the new test form and the old test form to which the new test is equated. These common items serve as a basis for estimating how each group would have performed on the test taken by the other which, in turn, is used to convert the scores on the new test form to the score scale used by the old test form. Although the number of common items necessary for effective equating is substantial, it is usually rather small when compared to the total test length.

In the spring of 1968, most of the examination committees for the GRE Advanced Tests expressed an interest in deriving one or more subscores from the various tests for which they were responsible. The feasibility of providing such subscores was considered, and a number of questions were raised, among

[1] This study has had a rather dynamic history having been conceived by Gary Lutz, with the equatings being supervised by Susan Ford. The author inherited the project at the time of reorganization within the company and did no work on the project other than write this report. He is solely responsible for its contents.

those being the equating of subscores. If the traditional common-item equating were to be used, the number of common items required for equating subscores would be so large that proportionally few new items would result in the test-form. As an alternative to common-item equating, equating through the Verbal and Quantitative scores from the Aptitude Test was suggested.

Statement of the Problem

In order to study some of the difficulties in equating subscores, using Verbal and Quantitative test scores, a study was undertaken to answer the question: How does equating the total score of the GRE Advanced Tests using the Verbal and Quantitative test score from the GRE Aptitude Test compare with the traditional common-item equating for these same tests; more specifically are there practical differences between the two equating methods from the standpoint of reported scores? Is the relationship between the two equating methods constant for all Advanced Tests or is Verbal and Quantitative equating more suitable for some tests than others? Are there differences between the two methods over various administration months? Are there differences in equatings across various educational levels?

It was hypothesized that equating through the Verbal and Quantitative scores would be similar to common-item equating especially when the correlation between the Aptitude Test scores and an Advanced Test score was high. For example, the Verbal and Quantitative equating should prove approximately the same as common-item equating for the Advanced Tests in Economics, Sociology, Philosophy, and Biology as these tests correlate highest with the Verbal and Quantitative scores, while, on the other hand, Spanish, French, and Physics Advanced Tests correlate lowest with the Aptitude Test scores; thus, less similarity was hypothesized. Further, when a test was equated to itself, using Verbal and Quantitative equating, the parameters should be approximately one and zero.[2] The degree to which this is not true reflects the error in the equating parameters.

One problem that occurs when using Verbal and Quantitative Tests scores for equating is that of different levels of candidate preparation. As an example, consider two examinees, one seeking admission to graduate school for the first time, the other having completed a Masters program seeking entry into a doctoral program. These two candidates are likely to score very differently on an Advanced Test although their Aptitude Test scores are the same. This fact lowers the correlation between the Advanced Test and the Aptitude Test scores weakening the strength of the equating.

The Sample

All candidates who took one of 17 GRE Advanced Tests between October 1967 and September 1968 inclusive and who were registered as regular national candidates, candidates for special administration, National Science Foundation candidates, or Oak Ridge Institute of Nuclear Science candidates were selected. A further constraint on the sample was that each candidate for further study had Aptitude Test scores earned no more than three months prior to the Advanced Test score.

Multiple scores for either the Advanced Tests or the Aptitude Tests were treated as follows: In the event of multiple Aptitude Test scores the Aptitude

[2] The equating parameters are of the form Y = A + BX, where Y is the old form scale and X the new form scale. When we say the parameters should be one and zero when a test is equated to itself, that is A = 0; B = 1, we mean Y = X, the two score scales should be the same.

Test score nearest the first Advanced Test score was taken. Multiple Advanced Test scores could not be identified since Advanced Test scores were sampled rather than the candidates themselves.

Since both old and new equating forms had to appear during the period under study, only candidates who took these forms were selected. This action resulted in some candidates taking forms in Economics, Political Science, and Spanish not being selected. The total obtained sample size was 85,111 for 17 Advanced Tests. The Advanced Test with the highest volume was Education with 13,516 candidates selected while the Advanced Test in Geology recorded the lowest volume with 961 candidates. Between four and six test dates were considered for each Advanced Test.

Methodology

The genealogical charts, Appendix 1, for the Advanced Tests were used to determine the forms to be equated using Verbal and Quantitative equating from the totality of test forms given between October and September of the test year. The rule was to duplicate any past common-item equating with Verbal and Quantitative Aptitude score equating. Thus, for example, in economics the form was equated to forms through Verbal and Quantitative Aptitude scores since the traditional common-item equating was accomplished by equating these same forms in 1966.[3]

[3] When one test form is equated to two other test forms through common items, the results are two equating lines of the form $A_1 + B_1X$ and $A_2 + B_2X$ for converting raw scores to scaled scores. To obtain one operational conversion line, the angle between these two conversion lines is bisected and that bisector becomes the conversion line for score reporting.

It should be emphasized here that the sample used for Verbal and Quantitative equating was the sample described previously, while the common-item equatings used as comparisons were the original equatings carried out at various times in the past. Thus, the comparisons of common-item equating with Verbal and Quantitative equating were valid only to the extent of the stability of common-item equating from sample to sample. In order to investigate the stability question for Verbal and Quantitative equating to some extent, whenever a particular form was given more than once during the year under study, those tests were equated to each other.

The procedure for equating using Verbal and Quantitative Aptitude scores is described completely in Appendix [illegible] and is similar to the traditional common-item approach. Generally speaking, though, the process of Verbal and Quantitative equating goes as follows. For both the new form and the old form to which the new form is being equated, conceptualize two regression planes for predicting Advanced Test scores from Verbal and Quantitative Aptitude Test scores, one for the group of examinees taking the new form only, the other for the group taking the new and old forms. We assume these regression planes to be identical in their intercepts, slopes, and errors of estimate. From these assumptions, equations for estimating the mean score and variance on the new form for the total group (new form examinees and old form examinees) are developed. Similarly for the old form, this same procedure is carried out and estimates for the total group are obtained for the old form mean and variance. These two distributions are standardized and set equal to each other after which the new form raw scores are given as a function of the old form raw scores and the equating is completed for raw scores. These equated raw scores are then converted to scaled scores using the old form scaled conversion parameters and the equating is complete.

In order to make comparisons of the Verbal and Quantitative score equating and common-item equating, the equating lines for both methods were graphed for obtaining scaled scores from raw scores. There were separate graphs for each Advanced Test and each particular equating using common old forms within each Advanced Test. The Advanced Tests were then classified into one of three categories depending on the difference at the extreme raw scores between the two methods of equating under study. Those were

Class I. No extreme differences of greater than 50 score points at either extreme

Class II. A difference of more than 50 points at only one extreme

Class III. A difference of more than 50 score points at both extremes.

It was assumed Advanced Tests falling into Class I would be most amenable to equating through Verbal and Quantitative Aptitude Test scores while the other test would be less favorable for that method. For Advanced Tests falling in Class III justifying the use of Verbal and Quantitative equatings would be particularly difficult.

The sample was further partitioned by educational level for each Advanced Test. The educational levels were indicated by every candidate at the time he took the test and are: not now in college, sophomore, junior, senior, first and second year graduate students. Equatings using Verbal and Quantitative Test scores were to be completed for every Advanced Test and every educational level.

Results

A total of 128 equatings were accomplished for 17 different Advanced Tests. The equations for converting raw scores on the new forms to the scaled scores were tabulated and graphed, the graphs appearing in Appendix 3. For each conversion equation obtained, using Verbal and Quantitative equating, four scores were obtained, those being the scaled score when a candidate answers no items correctly, i.e., raw score zero; the scaled score when the candidate answers every item correctly; the scaled score corresponding to the lowest raw score found in the equating sample; and the scaled score corresponding to the highest raw score found in the equating sample. These last two scores were included in an attempt to make the comparisons more valid in a "practical" sense, for example, no one obtains the highest theoretical score for most advanced tests; therefore, the obtained extreme score might provide a better location for obtaining greater insight as to the practical differences between the methods.

Equivalent scale scores were obtained using the common-item conversions for the same raw scores. Values obtained from the Verbal and Quantitative conversions were then subtracted from the values obtained from the common-item conversions and tabulated in Table 1. The differences obtained at zero raw score and at the maximum raw score were termed possible score differences, while the remaining two differences were for observed scores. The subscripts represent the number of the administration month. Using the classification system previously described on the possible scaled score differences, five Advanced Tests fall into Class I, having differences less than 50 scaled score points. The Tests classified in Class I were Education, History, Literature, Political Science, and Spanish. Using the classification scheme on the obtained scale score differences, Biology, French, and Psychology Advanced Tests join the previously mentioned in the first classification.

Of the Advanced Tests classified in Class II, only one Advanced Test displayed differences of larger than 50 scaled score points at the lower end of both the possible and observed scaled scores, that test being music. The difference in this case would be that, were the Verbal and Quantitative equating used, examinees would obtain higher scaled scores at the lower end of the score range than they would had common-item equating been used.

The remaining tests falling in Class II on the possible scaled score differences were Biology, Chemistry, Economics, Engineering, Mathematics, and Psychology. These tests all were characterized by differences of more than 50 scaled score points at the top end of the possible scaled score range.

Advanced Tests in French, Geology, Philosophy, Physics, and Sociology were all classified in Class III on the possible scaled score criterion with only French changing classification on the observed scaled score criterion. A complete classification for both observed and possible scaled score differences appears in Table [illegible].

An interesting event did occur when the Verbal and Quantitative equatings were compared with the common-item equatings in a way other than measuring the endpoint differences of each old form. Most of the common-item equatings involved two old forms as the genealogical charts indicate. If the Verbal and Quantitative equating through two old forms is performed as for common-item equating, that is, bisecting the two obtained equating lines and using the bisector for score reporting, the results change a little as demonstrated in Appendix 5. The score differences at the extremes for Advanced Tests in Biology, Engineering Form Q, French, Geology, Music, Sociology Form Q, and Psychology are each less than 50 points. It is also noteworthy that when Advanced Tests were equated to only one form, poor agreement between equating methods was found.

When a test form was used more than once during the testing year under study, these tests were equated to themselves. Differences between the Verbal and Quantitative equatings and the common-item equatings were calculated and tabulated. These differences provided a rough estimate of how Verbal and Quantitative equatings varied from one equating sample to another. Unfortunately, there are no comparable figures available for common-item equating. The results appear in Table 5.

The results of these calculated differences were mixed. In considering the differences over all Advanced Tests, the process of Verbal and Quantitative equating seems to be unstable as 6 of the 23 equatings resulted in score differences of 50 points or more roughly amounting to about 25 per cent of the equatings. On the other hand, when the equatings were taken by individual Advanced Tests, the number of equatings performed was insufficient for drawing any meaningful conclusions.

For those same Verbal and Quantitative equatings correlations, both first order and multiple correlations were calculated for both the group making up the old and new form equating samples. These correlations were between the form and Verbal Aptitude, the form and Quantitative Aptitude, Verbal and Quantitative Aptitude and the multiple correlation of the form with the two aptitude test scores. These correlations tend to remain stable from old to new form with the exception of the correlation between the form and Verbal in the cases of Chemistry and Mathematics, between Verbal and Quantitative for Spanish, and the multiple correlations for Mathematics and Spanish. These results appear in Table 4.

The sample was partitioned by educational level for each Advanced Test. Counts for each educational level of every Advanced Test were obtained and based on these counts and cost factors; no equatings were performed by educational level. The counts showed that most everyone who took Advanced Tests were seniors and that equatings for the other educational levels were prohibitive based on the small sample numbers.

Discussion and Comments

The question now arises of whether the study accomplished the objective it set up. Clearly, some practical differences were found between common-item and Verbal and Quantitative equating methods in terms of the 50-point classification scheme. One difficulty in interpreting these differences comes about when the samples used for equating are considered. Since different samples were used for each equating, one could logically suspect these differences. The question of comparing the two types of equating lines using identical samples cannot be answered. Common-item equatings corresponding to Verbal and Quantitative equatings could have been performed using the same samples had there been funds for rescoring all answer sheets and 158 additional equatings.

Another question arising in the interpretation of the equating line differences was the significance of the differences obtained. Fifty points was the criterion for significance in this study but was that too much or was that enough? No probability statements can be made concerning statistical significance, and one is forced to use "careful human judgment." Since nothing is known of how sampling differences affect common-item equating and only very limited evidence is available for Verbal and Quantitative equating, no statistical test can be made.

The differences obtained were assessed at the endpoints of the possible score ranges. One might question the need for difference to be calculated here. For example, which end of the score scale is most damaged by a lack of agreement between equating methods? It might be that the need to differentiate among candidates scoring at the highest end of the scale is not necessary thus allowing a relaxation of the 50-point score difference at the high end. Also, one might reason, no one scores at the highest possible score anyway and no one cares whether that score is 990 or 1050 in most selection or diagnostic cases. Therefore, one might question using the possible endpoints as difference criteria and suggest some other less conservative points for assessing practical differences.

Was the relationship between the two equating methods constant for all tests? This we conclude was not the case.. Had the relationship been constant, we would have expected all the Advanced Tests to fall in the same classification. Also, a look through Appendix 5 will illustrate the variability of the Verbal and Quantitative equating lines with respect to the common-item equating line. Clearly, the Verbal and Quantitative equating is more suitable for those Advanced Tests falling in Class I than those falling in Class III with respect to agreement with common-item equating.

The last two questions, differences across various administrations and educational levels, were not answered at all. In order to answer the first question, all forms equated through Verbal and Quantitative scores would have had to be equated using common items. The second question could not be answered due to the relatively small sample sizes obtained for the various educational levels.

The main difficulty this study encountered involved the lack of knowledge of the properties of the common-item equating method. For example, consider the comparison of the operational common-item equating line with a Verbal and Quantitative equating line. Where do we want to evaluate their differences? What first blocks our progress is our not knowing how the common-item equating line varies from sample to sample. Comparisons between the two methods must be considered in light of the sampling variations of each method. The problem of sampling variation cannot be easily solved mathematically as the estimates of the slope and intercept of the equating line involves the ratio of two other estimates. The answer could be found in computer simulation of equatings. If many equatings were simulated under various conditions on the means, variances, and correlations between the anchor and the test, estimates of the equating line variation can be obtained and confidence bands drawn and comparisons made more easily.

Another area of concern should be that of the robustness of the equating procedure against violations in the three basic assumptions. By assessing the degree to which violations in the assumptions affect the equating outcome, the total variation in the equating procedure can be partitioned into two parts, one due to the lack of compliance with the assumptions, the other due to sampling variation. In practice one can do nothing about the second component, but one can select samples for equating where the assumptions are most likely to hold.

It is recommended that studies be undertaken to estimate the variability of the common-item equating line and its robustness against various violations in assumptions. Having accomplished that task, investigations of other methods of equating could be undertaken with meaningful comparisons arising.

Table 1

Common-Item Scaled Score Minus Verbal and Quantitative Scaled Scores at the Extreme Ends of the Scale

Test and Equating Forms	Possible Score Differences		Observed Score Differences	
	Hghst. Poss. Scaled Scores	Lwst. Poss. Scaled Scores	Hghst. Poss. Scaled Scores	Lwst. Poss. Scaled Scores
Biology				
P_{10} -- N_2	24	- 5	18	- 1
P_{10} -- N_4	41	-12	30	- 7
O_1 -- M_{12}	50	- 6	43	- 6
O_1 -- M_7	34	19	33	19
O_1 -- N_2	54	13	50	13
Chemistry				
P_{10} -- N_{12}	46	- 2	36	- 2
P_{10} -- O_1	8	- 1	4	- 1
O_1 -- L_2	125	10	92	12
O_1 -- L_7	37	5	40	15
P_4 -- N_{12}	- 80	4	-55	4
P_4 -- O_1	-126	11	-85	11
N_{12} -- L_7	- 3	22	2	5
N_{12} -- L_2	- 8	8	69	11
Economics				
P_{10} -- N_{12}	20	8	21	7
P_4 -- N_{12}	- 4	20	1	18
O_1 -- N_{12}	54	-23	8	-23
O_1 -- M_7	79	-28	72	-28

Table 1 Cont'd.

Test and Equating Forms	Possible Score Differences Hghst. Poss. Scaled Scores	Possible Score Differences Lwst. Poss. Scaled Scores	Observed Score Differences Hghst. Poss. Scaled Scores	Observed Score Differences Lwst. Poss. Scaled Scores
Education				
P_{12} -- M_{10}	- 21	20	-15	9
P_{12} -- L_{7}	1	15	0	6
O_{1} -- M_{10}	- 3	17	- 1	14
Q_{4} -- O_{1}	- 28	- 9	-21	-10
Q_{4} -- N_{2}	- 7	4	- 2	2
Engineering				
P_{10} -- N_{12}	- 78	20	-50	20
P_{10} -- M_{2}	22	10	21	10
Q_{1} -- P_{10}	- 8	14	0	10
Q_{1} -- O_{4}	81	9	61	5
O_{4} -- M_{2}	- 66	20	-35	22
Q_{7} -- O_{4}	29	- 9	18	-13
Q_{7} -- P_{10}	- 63	- 7	-44	-11
French				
P_{1} -- M_{10}	- 57	27	-37	21
P_{1} -- M_{2}	- 11	- 4	- 6	- 4
P_{1} -- N_{4}	- 54	61	-30	50
P_{1} -- N_{12}	- 12	3	- 9	1
P_{7} -- M_{10}	- 16	-22	-13	19
P_{7} -- M_{2}	27	-42	13	-33
P_{7} -- N_{4}	- 12	20	- 5	15
P_{7} -- N_{12}	24	-37	12	-31

Table 1 Cont'd

Test and Equating Forms	Possible Score Differences		Observed Score Differences	
	Hghst. Poss. Scaled Scores	Lwst. Poss. Scaled Scores	Hghst. Poss. Scaled Scores	Lwst. Poss. Scaled Scores
Geology				
P_{12} -- M_1	- 24	-12	-19	-11
P_{12} -- K_2	3	3	2	3
P_7 -- M_1	- 67	-53	-62	-53
P_7 -- K_2	29	-35	5	-35
History				
P_1 -- M_{12}	- 7	- 3	- 7	- 1
P_1 -- M_{10}	- 26	25	-14	26
P_1 -- M_2	30	10	24	15
Q_4 -- P_1	- 26	- 1	-17	0
Q_4 -- P_7	- 30	15	-14	16
Q_4 -- M_{12}	- 37	0	-23	1
P_7 -- M_{12}	- 4	-18	- 7	-15
P_7 -- M_{10}	- 21	10	-15	13
P_7 -- M_2	32	- 5	24	- 2
Literature				
O_{12} -- L_{10}	4	10	1	13
O_{12} -- L_2	11	18	7	22
P_1 -- O_{12}	- 43	9	-37	7
P_1 -- O_4	- 34	32	-26	30
O_4 -- L_{10}	- 7	-10	-10	- 6
O_4 -- L_2	- 2	- 2	- 5	3
P_7 -- O_{12}	- 45	- 3	-37	- 3
P_7 -- O_4	- 33	-22	-23	-22

Table 1 Cont'd

Test and Equating Forms	Possible Score Differences		Observed Score Differences	
	Hghst. Poss. Scaled Scores	Lwst. Poss. Scaled Scores	Hghst. Poss. Scaled Scores	Lwst. Poss. Scaled Scores
Mathematics				
P_{10} -- N_2	160	-16	157	-13
O_{12} -- N_2	136	20	138	21
Q_1 -- P_{10}	- 45	5	- 47	3
Q_1 -- P_7	161	- 3	155	- 5
P_7 -- N_2	- 22	- 5	- 28	- 2
Music				
O_{12} -- M_{10}	- 11	- 4	- 12	- 6
O_{12} -- M_1	15	-28	10	-26
O_{12} -- M_7	1	16	1	13
O_2 -- M_{10}	- 27	0	- 26	- 3
O_2 -- M_1	- 1	-26	- 4	-26
O_2 -- M_7	- 15	18	- 14	15
Q_4 -- M_{10}	- 31	-37	- 31	-37
Q_4 -- M_1	- 7	-57	- 13	-57
Q_4 -- M_7	- 20	-18	- 12	-18
Philosophy				
P_{12} -- J_{10}	4	30	7	30
P_{12} -- M_1	78	- 1	62	5
P_{12} -- J_2	143	12	24	20
P_{12} -- M_4	42	40	42	40
M_1 -- J_{10}	- 83	41	- 52	36
M_1 -- J_2	60	10	51	17
M_4 -- J_{10}	- 37	- 8	- 30	- 3

Table 1 Cont'd

Test and Equating Forms	Possible Score Differences		Observed Score Differences	
	Hghst. Poss. Scaled Scores	Lwst. Poss. Scaled Scores	Hghst. Poss. Scaled Scores	Lwst. Poss. Scaled Scores
Philosophy Cont'd				
P_7 -- J_{10}	-146	88	- 72	64
P_7 -- M_1	- 56	55	- 20	47
P_7 -- J_2	37	64	39	63
P_7 -- M_4	- 91	105	- 20	90
Physics				
P_{10} -- O_{12}	13	40	16	42
P_{10} -- N_4	126	- 18	116	-16
O_{12} -- K_7	131	10	122	14
Q_1 -- P_{10}	- 17	- 32	- 21	-29
Q_1 -- O_{12}	- 7	10	- 8	13
L_2 -- K_7	-105	56	- 60	55
Political Science				
P_1 -- N_{10}	- 29	44	- 17	43
P_1 -- N_2	- 11	32	15	33
P_7 -- N_{10}	- 44	35	- 24	32
P_7 -- N_2	1	23	7	23
Psychology				
O_1 - L_{10}	- 9	26	- 1	23
O_1 -- K_{12}	- 30	34	- 15	33
O_1 -- L_2	40	13	36	12
Q_4 -- L_{10}	0	- 16	- 3	-13
Q_4 -- O_1	10	- 39	- 1	-36
Q_4 -- L_2	55	- 25	37	-22

Table 1 Cont'd

Test and Equating Forms	Possible Score Differences		Observed Score Differences	
	Hghst. Poss. Scaled Scores	Lwst. Poss. Scaled Scores	Hghst. Poss. Scaled Scores	Lwst. Poss. Scaled Scores
Psychology Cont'd				
Q_4 -- O_7	-22	-3	-18	0
O_7 -- L_{10}	16	-13	12	-14
O_7 -- K_{12}	-2	-6	-3	-7
O_7 -- L_2	66	-23	50	-24
Sociology				
M_{12} -- J_{10}	-119	41	90	38
M_{12} -- J_4	-128	68	-81	63
O_1 -- J_{10}	-43	11	-33	-2
O_1 -- M_{12}	67	-29	45	-43
O_1 -- M_2	90	-37	61	-50
O_1 -- J_4	-51	37	-33	24
M_2 -- J_{10}	-148	82	-93	42
M_2 -- J_4	-59	80	-93	68
O_7 -- J_{10}	-37	-29	-37	-40
O_7 -- M_{12}	73	-63	30	-76
O_7 -- M_2	94	-70	43	-87
O_7 -- J_4	-40	-7	-32	-20
Spanish				
P_1 -- N_{10}	-27	-14	-27	-15
P_1 -- M_{12}	-27	3	-24	0
P_1 -- M_2	-2	21	0	19
P_1 -- N_7	-27	-7	-26	-11

Table 1 Cont'd

Test and Equating Forms	Possible Score Differences		Observed Score Differences	
	Hghst. Poss. Scaled Scores	Lwst. Poss. Scaled Scores	Hghst. Poss. Scaled Scores	Lwst. Poss. Scaled Scores
Spanish Cont'd				
P_4 -- N_{10}	- 27	- 15	- 28	-13
P_4 -- M_{12}	- 29	8	- 27	9
P_4 -- M_2	- 4	21	- 3	23
P_4 -- N_7	- 33	- 16	- 33	-15

Table 2

Classifications of Advanced Tests by Possible Scores and Observed Score Differences

Class I: No extreme differences of greater than 50 score points at either extreme.

Possible Scores	Observed Scores
Education	Biology
History	Education
Literature	French
Political Science	History
Spanish	Literature
	Political Science
	Psychology
	Spanish

Class II: A difference of more than 50 score points at only one extreme.

Possible Scores	Observed Scores
Biology	Chemistry
Chemistry	Economics
Economics	Engineering
Engineering	Mathematics
Mathematics	Music
Music	
Psychology	

Class III: A difference of more than 50 score points at both extremes.

Possible Scores	Observed Scores
French	Geology
Geology	Philosophy
Philosophy	Physics
Physics	Sociology
Sociology	

Table [illegible]

Observed Score Differences Between
Common Item Equating (CIE) and Verbal-Quantitative Equating (VQE)
(Presented in Terms of CIE minus VQE)

	Possible Scores		Observed Scores	
Test and Equating Forms	High	Low	High	Low
Biology				
[illegible] -- M_1	-20	30	-13	24
Chemistry				
P_{10} -- P_4	1[illegible]	- 8	88	- 8
Economics				
P_{10} -- P_4	32	-12	21	- 3
Engineering				
Q_1 -- Q_7	42	23	43	19
French				
P_1 -- P_6	-40	43	-26	37
Geology				
N_{10} -- N_2	- 4	-15	-12	-11
P_{12} -- P_7	-20	30	- 3	35
History				
M_{10} -- M_7	61	-13	43	- 2
P_1 -- P_7	- 2	13	1	15
Literature				
Q_{12} -- Q_4	15	17	12	22
P_1 -- P_7	- 8	13	0	13
Mathematics				
P_{10} -- P_7	193	-12	190	- 9
Music				
M_1 -- M_7	-14	50	- 7	60
Philosophy				
J_{10} -- J_2	-122	-26	100	-21
P_{12} -- P_7	114	-40	84	-37

Table 3 Cont'd

Test and Equating Forms	Possible Scores		Observed Scores	
	High	Low	High	Low
Political Science				
P_1 -- P_7	16	6	10	7
N_2 -- N_{10}	-30	28	-28	13
Psychology				
L_{10} -- L_7	48	- 8	34	- 8
O_1 -- O_7	-30	43	-14	42
Sociology				
J_{10} -- J_4	- 4	59	- 1	25
M_{12} -- M [illegible]	19	- 6	17	- 8
O_1 -- O_7	-11	56	2	37
Spanish				
P_1 -- P_4	5	1	3	3

Table 4

Correlations for Tests Equated to Themselves

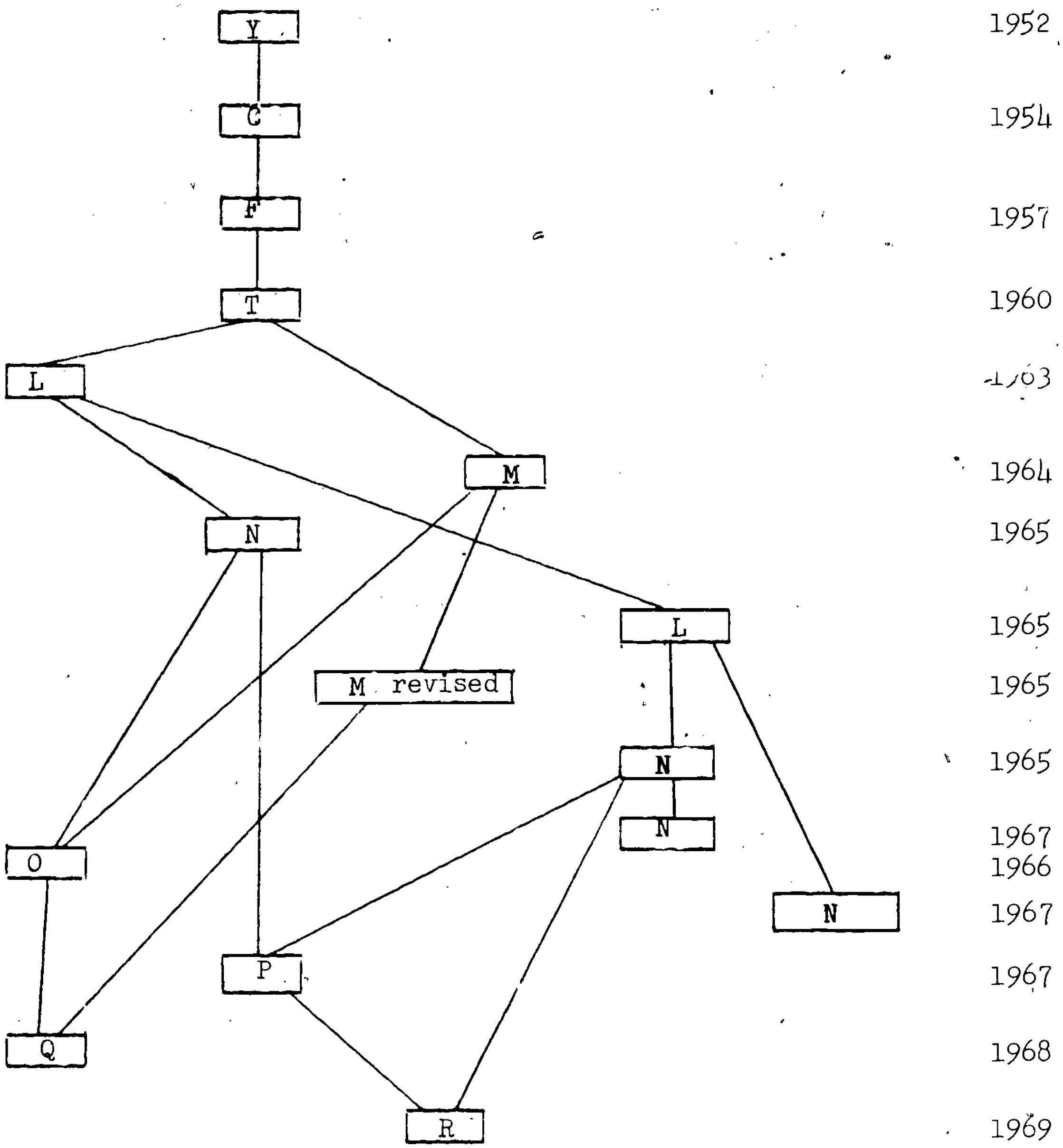
	"New Form"				"Old Form"			
Test	rxv	rxq	rvq	Rx. vq	ryv	ryq	rvq	Ry. vq
Biology	0.6478	0.5765	0.5761	0.6939	0.6141	0.5664	0.5488	0.6726
Chemistry	0.4624	0.5462	0.8065	0.5705	0.2669	0.5319	0.5287	0.5322
Economics	0.6183	0.6749	0.5953	0.7268	0.6038	0.5706	0.5521	0.6675
Engineering	0.4161	0.6047	0.5088	0.6306	0.4433	0.6247	0.4907	0.6491
French	0.6086	0.3295	0.5680	0.6089	0.5681	0.3213	0.5628	0.5681
Geology	0.5859	0.5105	0.5607	0.6258	0.5259	0.5421	0.4667	0.6237
History	0.6080	0.3270	0.4906	0.6088	0.6774	0.4253	0.5922	0.6781
Literature	0.7291	0.3472	0.5320	0.7307	0.7447	0.3570	0.5085	0.7451
Mathematics	0.5411	0.6622	0.6120	0.6841	0.3360	0.5622	0.5983	0.5622
Music	0.6278	0.5297	0.6652	0.6454	0.5401	0.4369	0.4942	0.5744
Philosophy	0.7281	0.5327	0.6004	0.7378	0.6611	0.5506	0.4982	0.7086
Political Science	0.6318	0.4478	0.6239	0.6355	0.7004	0.5394	0.5821	0.7189
Psychology	0.6907	0.5449	0.5653	0.7156	0.6297	0.4460	0.5112	0.6460
Sociology	0.7770	0.6577	0.6496	0.8026	0.7972	0.6866	0.6777	0.8217
Spanish	0.2595	0.0180	0.4212	0.2784	0.3610	0.0675	0.5603	0.3959

Appendix 1

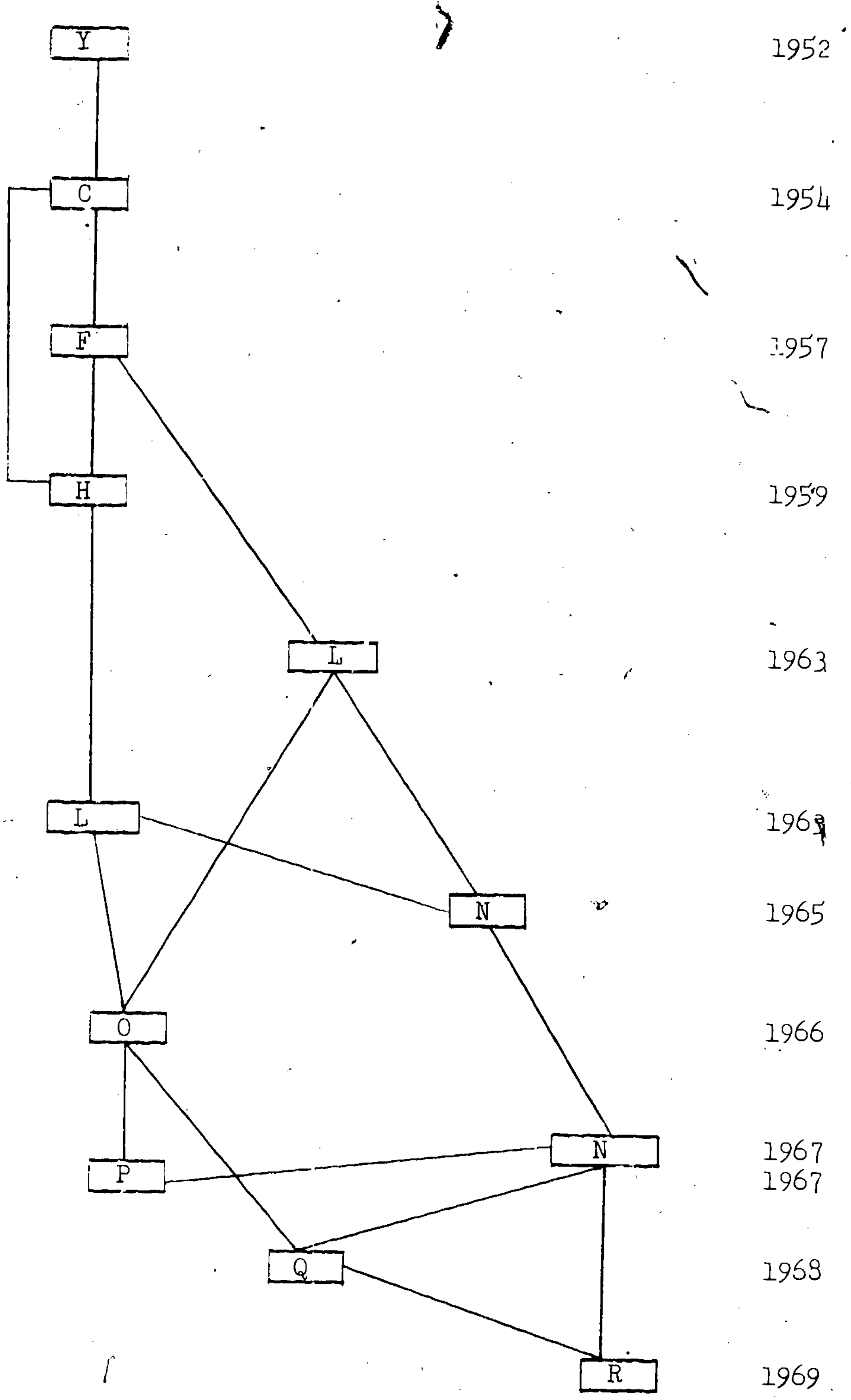
Appendix 1

The Graduate Record Examinations Genealogical Charts of

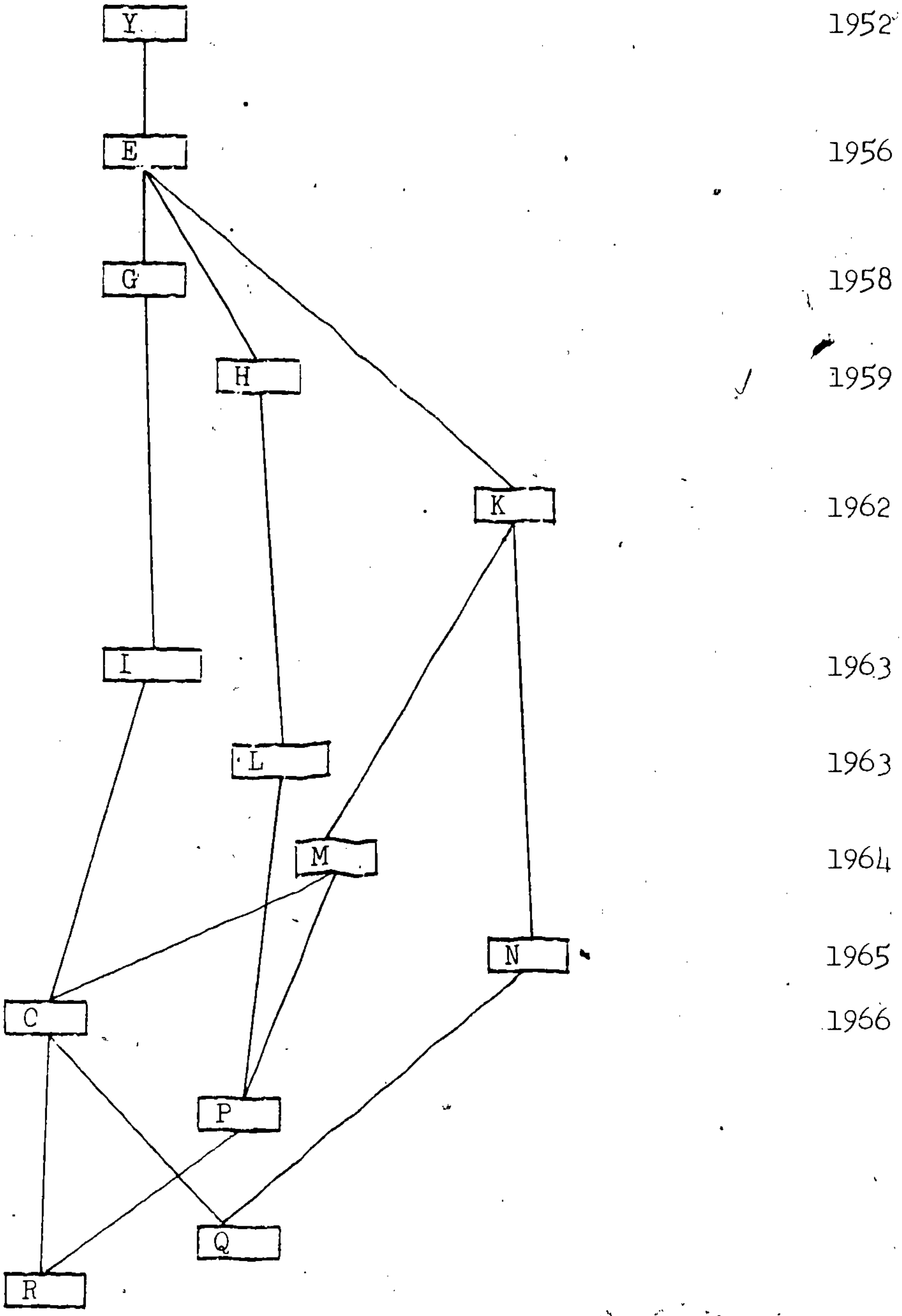
Advanced Biology Test



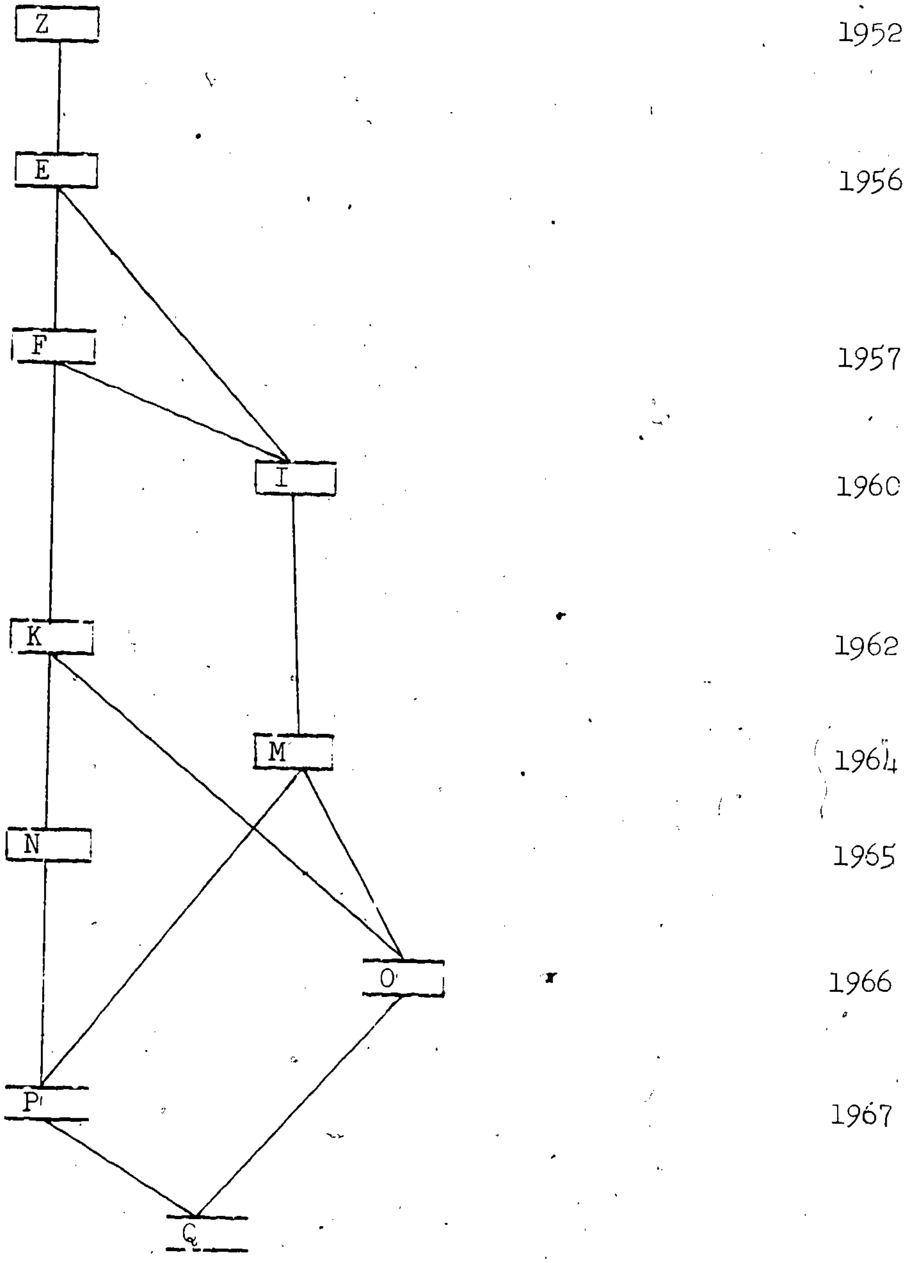
Advanced Chemistry Test



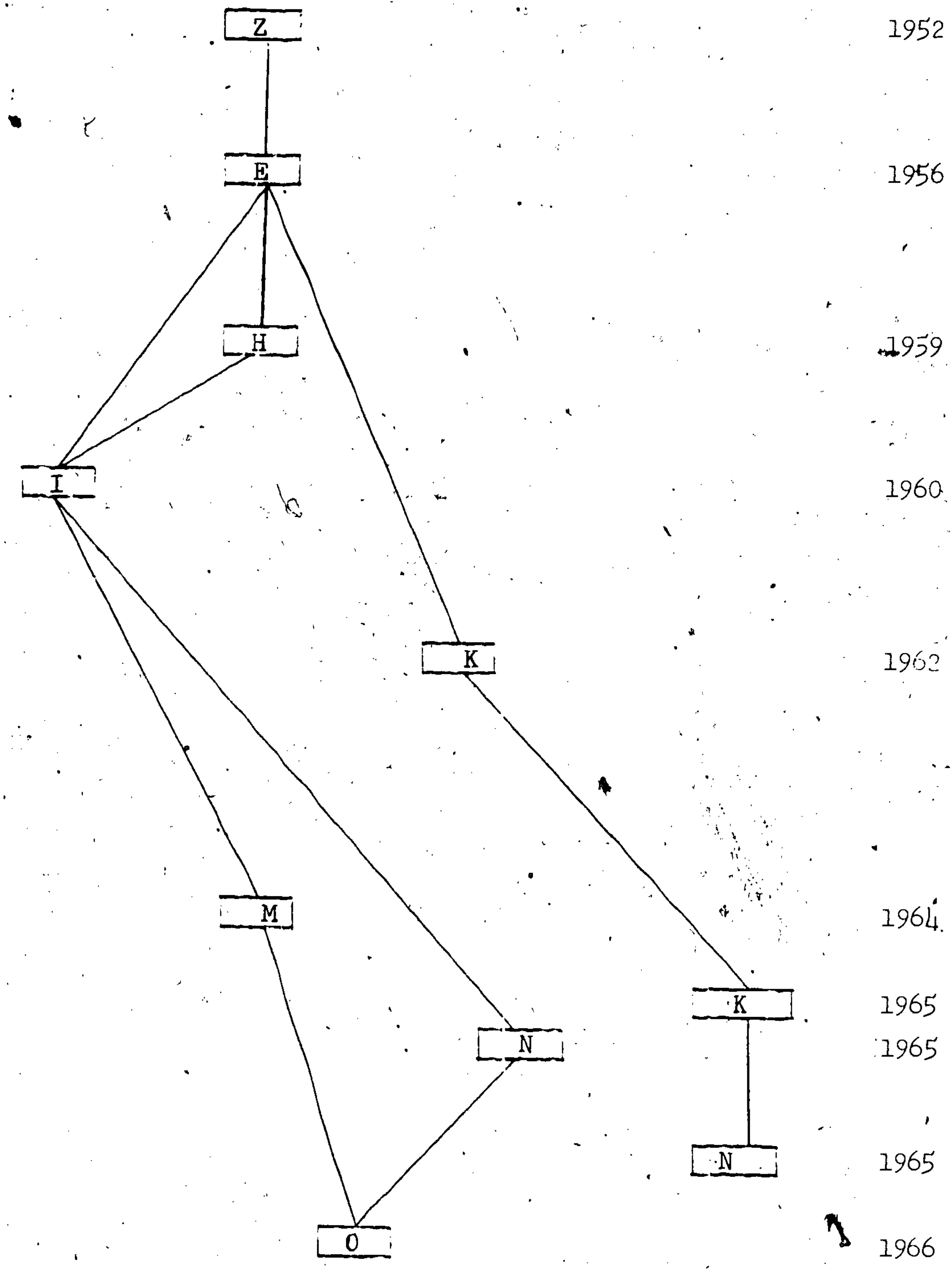
Advanced Education Test



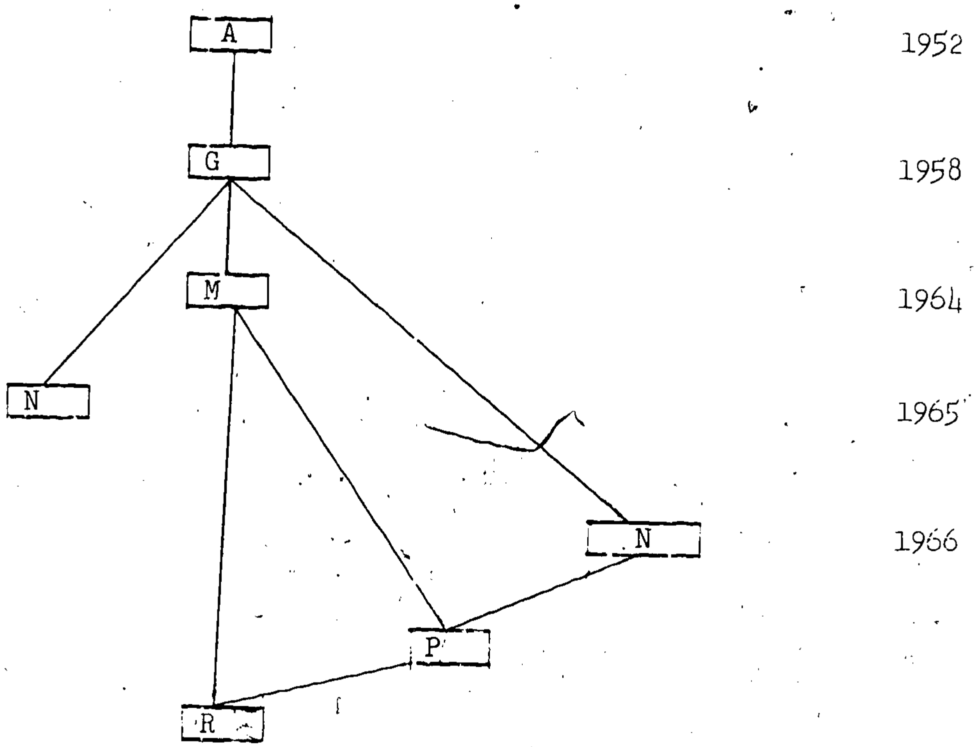
Advanced Engineering Test



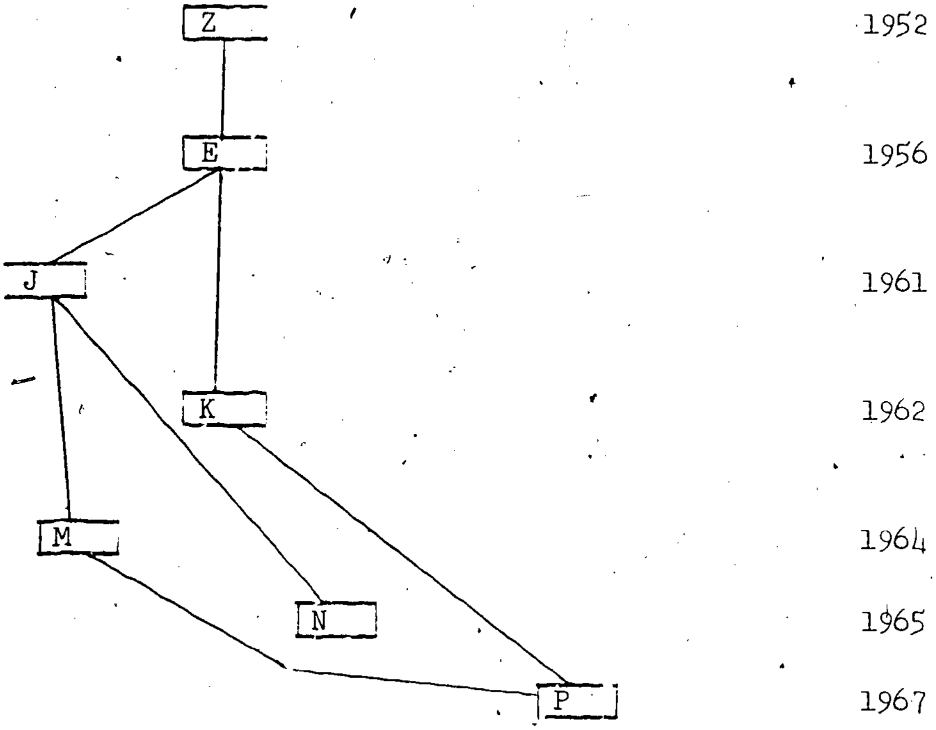
Advanced Economics Test



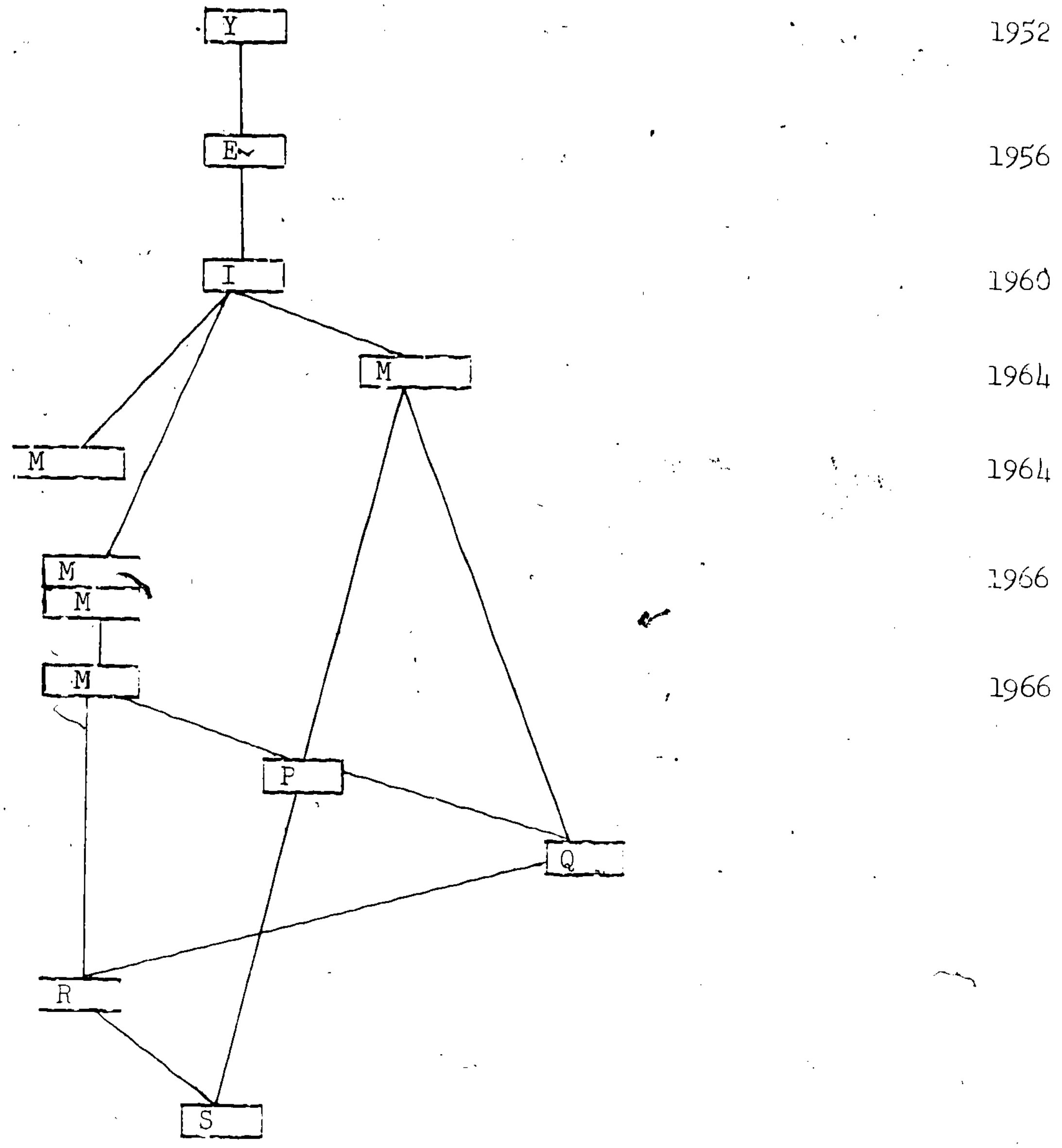
Advanced French Test



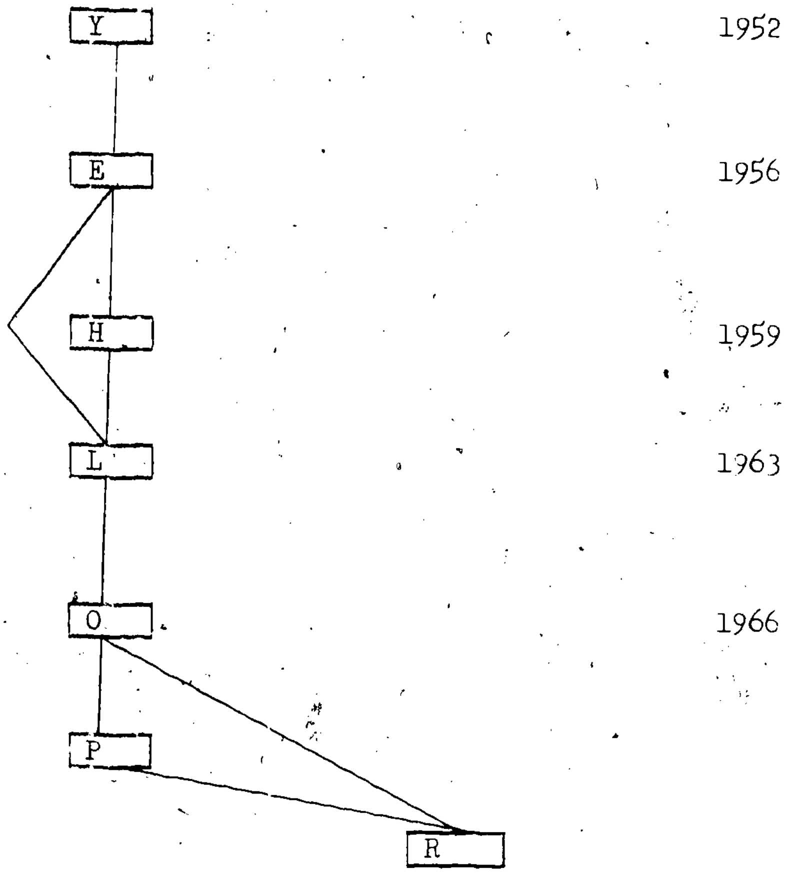
Advanced Geology Test



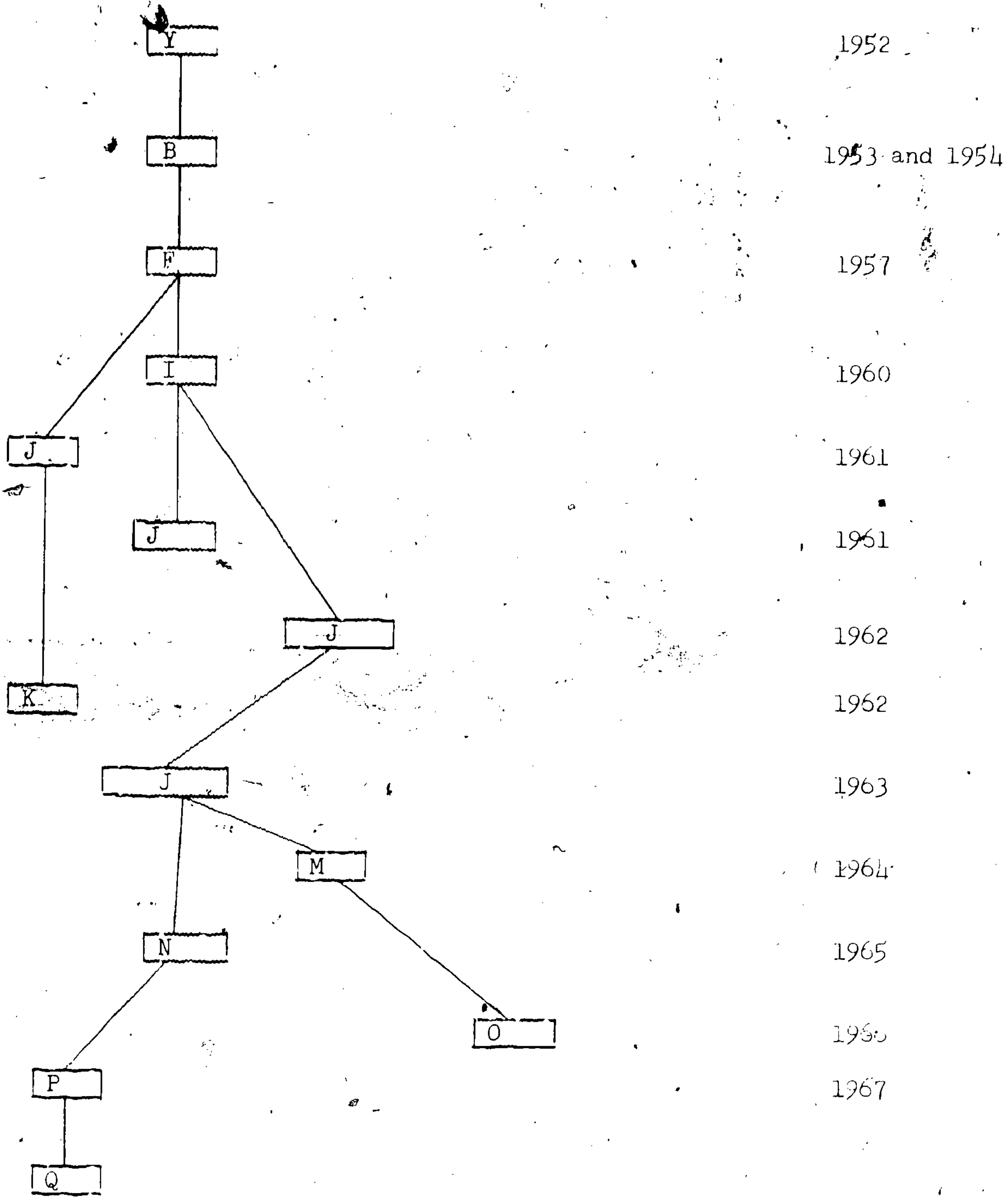
Advanced History Test



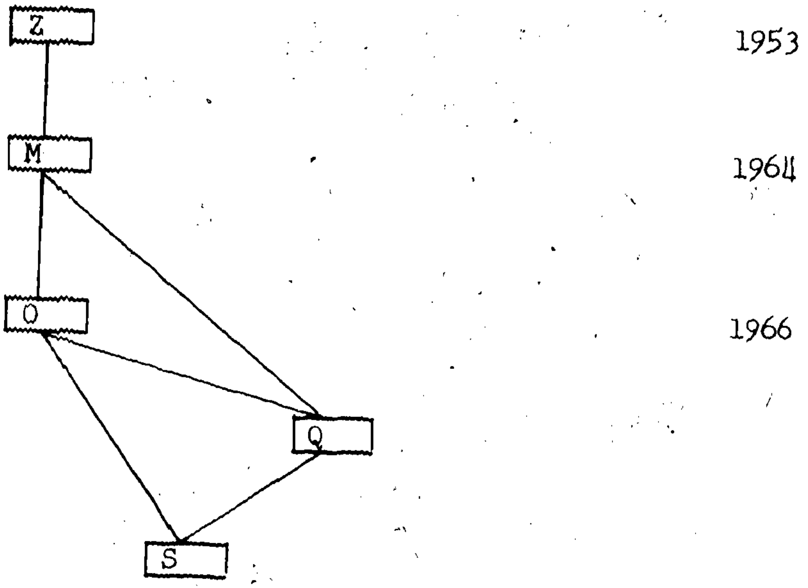
Advanced Literature Test



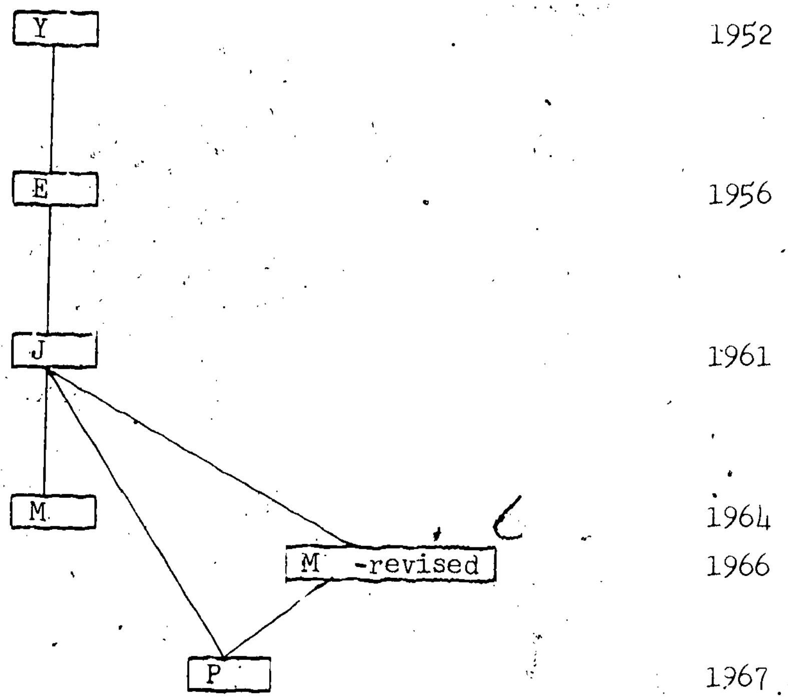
Advanced Mathematics Test



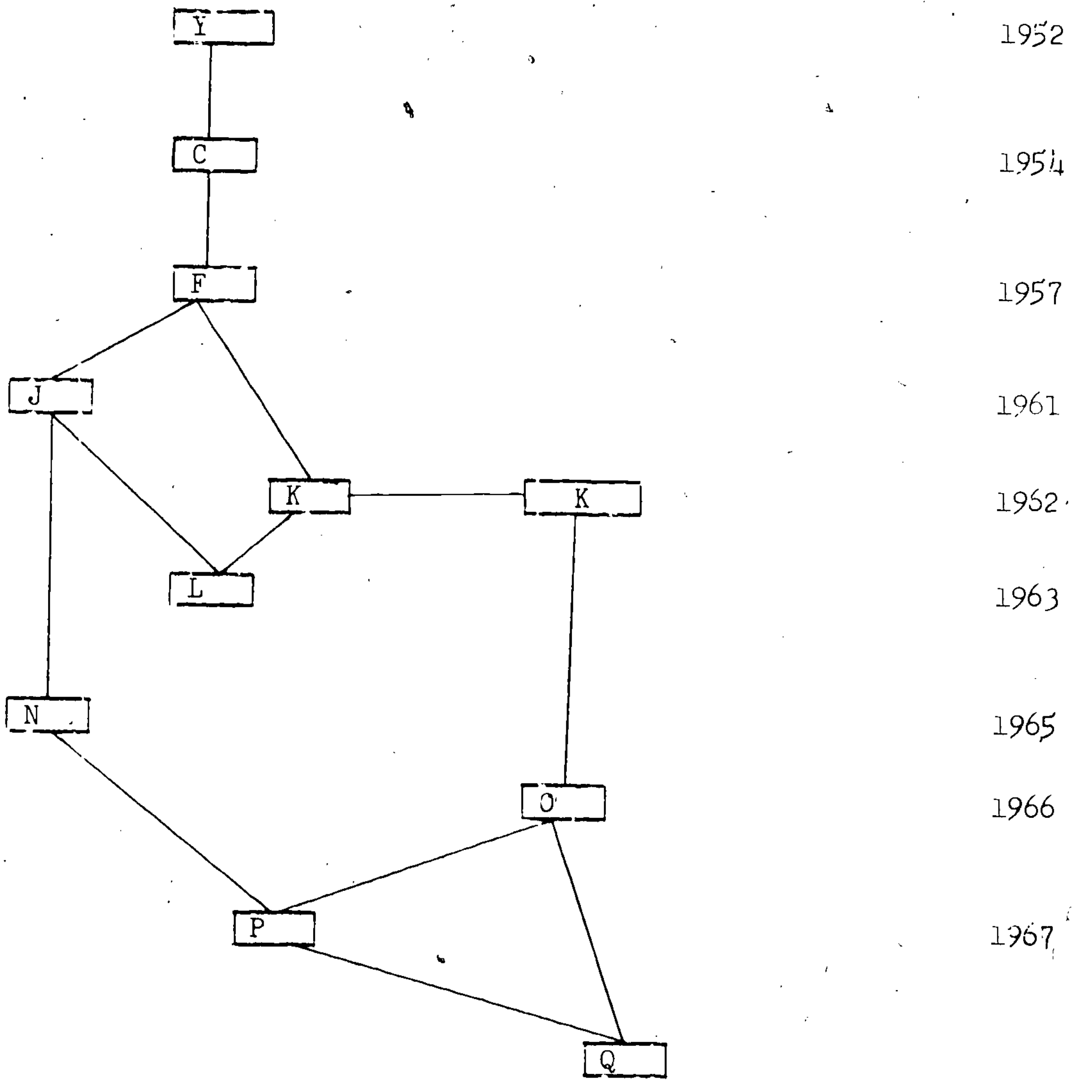
Advanced Music Test



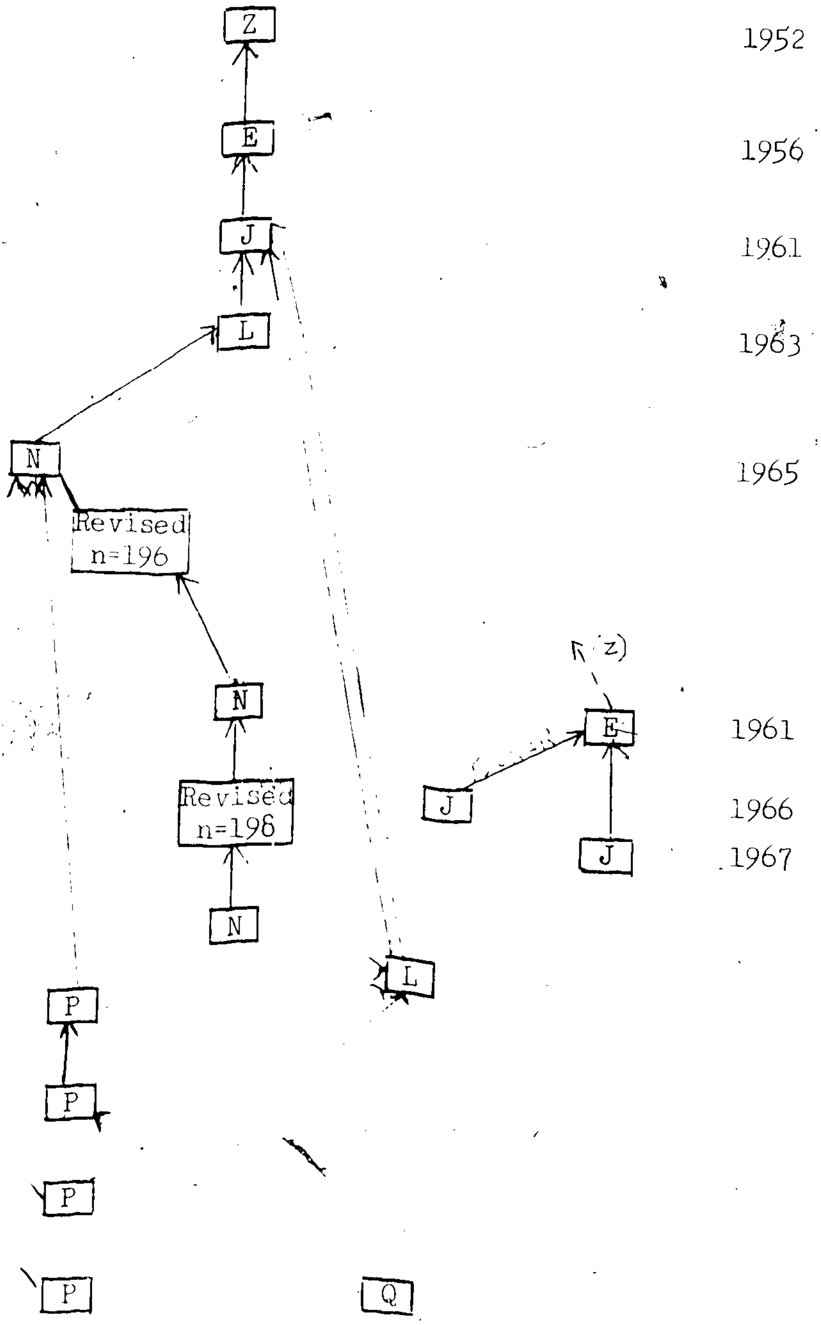
Advanced Philosophy Test



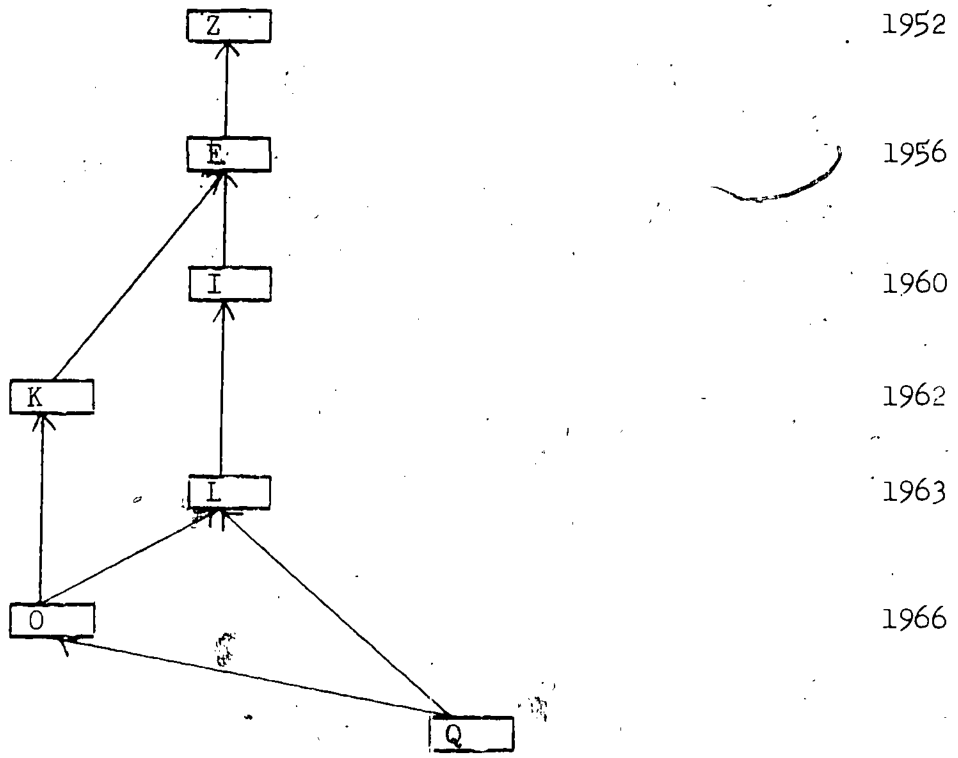
Advanced Physics Test



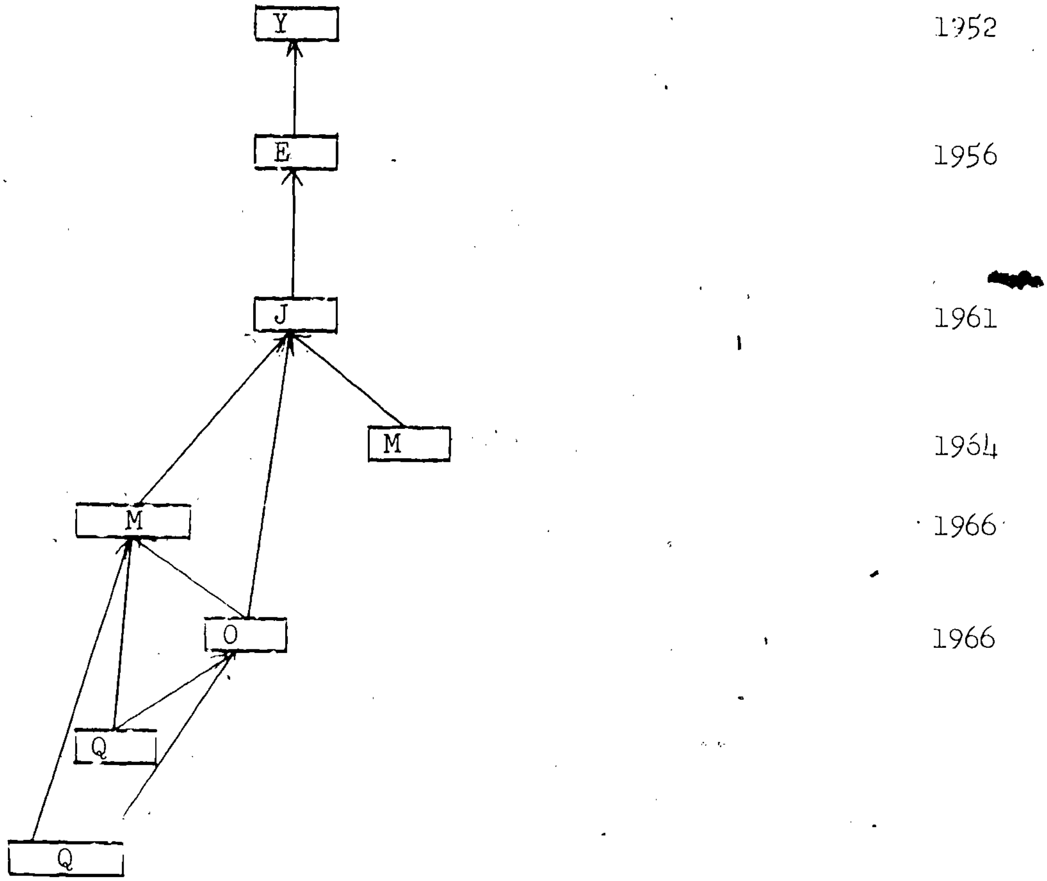
Advanced Political Science



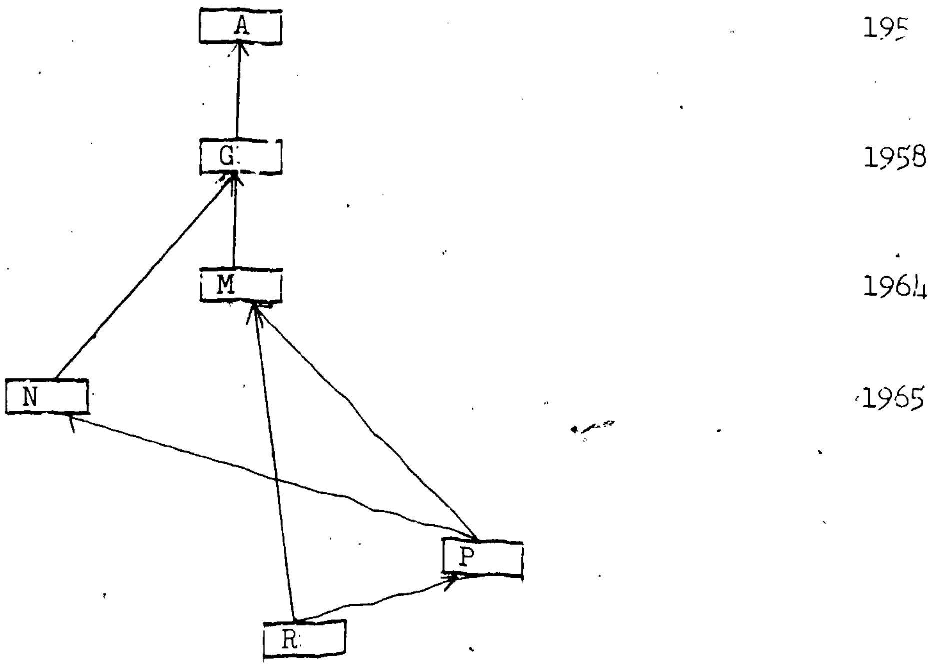
Advanced Psychology Test



Advanced Sociology Test



Advanced Spanish Test



Appendix 2

Appendix 2

Method Used for Equating GRE Advanced Tests Using Verbal and Quantitative Aptitude Test Scores as Anchors

Suppose two different groups of candidates take two different test forms designated as form X and form Y. We denote the group taking test X as group r and the group taking test Y as group s. Suppose further that test Y has been given sometime in the past and that test X has been recently administered and that both groups have taken a Verbal and Quantitative test denoted V and Q respectively. Thus, a group r has scores on tests X, V, and Q and group s has scores on tests Y, V, and Q.

We call form X the "new form" and form Y the "old form" and desire to make scores on test X comparable to scores on test Y. To do this, we conceptualize two regressions for each test form. For form X we consider the regression of the score on test X on the scores of V and Q for the group r, and do the same for the total group t = r + s, even though the total group did not take test X. These two regressions are denoted by

$$\tilde{X}_r = a_r + b_{1r}V_{1r} + b_{2r}Q_r \qquad (1)$$

and

$$\tilde{X}_t = a_{1t} + b_{1t}V_t + b_{2t}Q_t \qquad (2)$$

We now make three assumptions, the first being that the slopes for the two groups, r and t, are the same, i.e.,

$$\tilde{X}_r - b_{1r}\bar{V} - b_{2r}\bar{Q}_r = \tilde{X}_t - b_{1t}\bar{V}_t - B_{2t}\bar{Q}_t \quad , \qquad (3)$$

and the second being that the regression coefficients are the same, i.e.,

$$b_{1r} = b_{1t} \qquad (4)$$

$$b_{2r} = b_{2t} \qquad (5)$$

And finally the variance error of estimate, the expected squared error from prediction denoted VE,

$$VE = S_x^2 - bCb'$$

where $b = (b_1, b_2)$

C = the covariance matrix of V and Q ,

is the same for both groups,

$$S_{x_r}^2 - b_r C_r b_r' = S_{x_t}^2 - b_t C_t b_t \quad . \qquad (6)$$

Substituting equations (4) and (5) into (3) and solving for X_t we obtain

$$\tilde{X}_t = \tilde{X}_r + b_{1r}(\tilde{V}_t - \tilde{V}_r) + b_{2r}(\tilde{Q}_t - \tilde{Q}_r) \quad , \qquad (7)$$

and since we know all of the terms on the right hand side of the equations, we have an estimate of how the total group would have done on test X.

Substituting (4) and (5) into (6) and solving for $S_{x_t}^2$ we obtain

$$S_{x_t}^2 = S_{x_r}^2 + b_r(C_t - C_r)\, b_r' \quad . \qquad (8)$$

Using exactly the same assumptions and development for the relationship between Forms Y and V and Q with groups s and t we obtain estimates $\bar{Y}_t$ and $S_{y_t}^2$ for the mean and variance using the total group.

The conversion of the scores on test X to the corresponding scores on test Y is found by

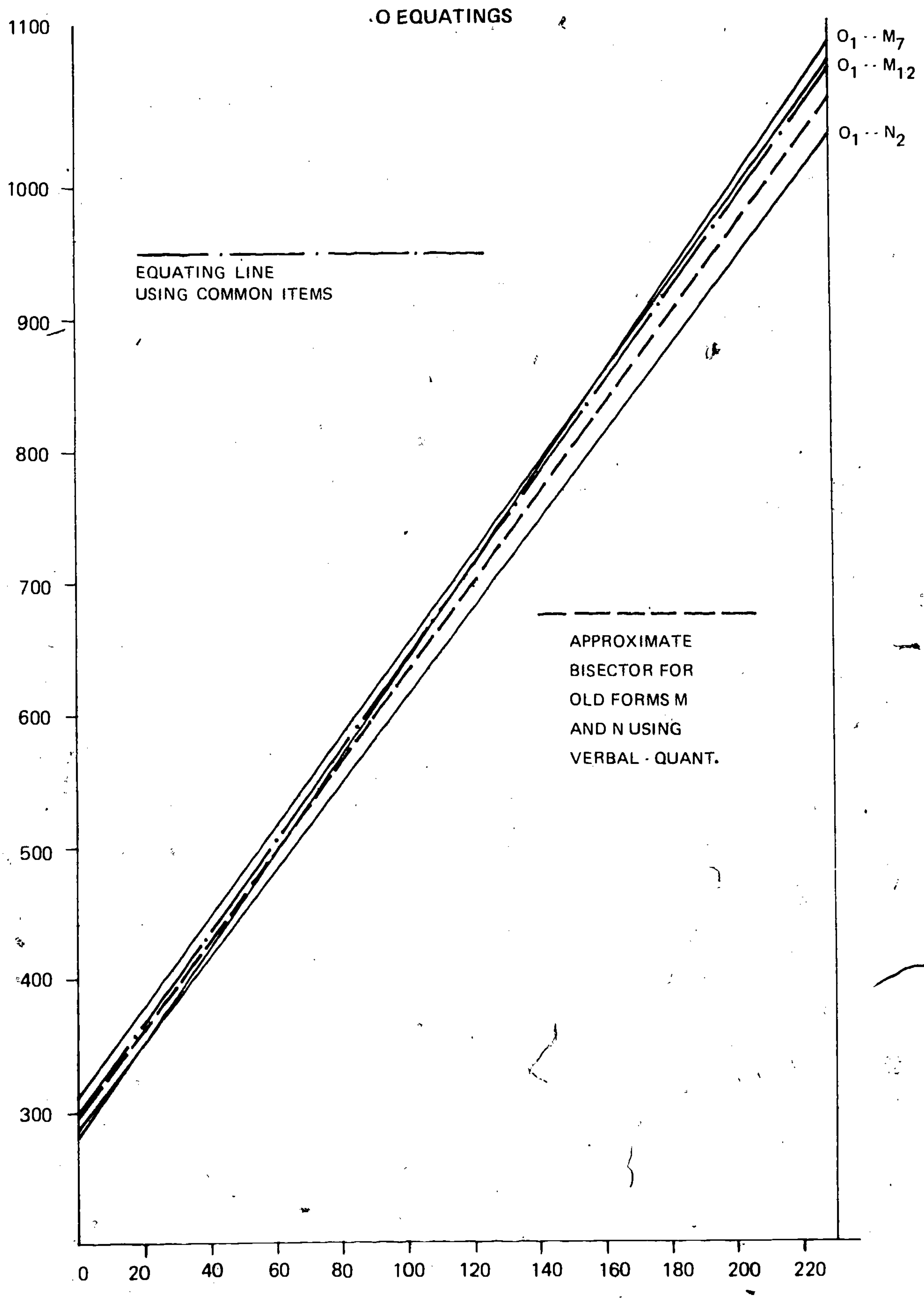
$$Y = a' + b' X$$

where $b' = \dfrac{S_{y_t}}{S_{x_t}}$ and $a' = Y_t - b'X_t$.

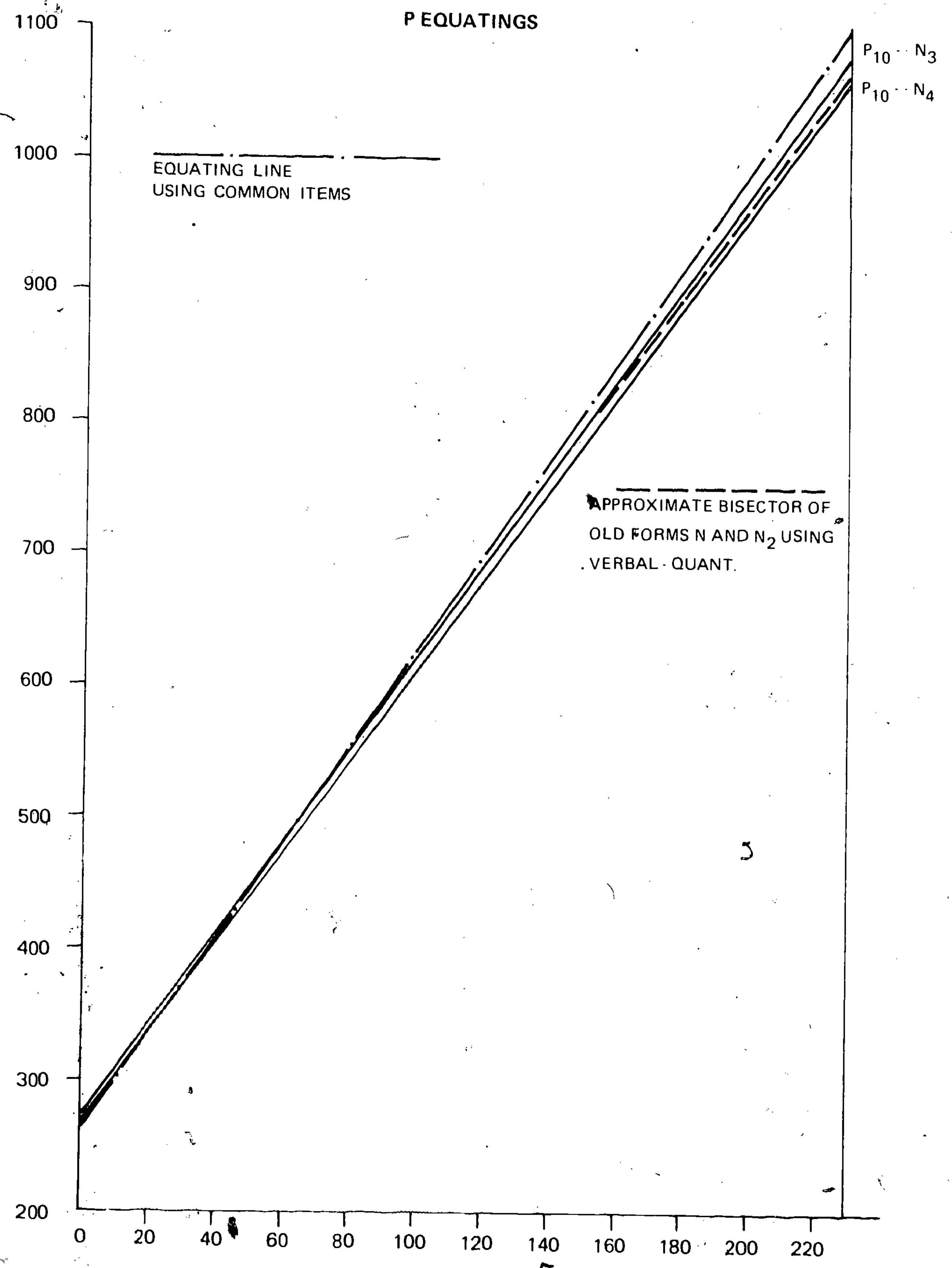
The common-item approach utilizes exactly the same approach only using an anchor test (usually common items) denoted Z instead of V and Q.

Appendix 3

BIOLOGY

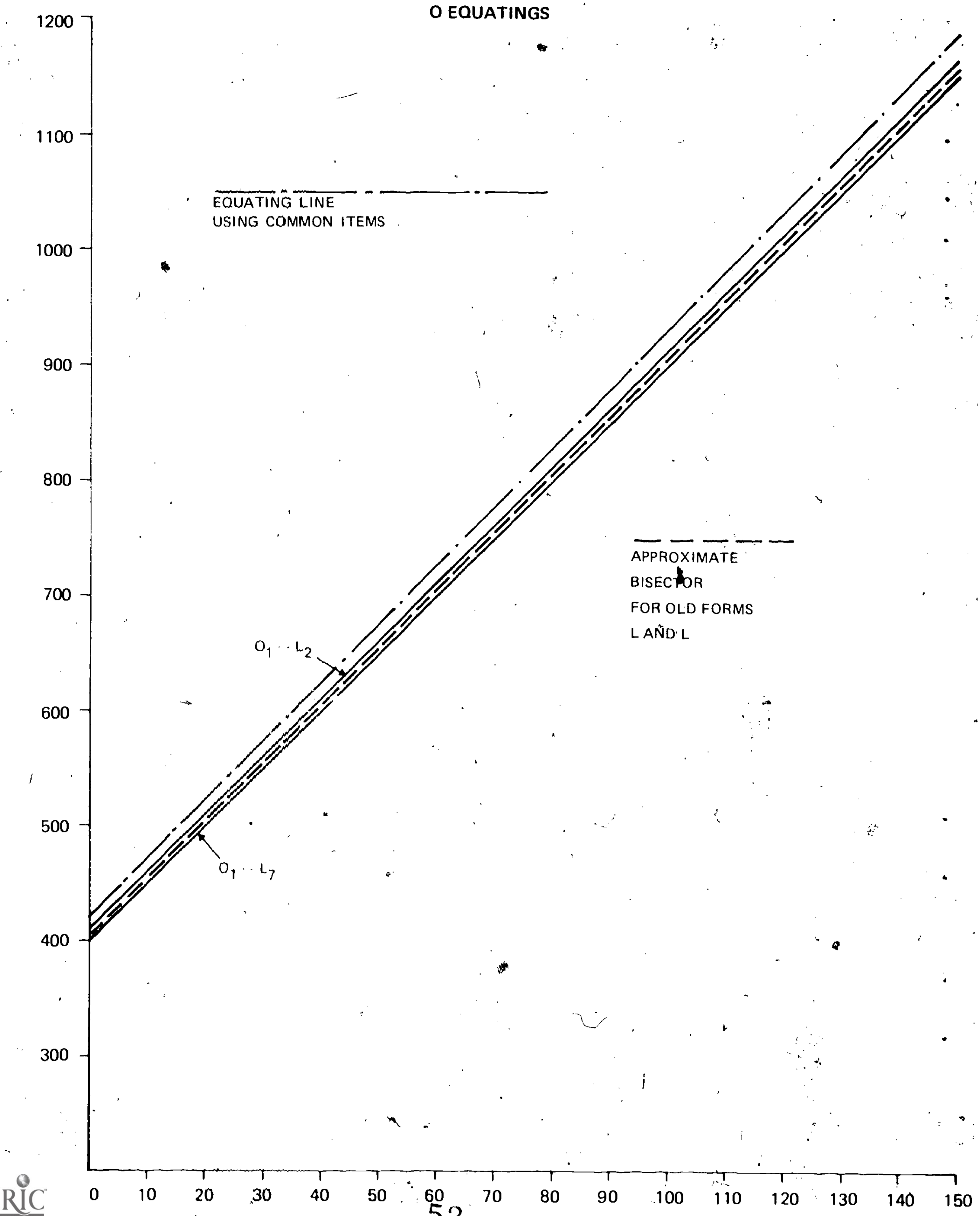

.O EQUATINGS
1100
1000
900
800
700
600
500
400
300
0
20
40
60
80
100
120
140
160
180
200
220
$O_1 \cdot\cdot M_7$
$O_1 \cdot\cdot M_{12}$
$O_1 \cdot\cdot N_2$
EQUATING LINE
USING COMMON ITEMS
APPROXIMATE
BISECTOR FOR
OLD FORMS M
AND N USING
VERBAL - QUANT.

BIOLOGY

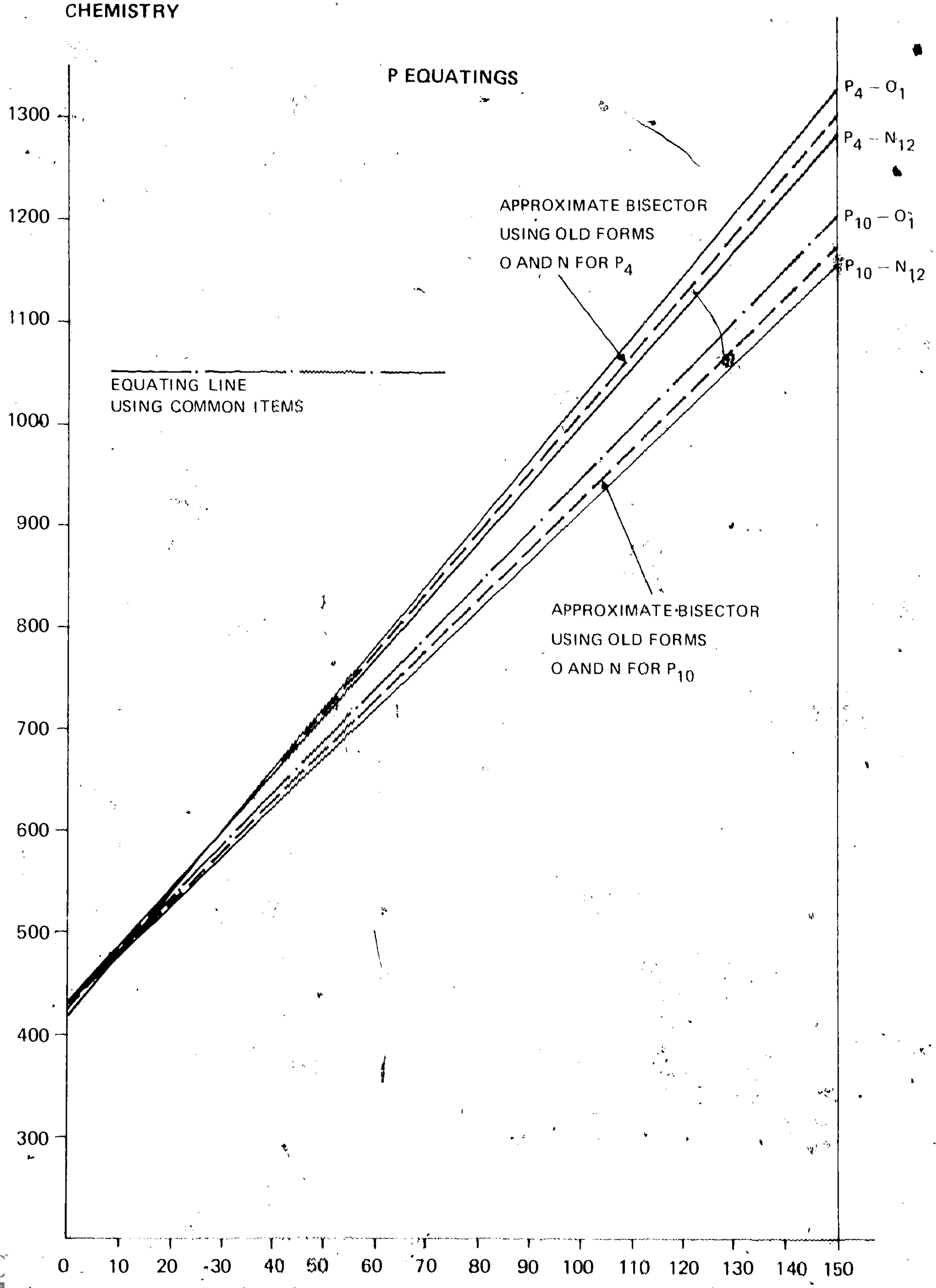

P EQUATINGS
EQUATING LINE
USING COMMON ITEMS
APPROXIMATE BISECTOR OF
OLD FORMS N AND N2 USING
VERBAL - QUANT.
P10 - - N3
P10 - - N4
1100
1000
900
800
700
600
500
400
300
200
0
20
40
60
80
100
120
140
160
180
200
220

CHEMISTRY

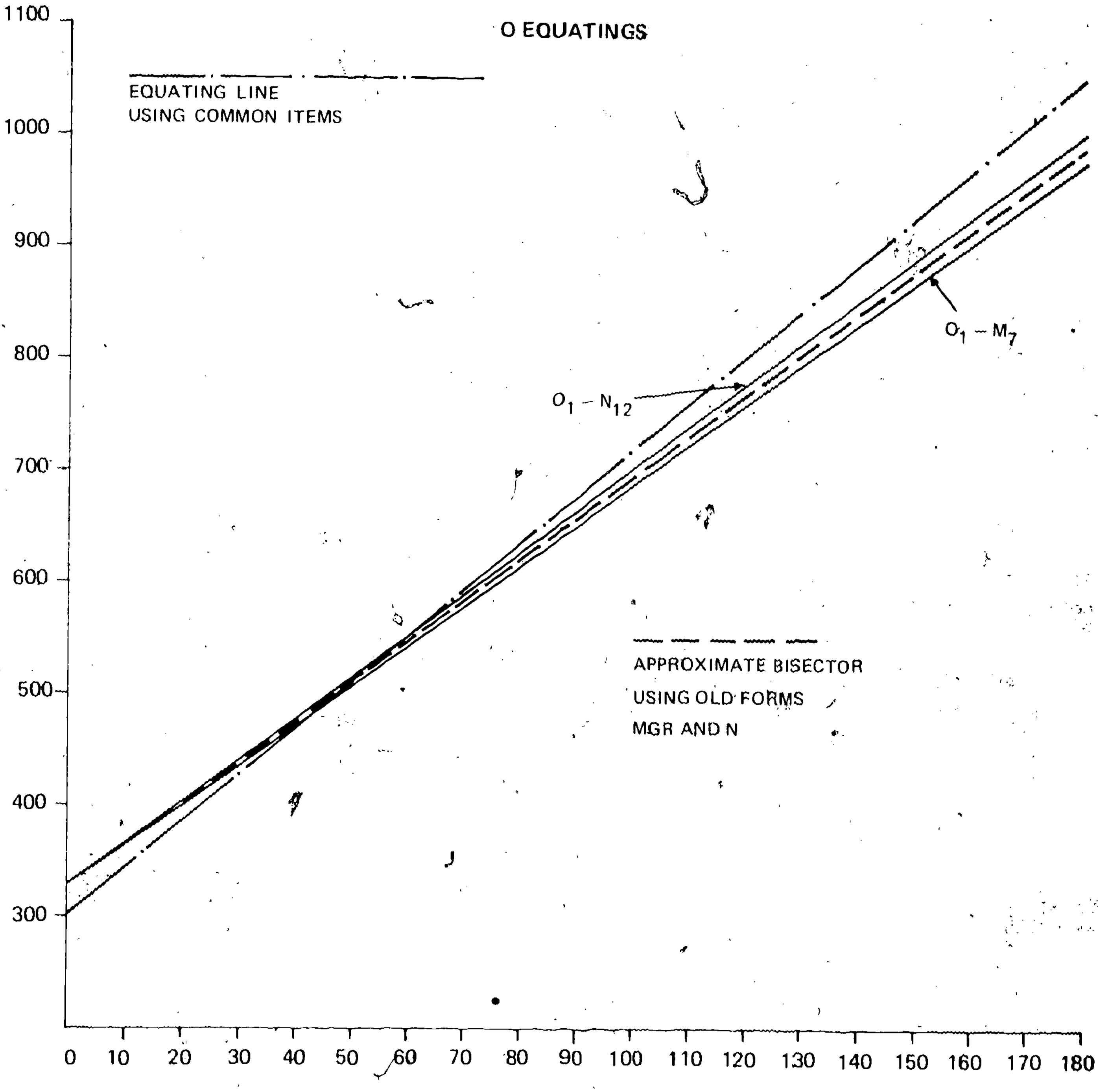
O EQUATINGS



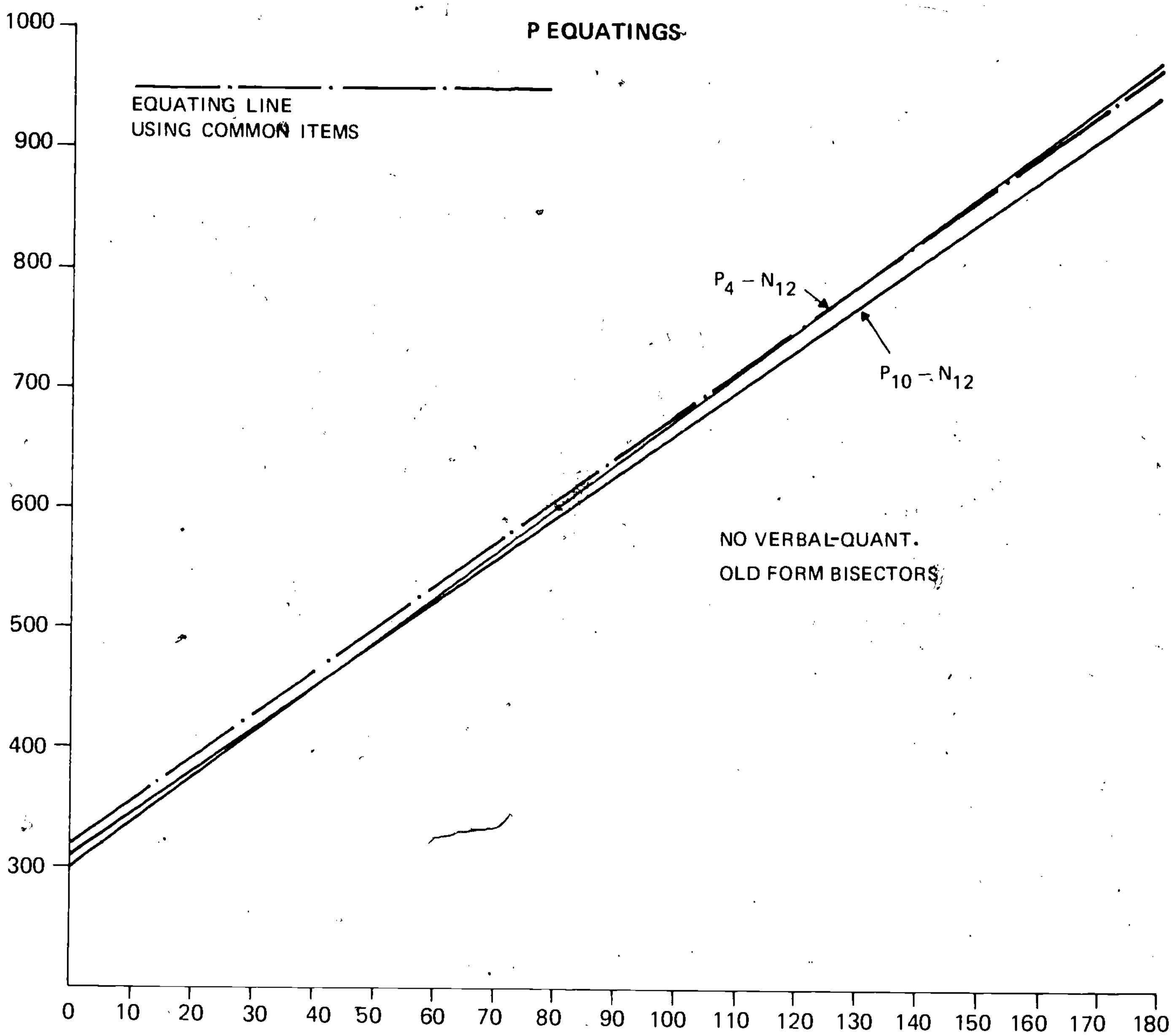
CHEMISTRY


P EQUATINGS
$P_4 - O_1$
$P_4 - N_{12}$
$P_{10} - O_1$
$P_{10} - N_{12}$
APPROXIMATE BISECTOR USING OLD FORMS O AND N FOR P_4
APPROXIMATE BISECTOR USING OLD FORMS O AND N FOR P_{10}
EQUATING LINE USING COMMON ITEMS
1300
1200
1100
1000
900
800
700
600
500
400
300
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150

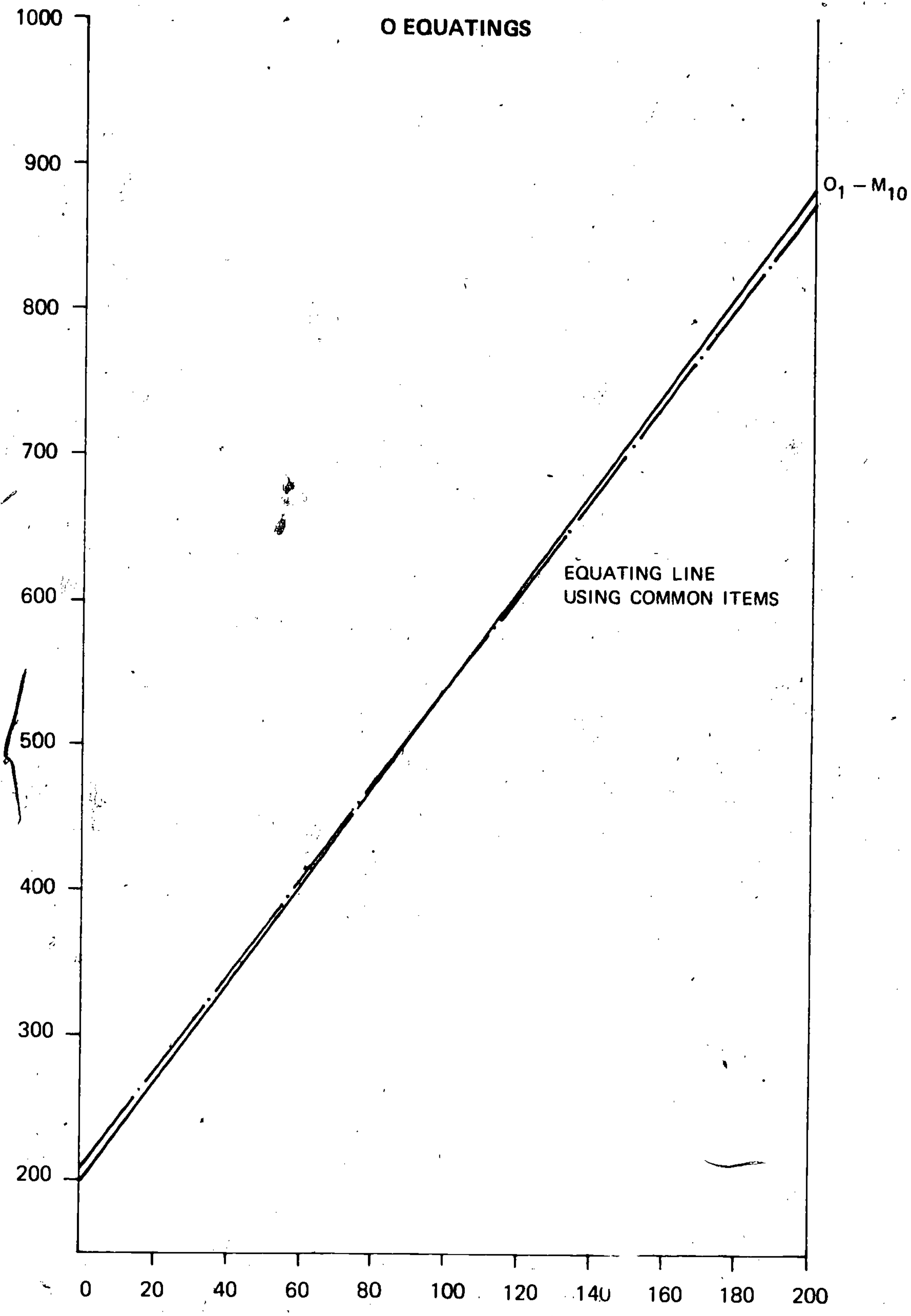
ECONOMICS


O EQUATINGS
EQUATING LINE
USING COMMON ITEMS
APPROXIMATE BISECTOR
USING OLD FORMS
MGR AND N
$O_1 - N_{12}$
$O_1 - M_7$
1100
1000
900
800
700
600
500
400
300
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180

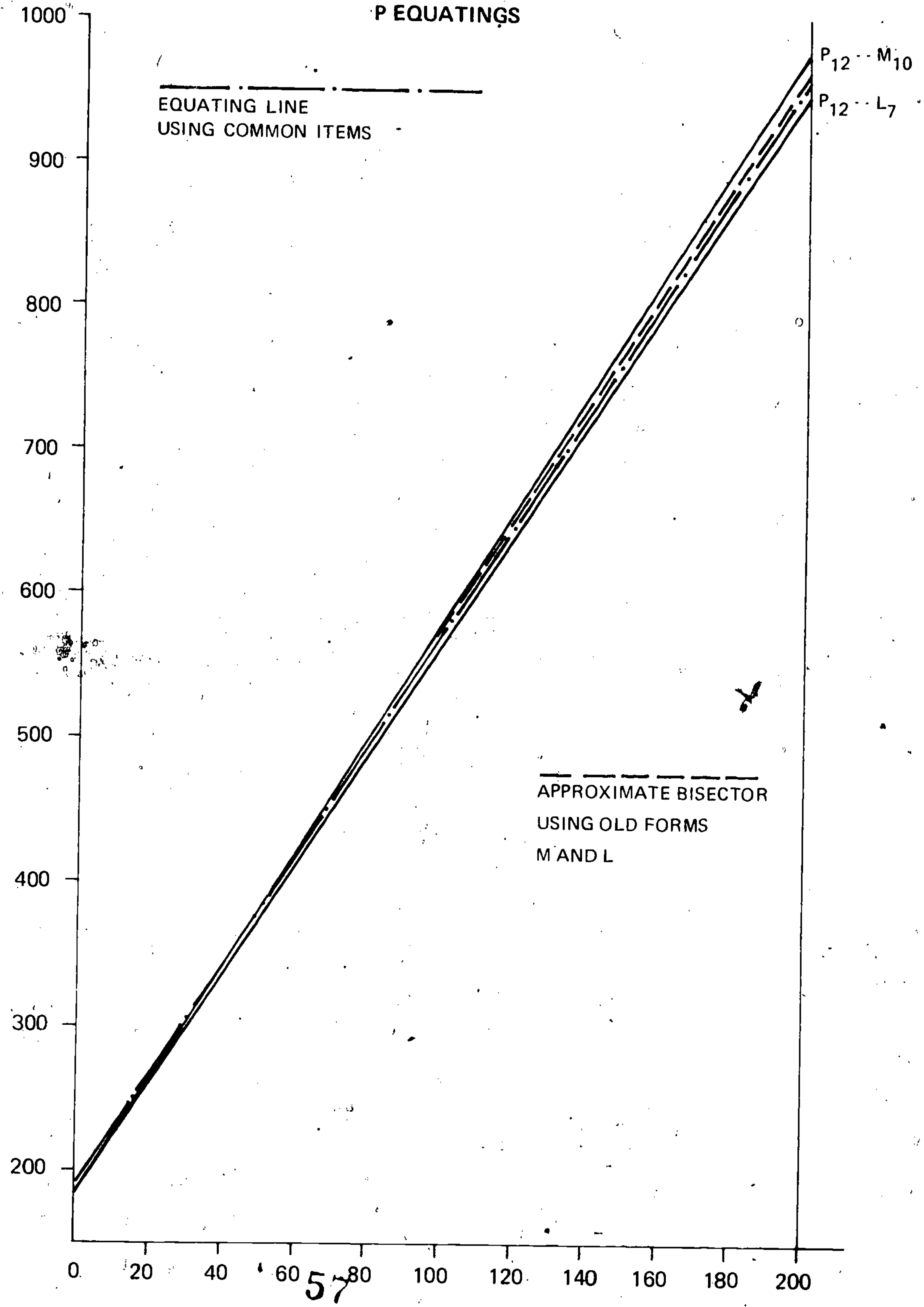
ECONOMICS


P EQUATINGS
EQUATING LINE
USING COMMON ITEMS
$P_4 - N_{12}$
$P_{10} - N_{12}$
NO VERBAL-QUANT.
OLD FORM BISECTORS
1000
900
800
700
600
500
400
300
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180

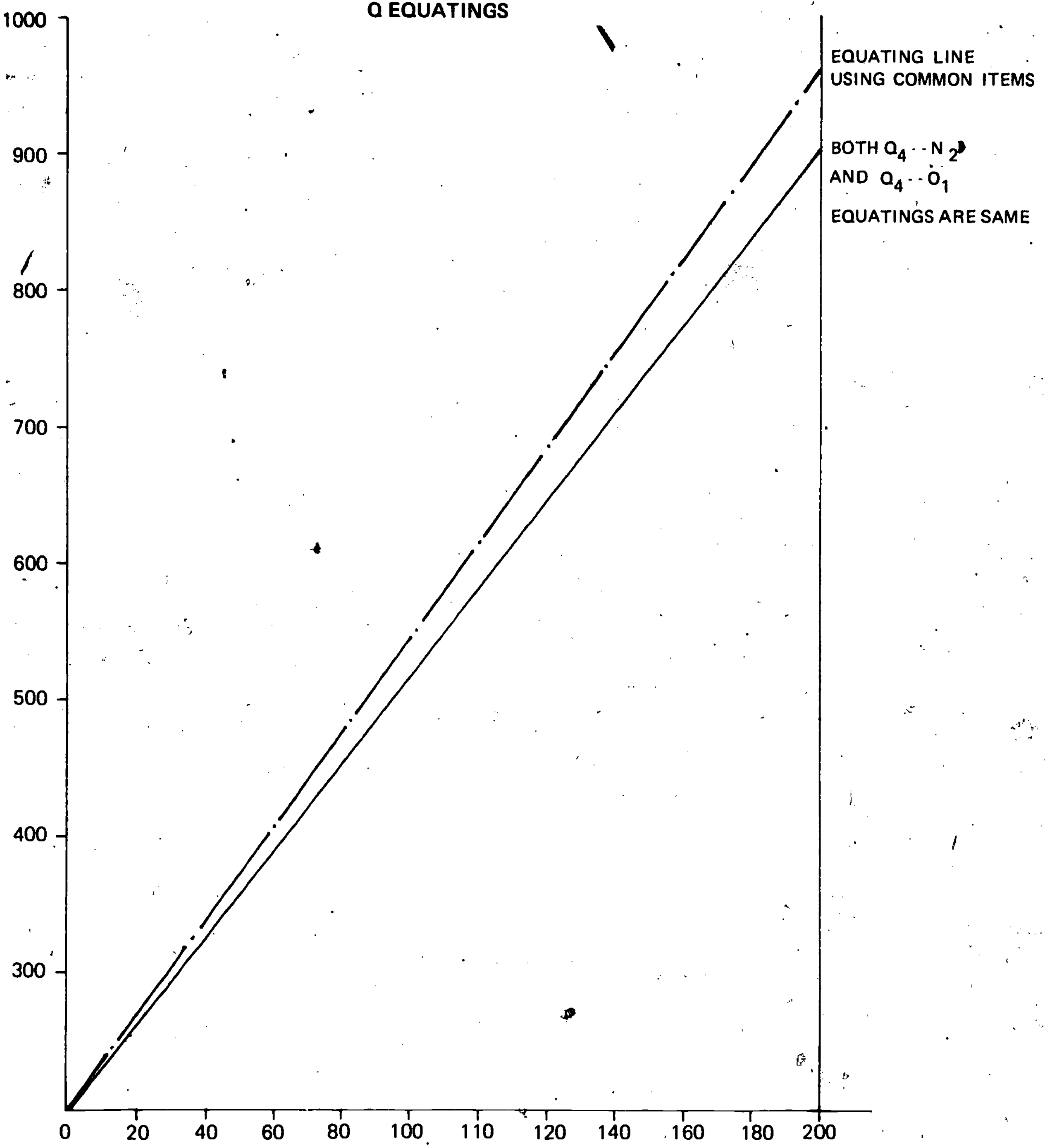
EDUCATION


O EQUATINGS
1000
900
800
700
600
500
400
300
200
$O_1 - M_{10}$
EQUATING LINE
USING COMMON ITEMS
0
20
40
60
80
100
120
140
160
180
200

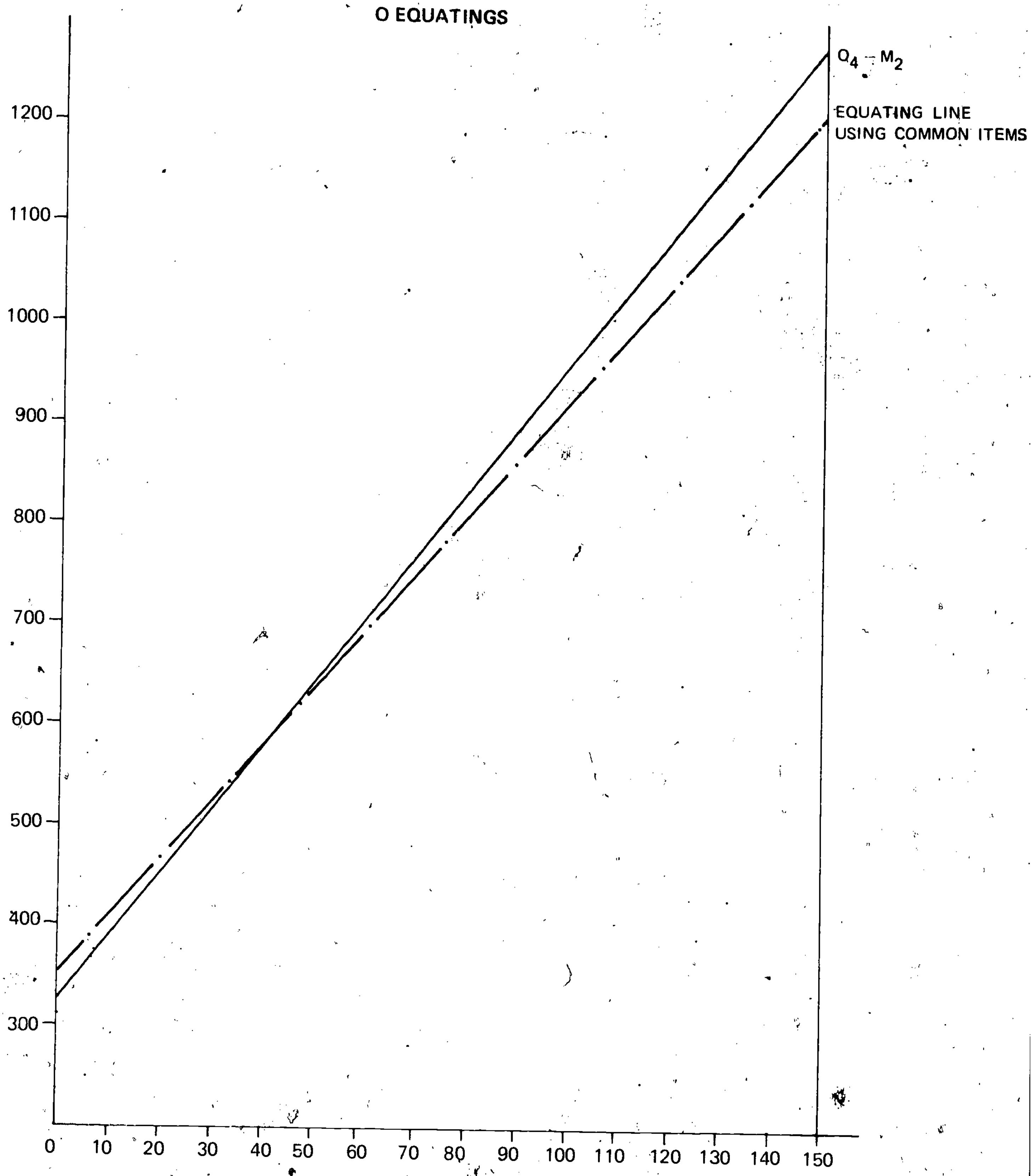
EDUCATION


P EQUATINGS
EQUATING LINE
USING COMMON ITEMS
APPROXIMATE BISECTOR
USING OLD FORMS
M AND L
P_{12} - - M_{10}
P_{12} - - L_7
1000
900
800
700
600
500
400
300
200
0
20
40
60
80
100
120
140
160
180
200

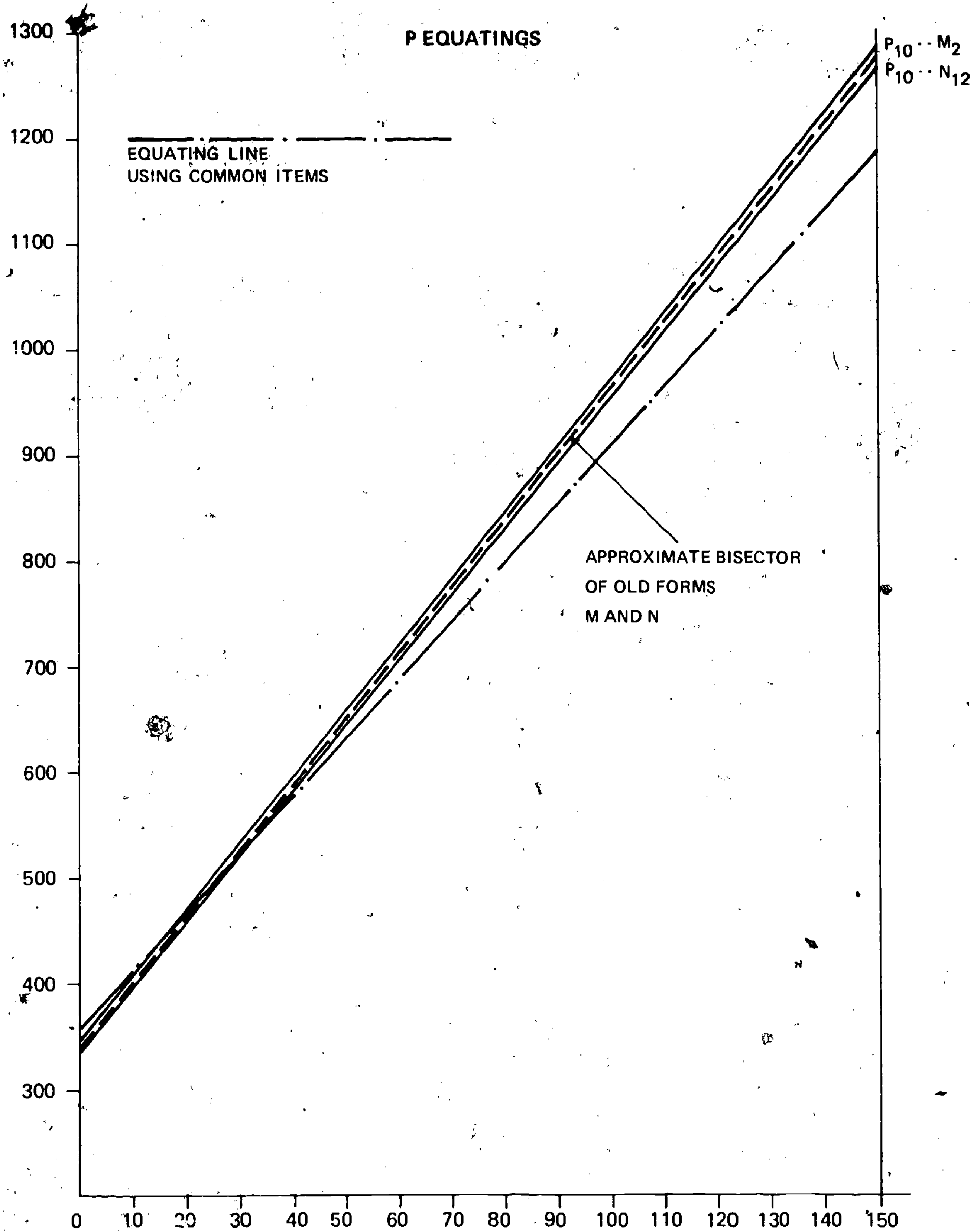
EDUCATION


Q EQUATINGS
1000
900
800
700
600
500
400
300
0
20
40
60
80
100
110
120
140
160
180
200
EQUATING LINE
USING COMMON ITEMS
BOTH Q_4 - - N_2
AND Q_4 - - O_1
EQUATINGS ARE SAME

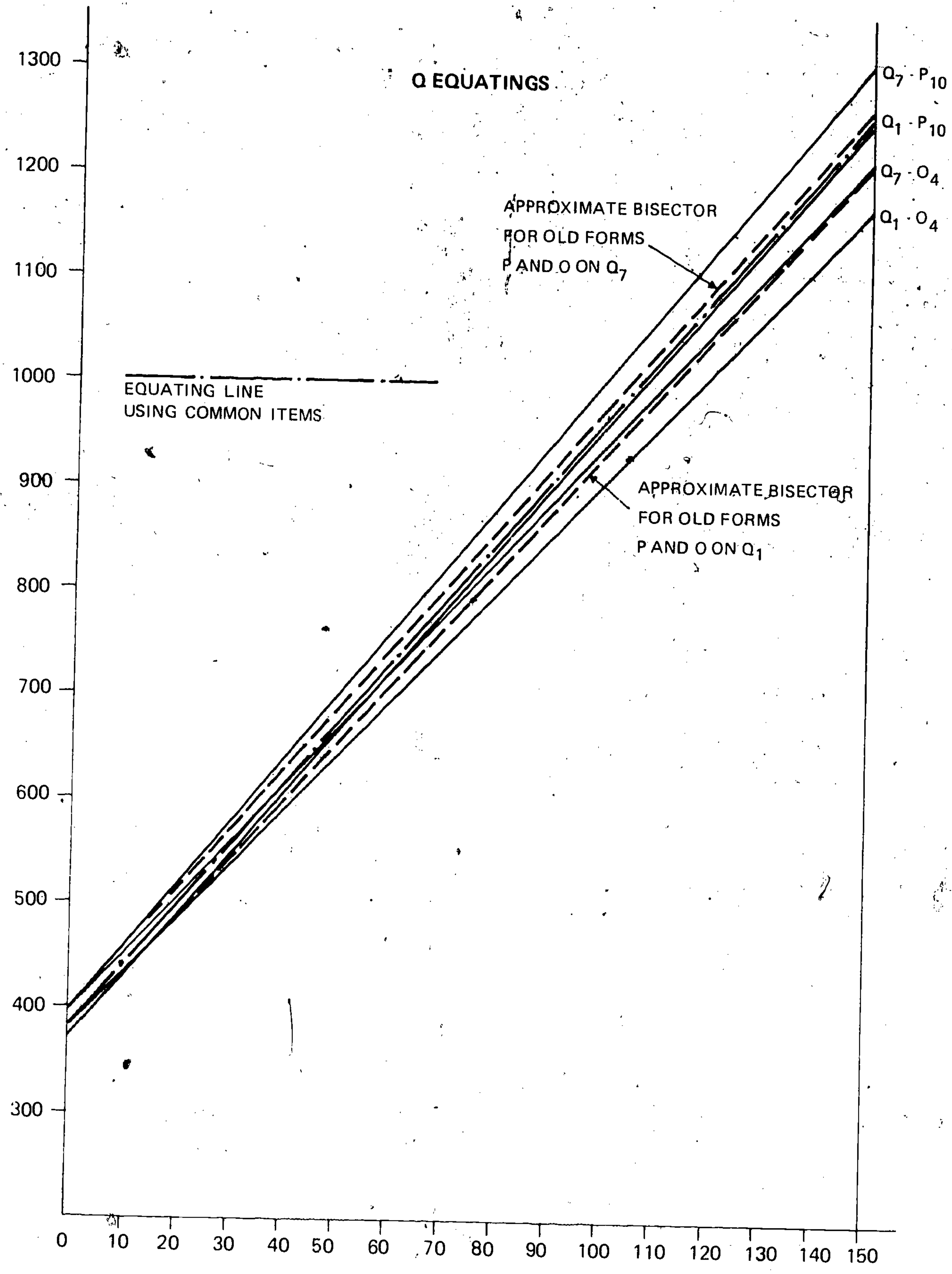
ENGINEERING



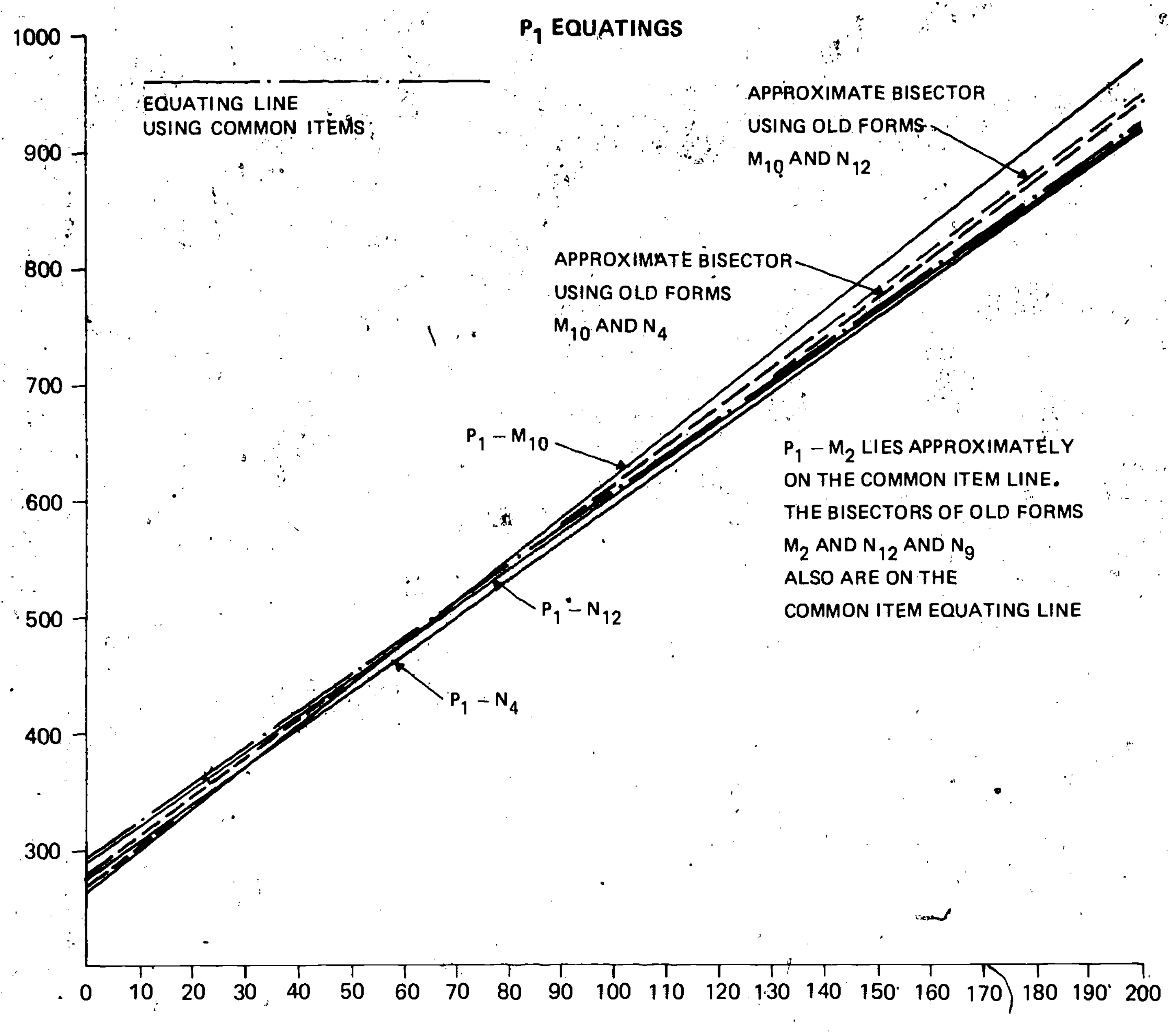
ENGINEERING



ENGINEERING

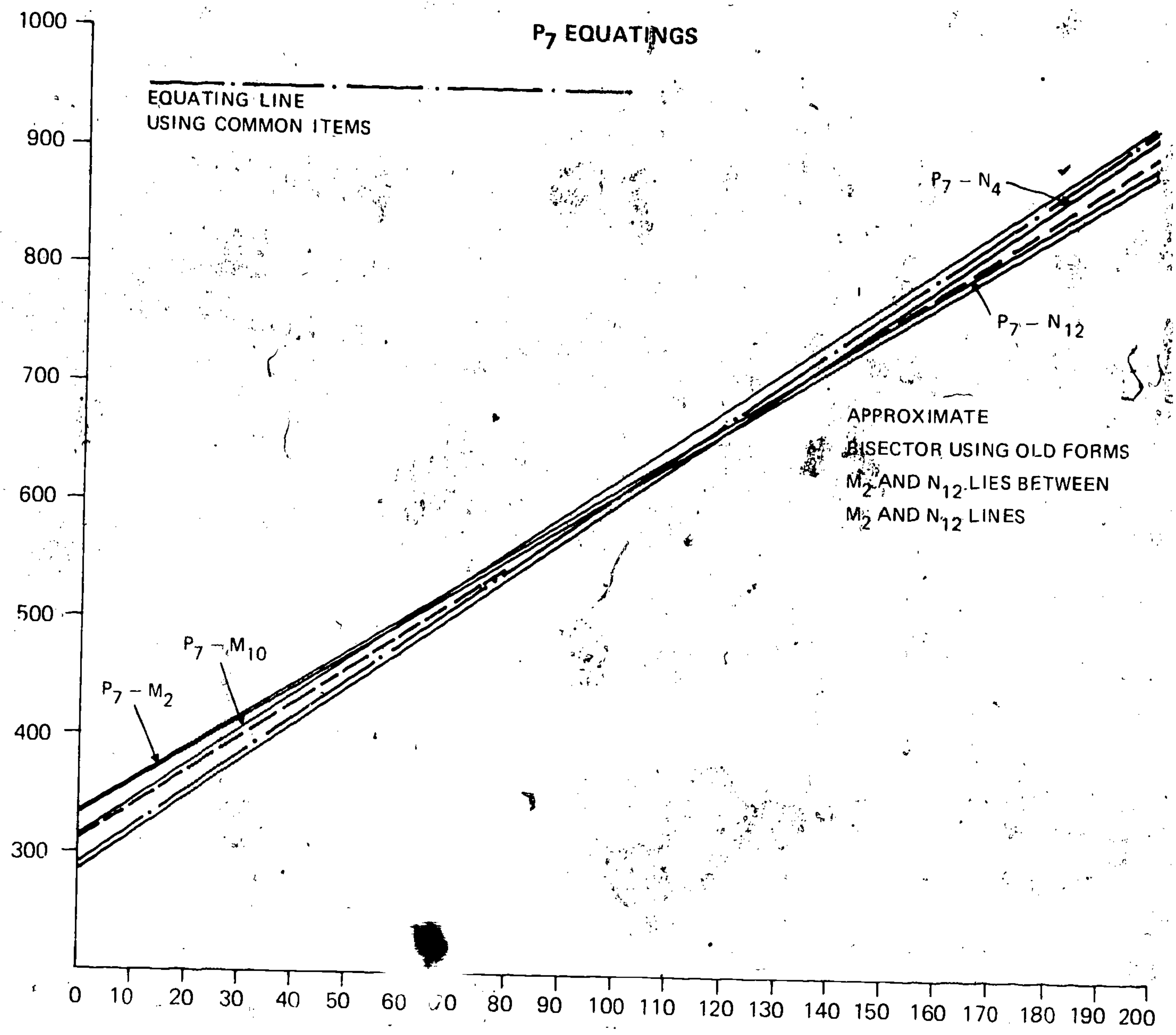


FRENCH

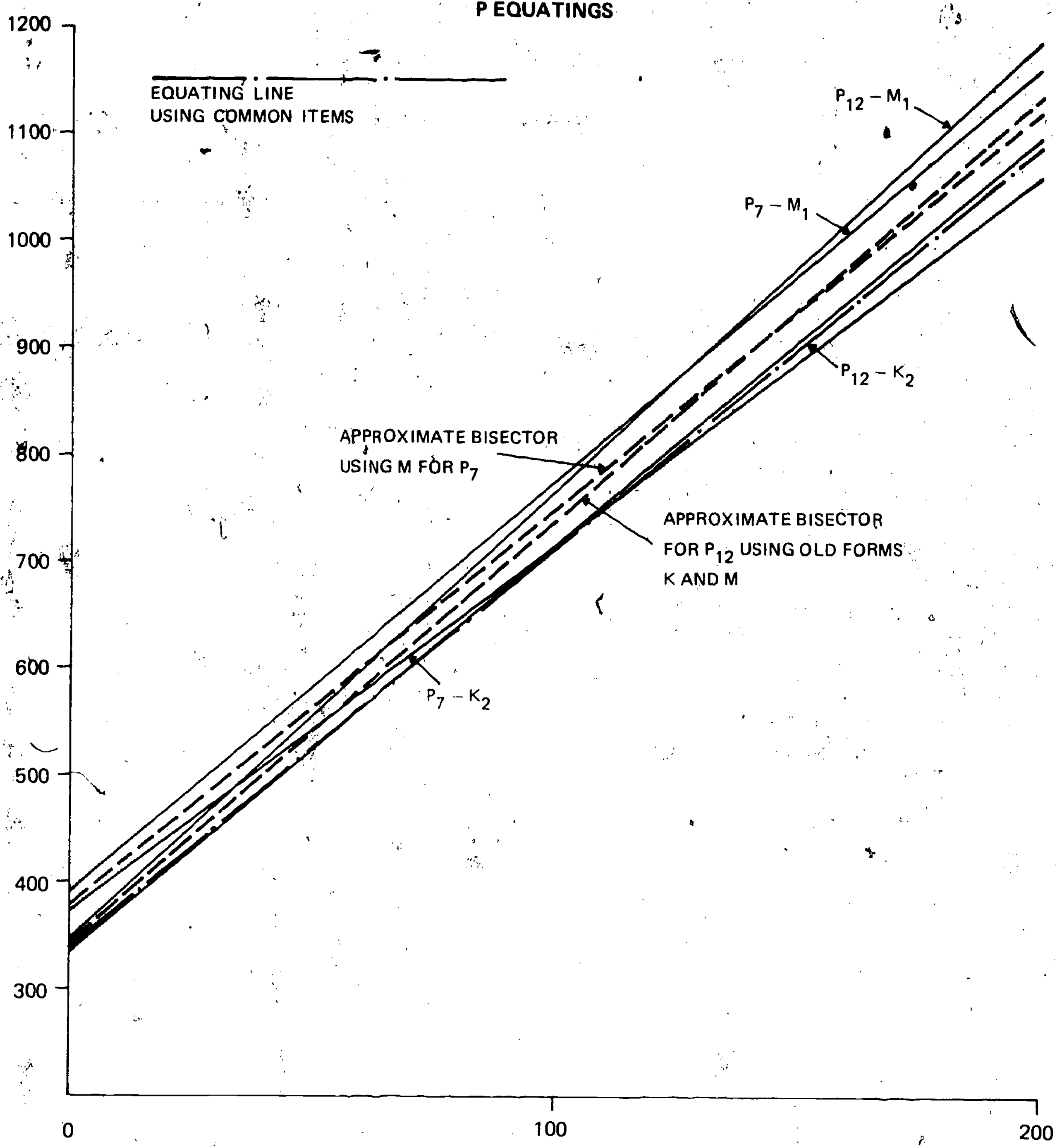


P_1 EQUATINGS
EQUATING LINE USING COMMON ITEMS
APPROXIMATE BISECTOR USING OLD FORMS M_{10} AND N_{12}
APPROXIMATE BISECTOR USING OLD FORMS M_{10} AND N_4
$P_1 - M_{10}$
$P_1 - N_{12}$
$P_1 - N_4$
$P_1 - M_2$ LIES APPROXIMATELY ON THE COMMON ITEM LINE. THE BISECTORS OF OLD FORMS M_2 AND N_{12} AND N_9 ALSO ARE ON THE COMMON ITEM EQUATING LINE
1000
900
800
700
600
500
400
300
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200

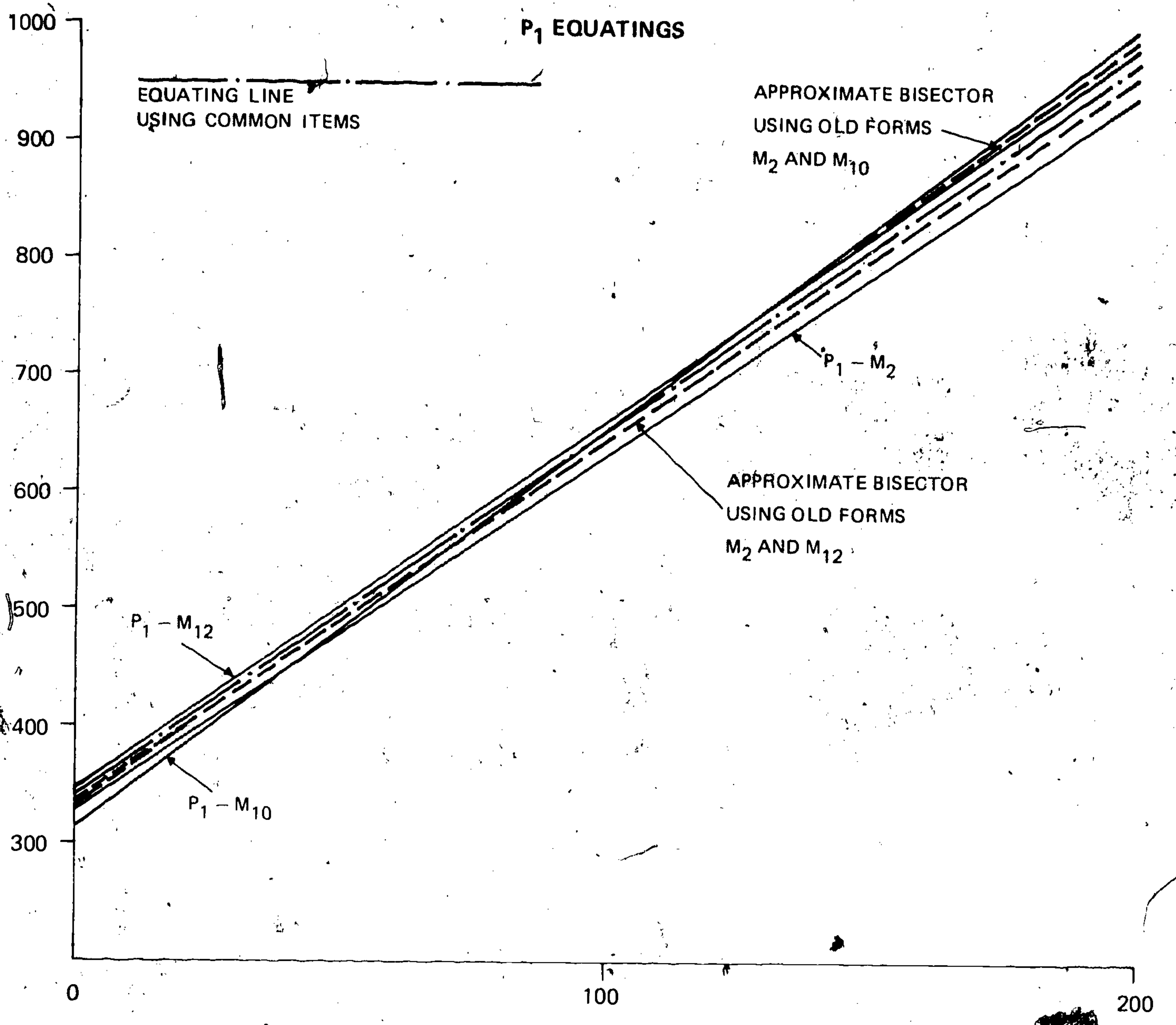
FRENCH


P$_7$ EQUATINGS
EQUATING LINE
USING COMMON ITEMS
P$_7$ – N$_4$
P$_7$ – N$_{12}$
APPROXIMATE
BISECTOR USING OLD FORMS
M$_2$ AND N$_{12}$ LIES BETWEEN
M$_2$ AND N$_{12}$ LINES
P$_7$ – M$_{10}$
P$_7$ – M$_2$
1000
900
800
700
600
500
400
300
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200

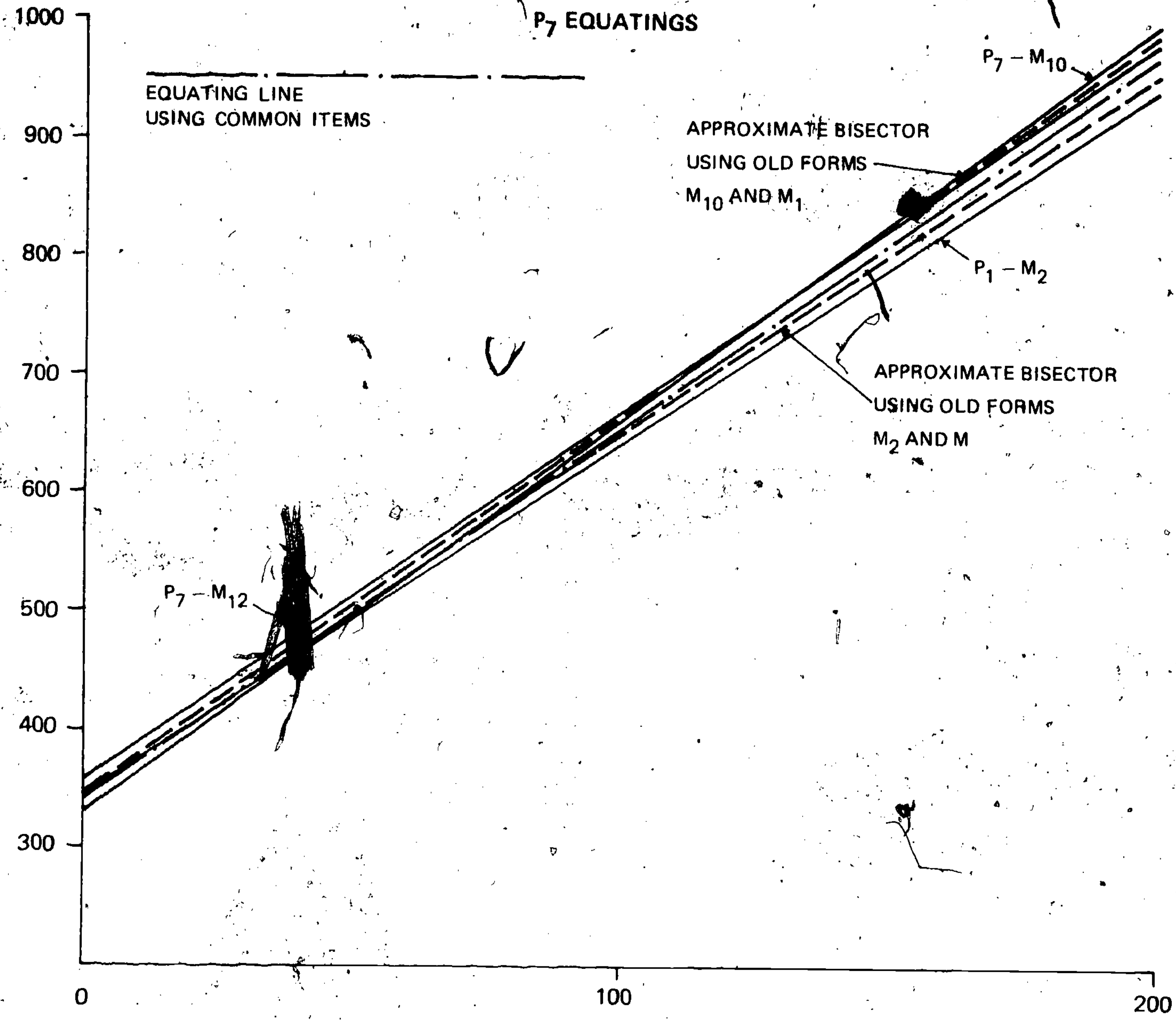
GEOLOGY


P EQUATINGS
EQUATING LINE
USING COMMON ITEMS
P12 – M1
P7 – M1
P12 – K2
P7 – K2
APPROXIMATE BISECTOR
USING M FOR P7
APPROXIMATE BISECTOR
FOR P12 USING OLD FORMS
K AND M
1200
1100
1000
900
800
700
600
500
400
300
0
100
200

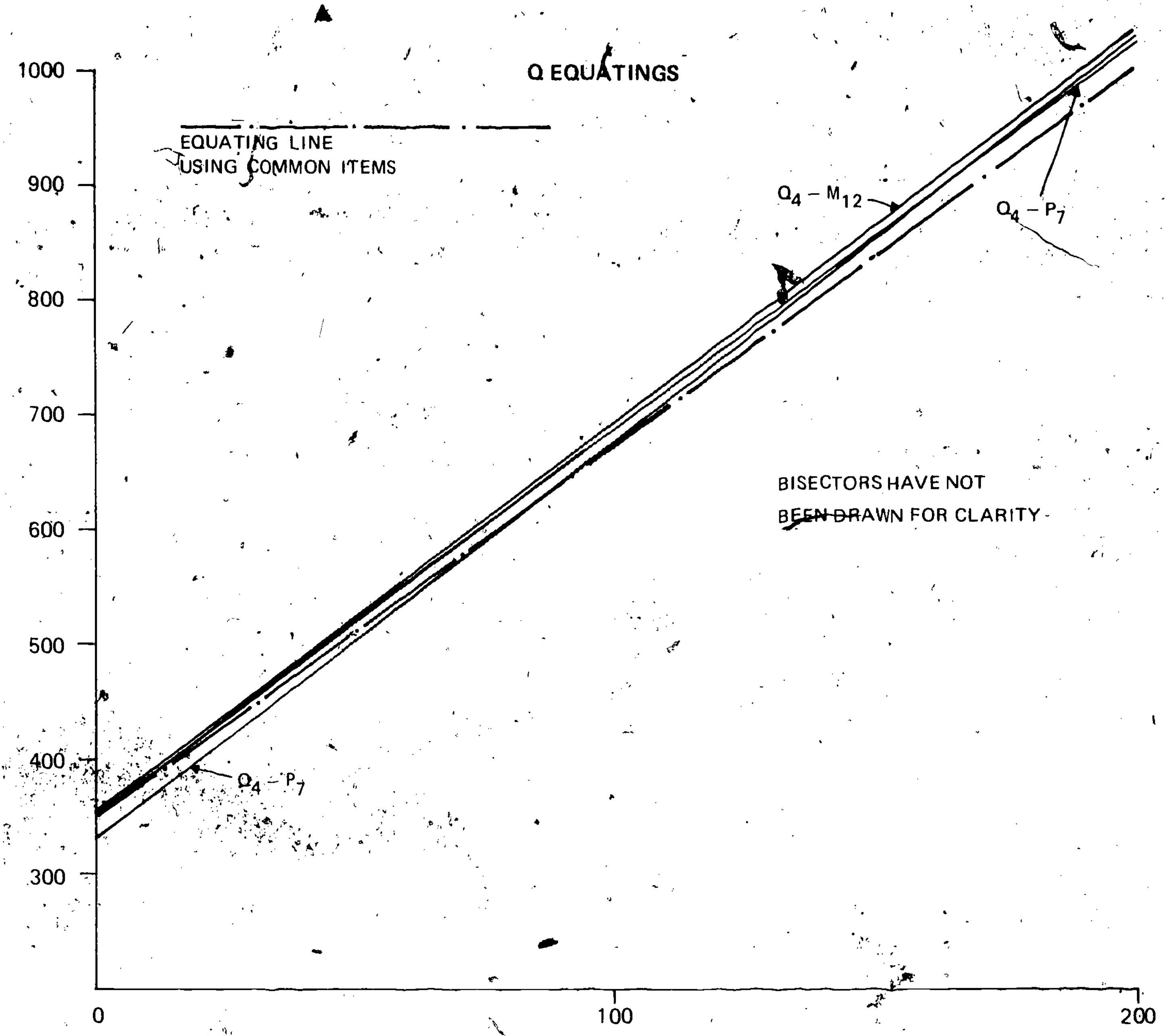
HISTORY


P₁ EQUATINGS
EQUATING LINE
USING COMMON ITEMS
APPROXIMATE BISECTOR
USING OLD FORMS
M_2 AND M_{10}
$P_1 - M_2$
APPROXIMATE BISECTOR
USING OLD FORMS
M_2 AND M_{12}
$P_1 - M_{12}$
$P_1 - M_{10}$
1000
900
800
700
600
500
400
300
0
100
200

HISTORY

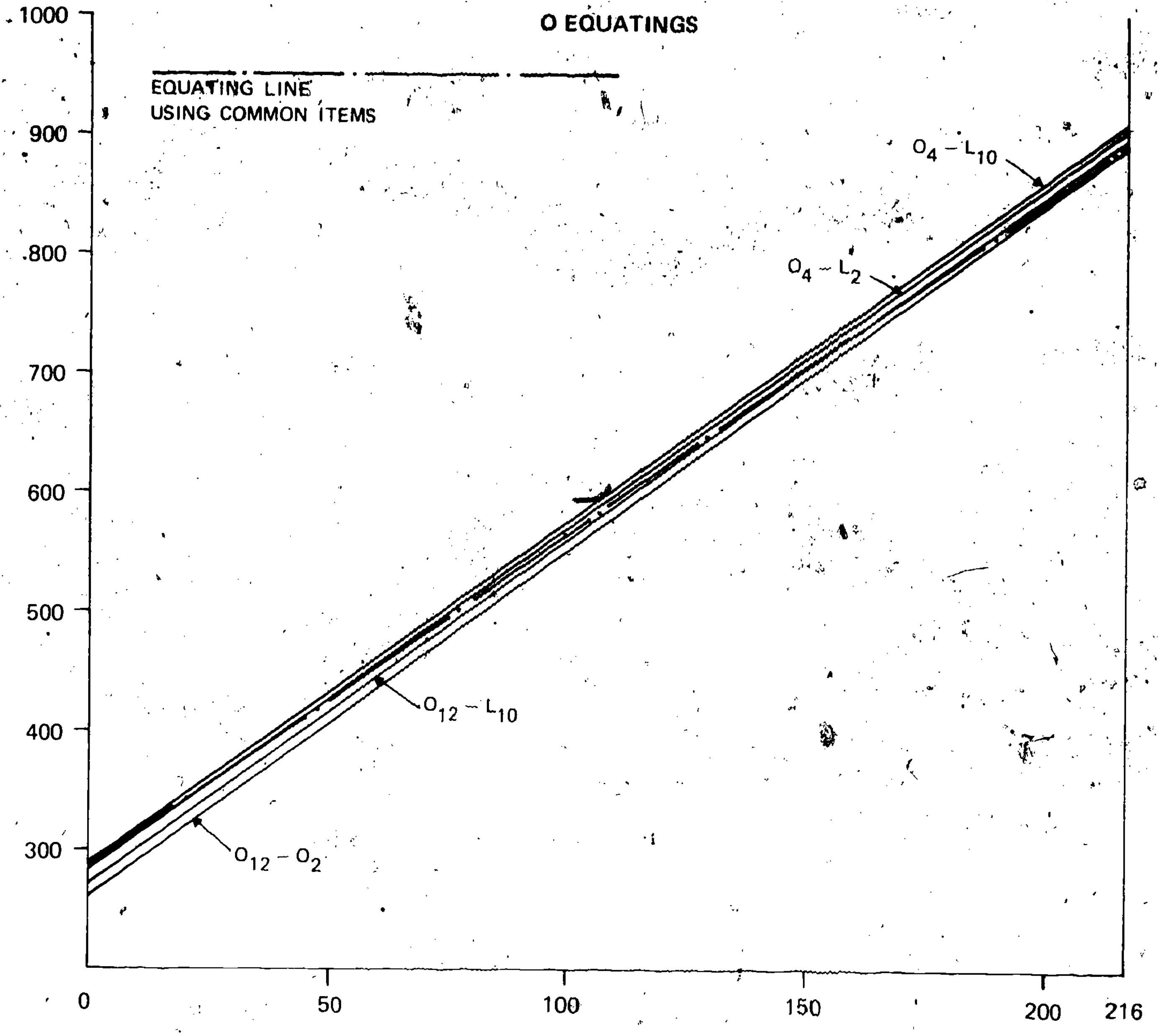

P7 EQUATINGS
EQUATING LINE
USING COMMON ITEMS
APPROXIMATE BISECTOR
USING OLD FORMS
M10 AND M1
P7 – M10
P1 – M2
APPROXIMATE BISECTOR
USING OLD FORMS
M2 AND M
P7 – M12
1000
900
800
700
600
500
400
300
0
100
200

HISTORY

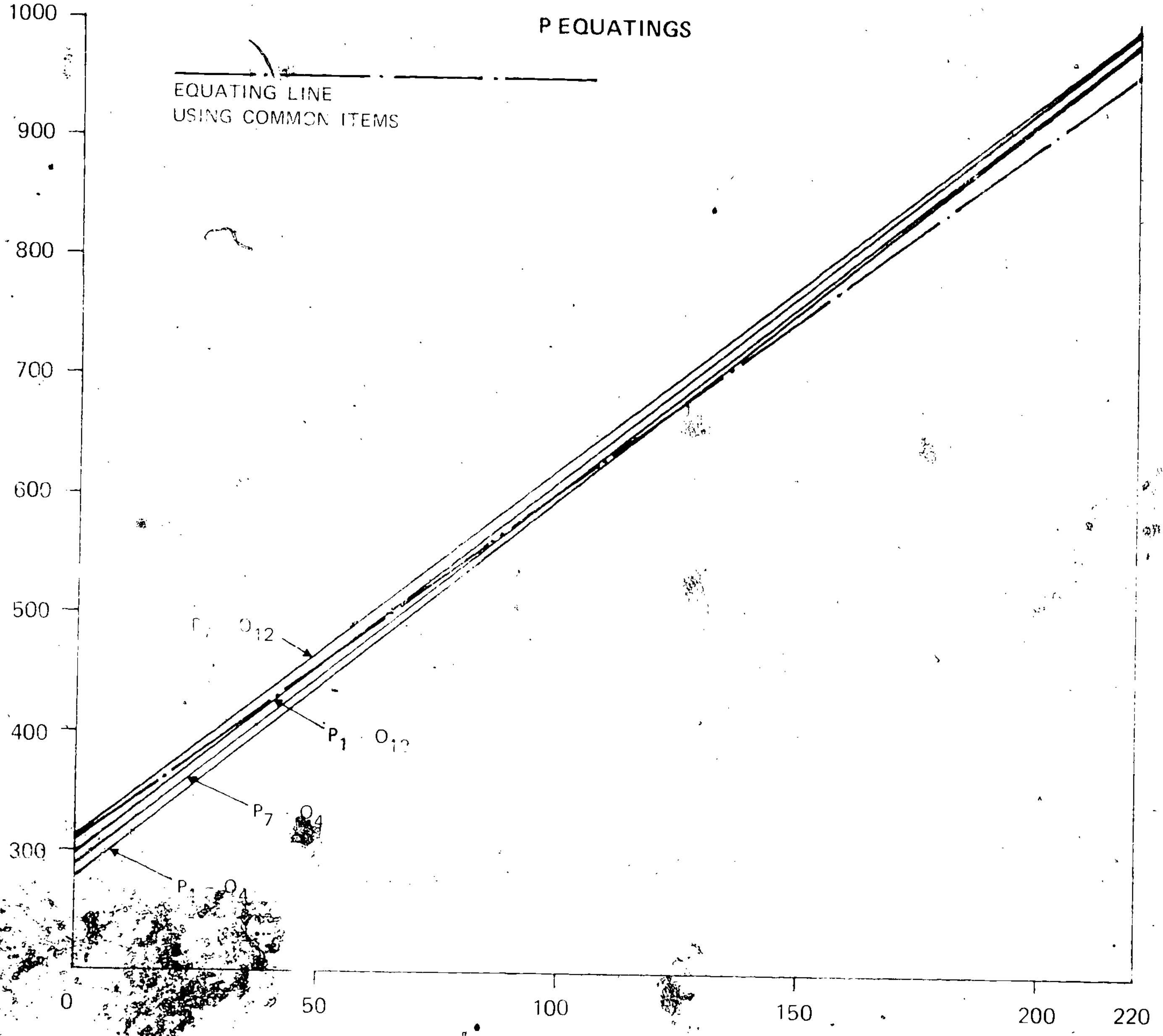


BISECTORS HAVE NOT BEEN DRAWN FOR CLARITY

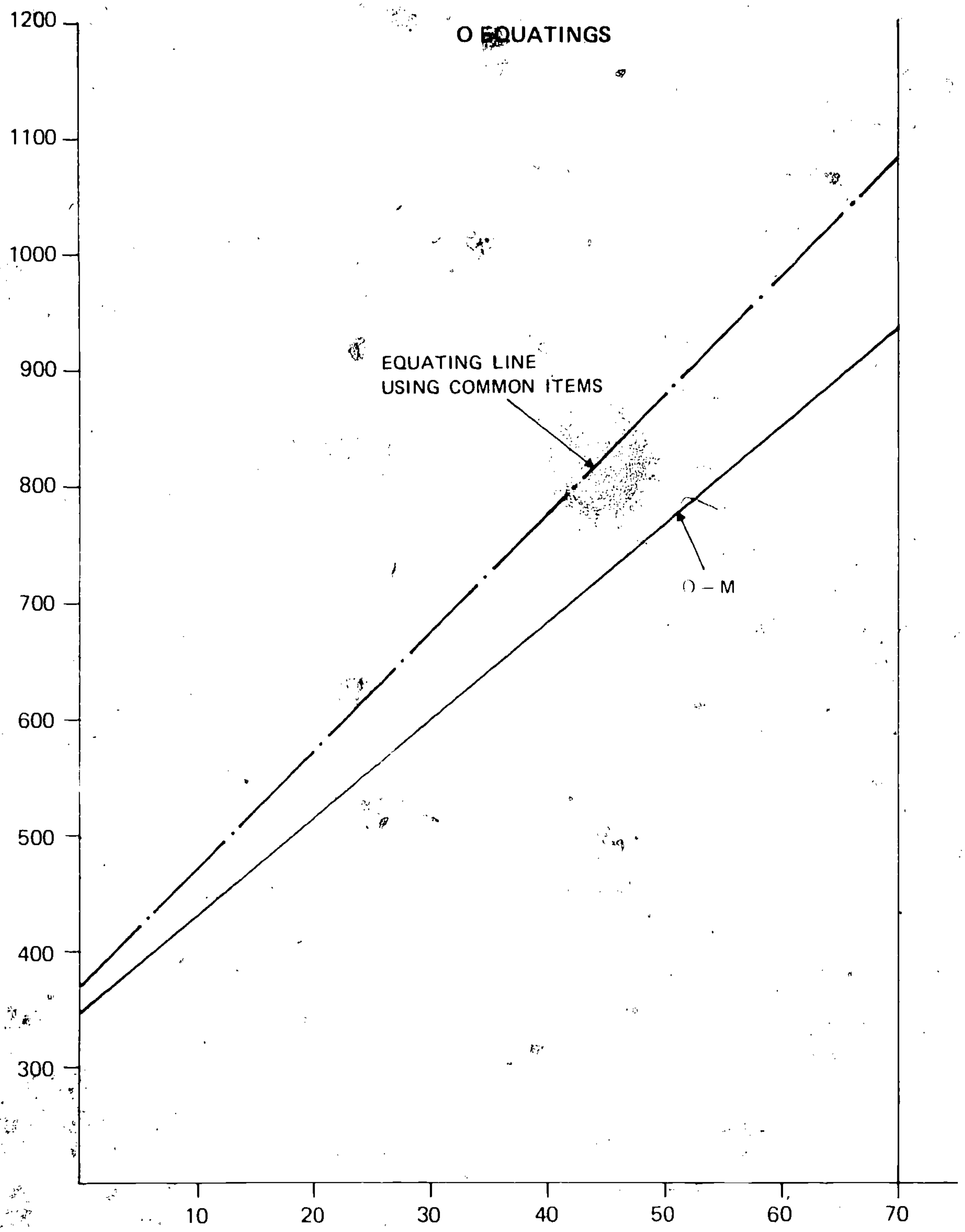
LITERATURE


O EQUATINGS
EQUATING LINE
USING COMMON ITEMS
$O_4 - L_{10}$
$O_4 - L_2$
$O_{12} - L_{10}$
$O_{12} - O_2$
1000
900
800
700
600
500
400
300
0
50
100
150
200
216

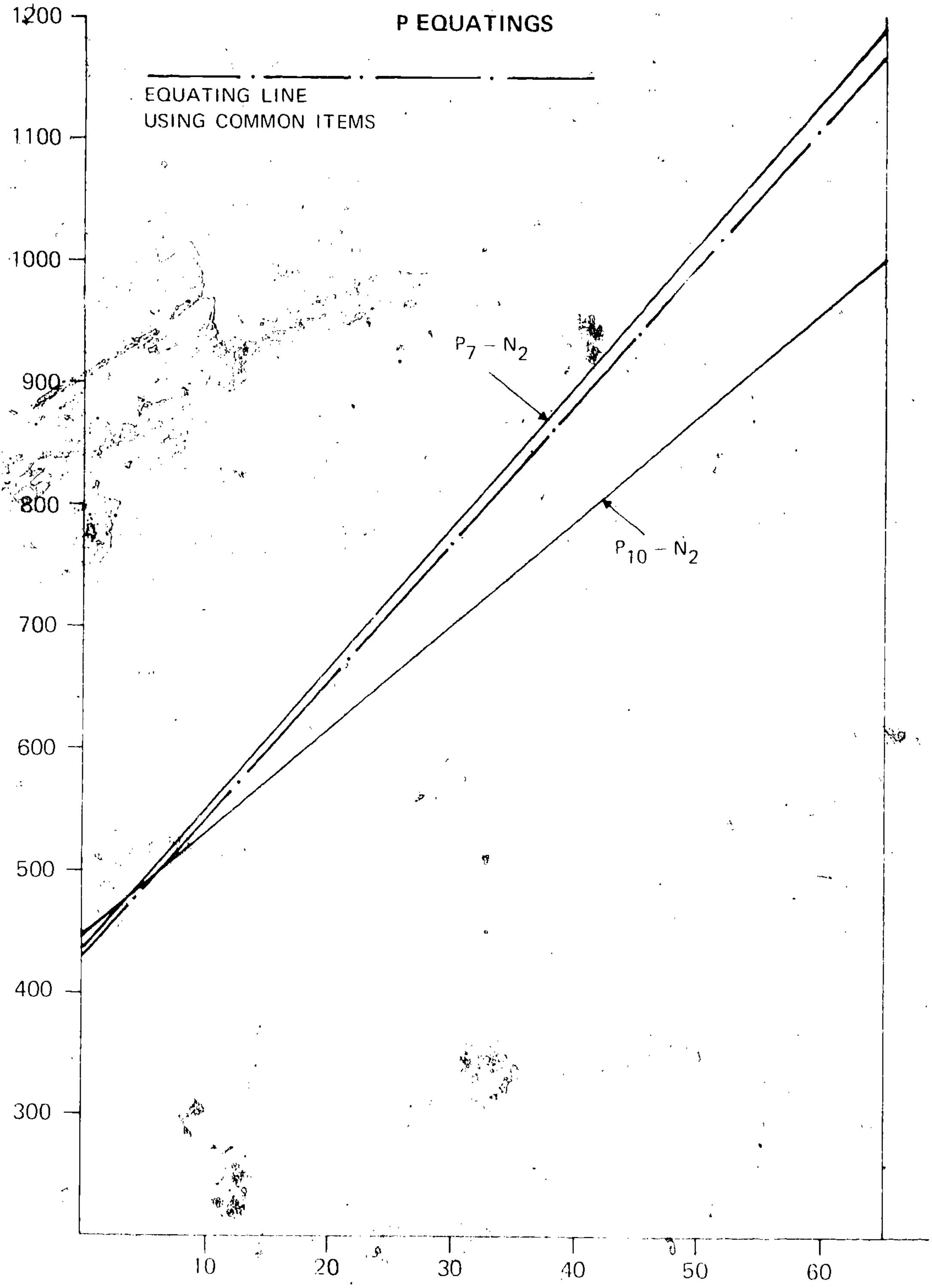
LITERATURE


P EQUATINGS
EQUATING LINE
USING COMMON ITEMS
P_7 O_{12}
P_1 O_{12}
P_7 O_4
P_1 – O_4
1000
900
800
700
600
500
400
300
0
50
100
150
200
220

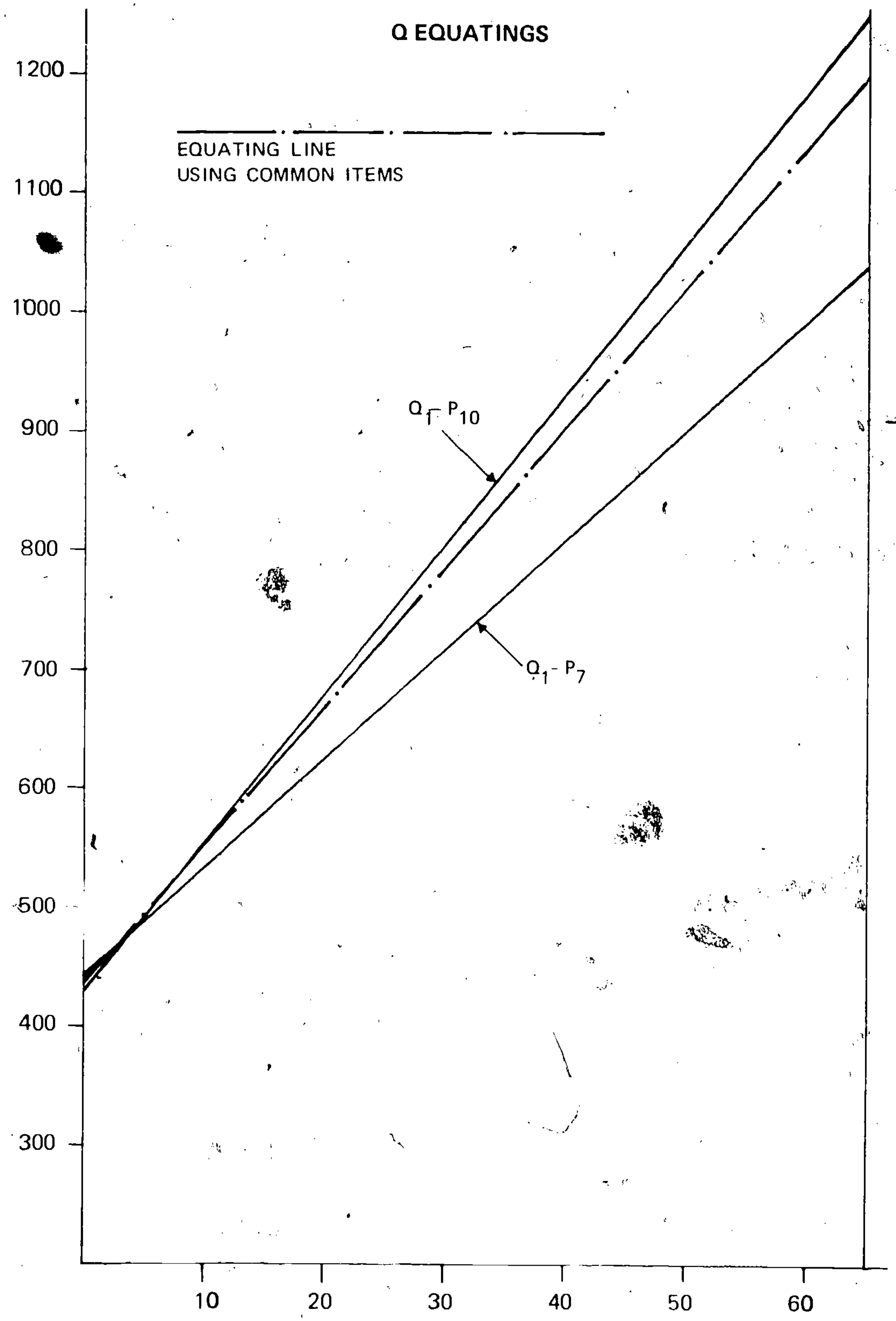
MATHEMATICS



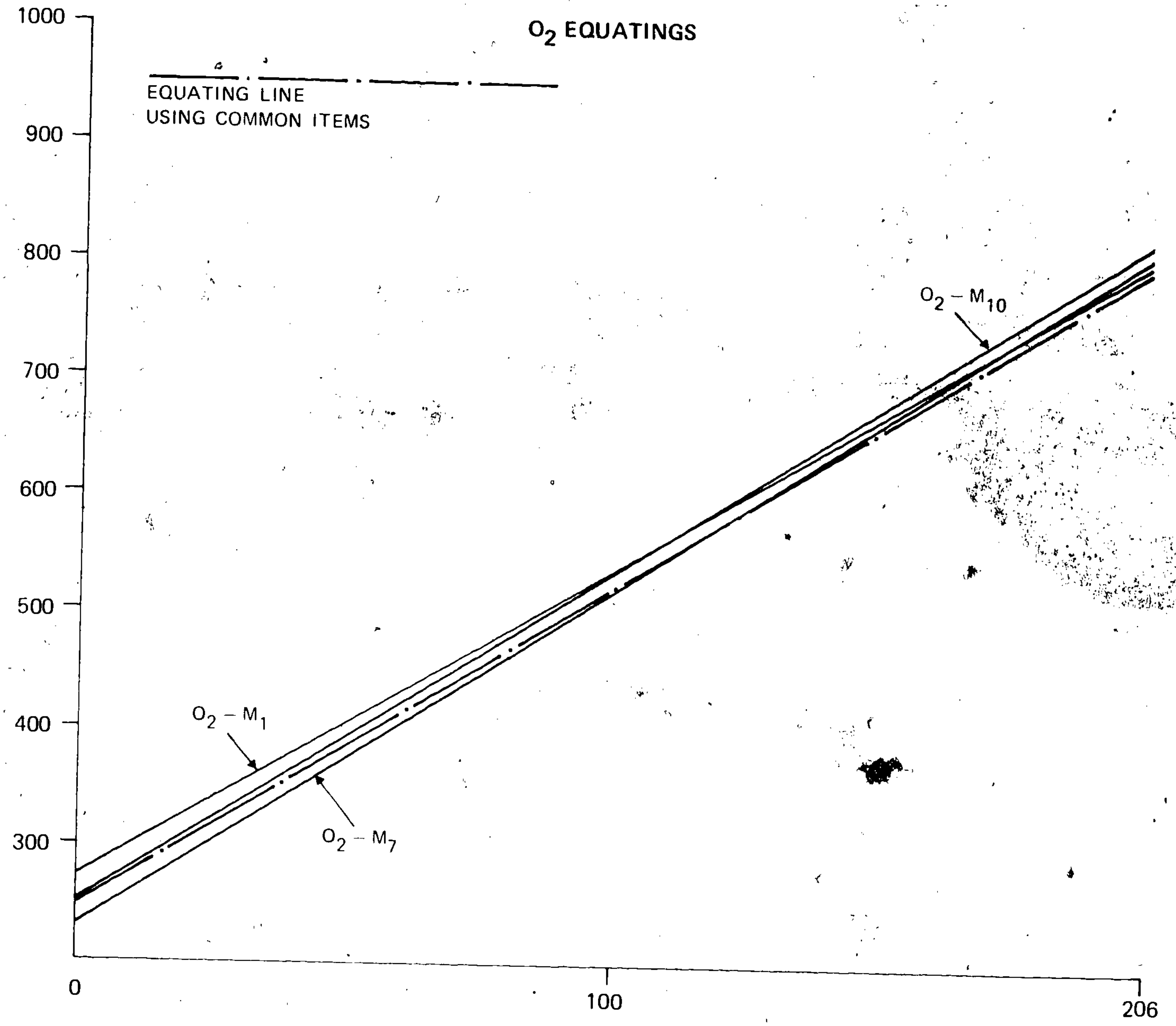
MATHEMATICS



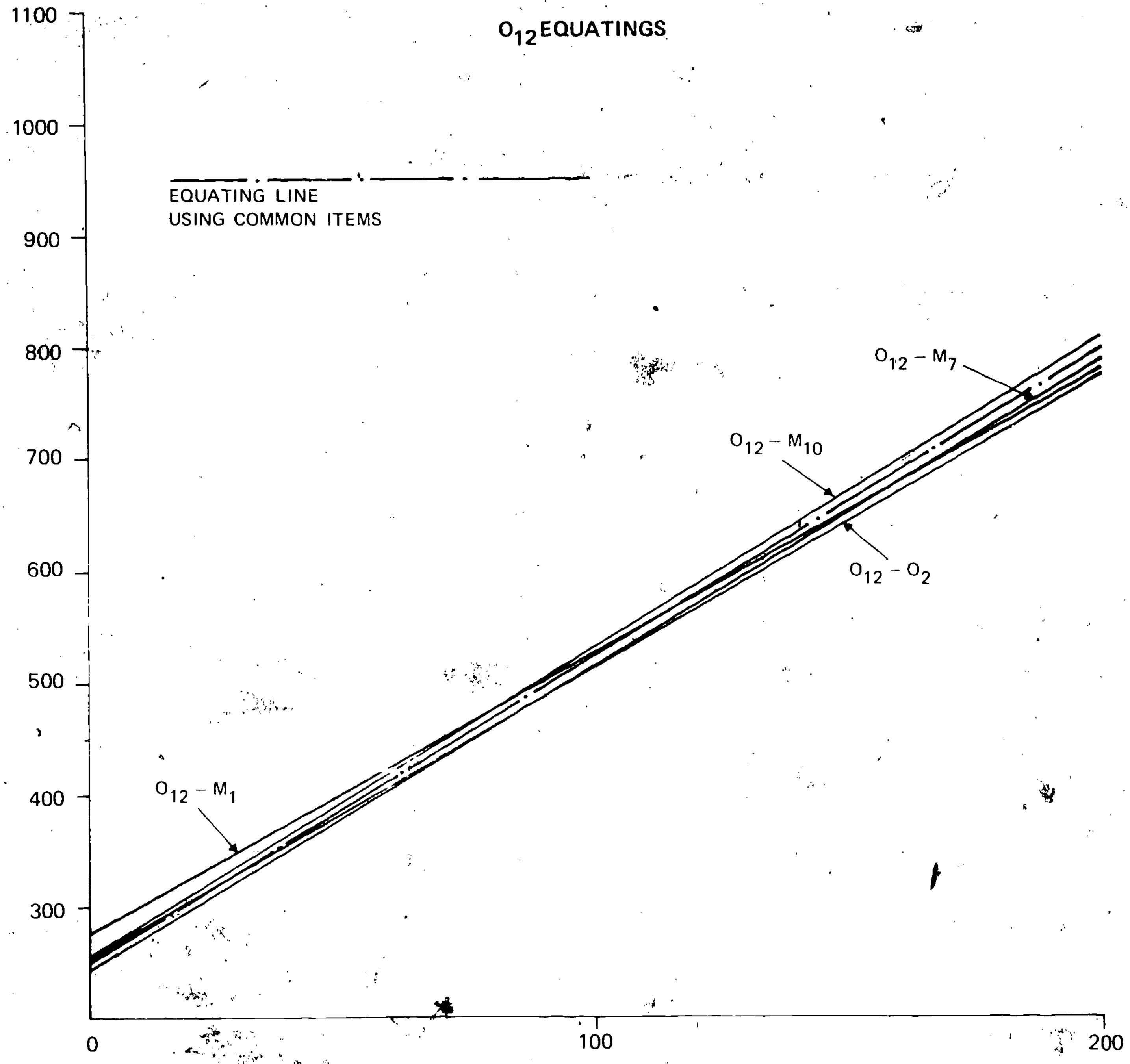
MATHEMATICS



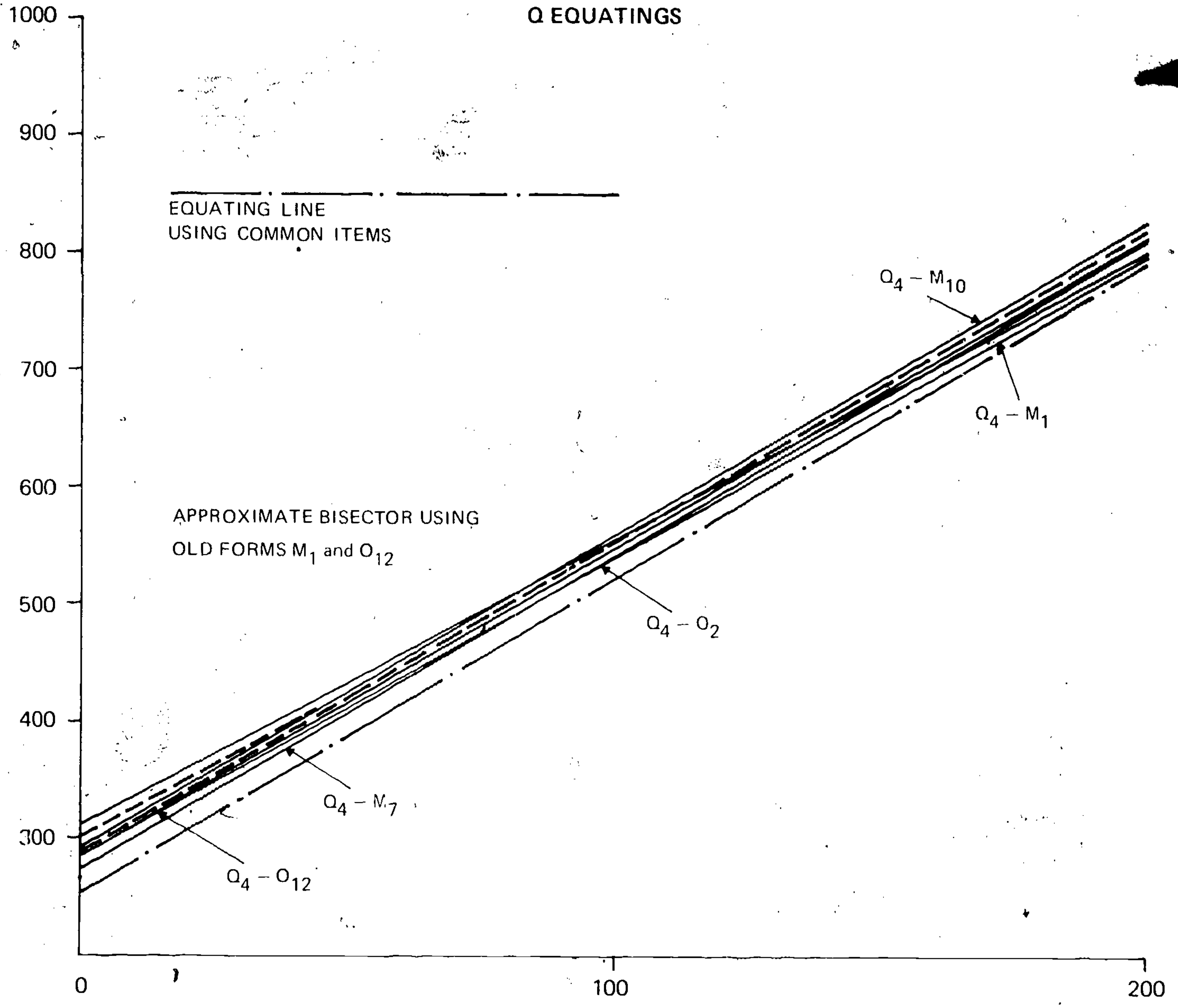
MUSIC


O_2 EQUATINGS
EQUATING LINE
USING COMMON ITEMS
$O_2 - M_{10}$
$O_2 - M_1$
$O_2 - M_7$
1000
900
800
700
600
500
400
300
0
100
206

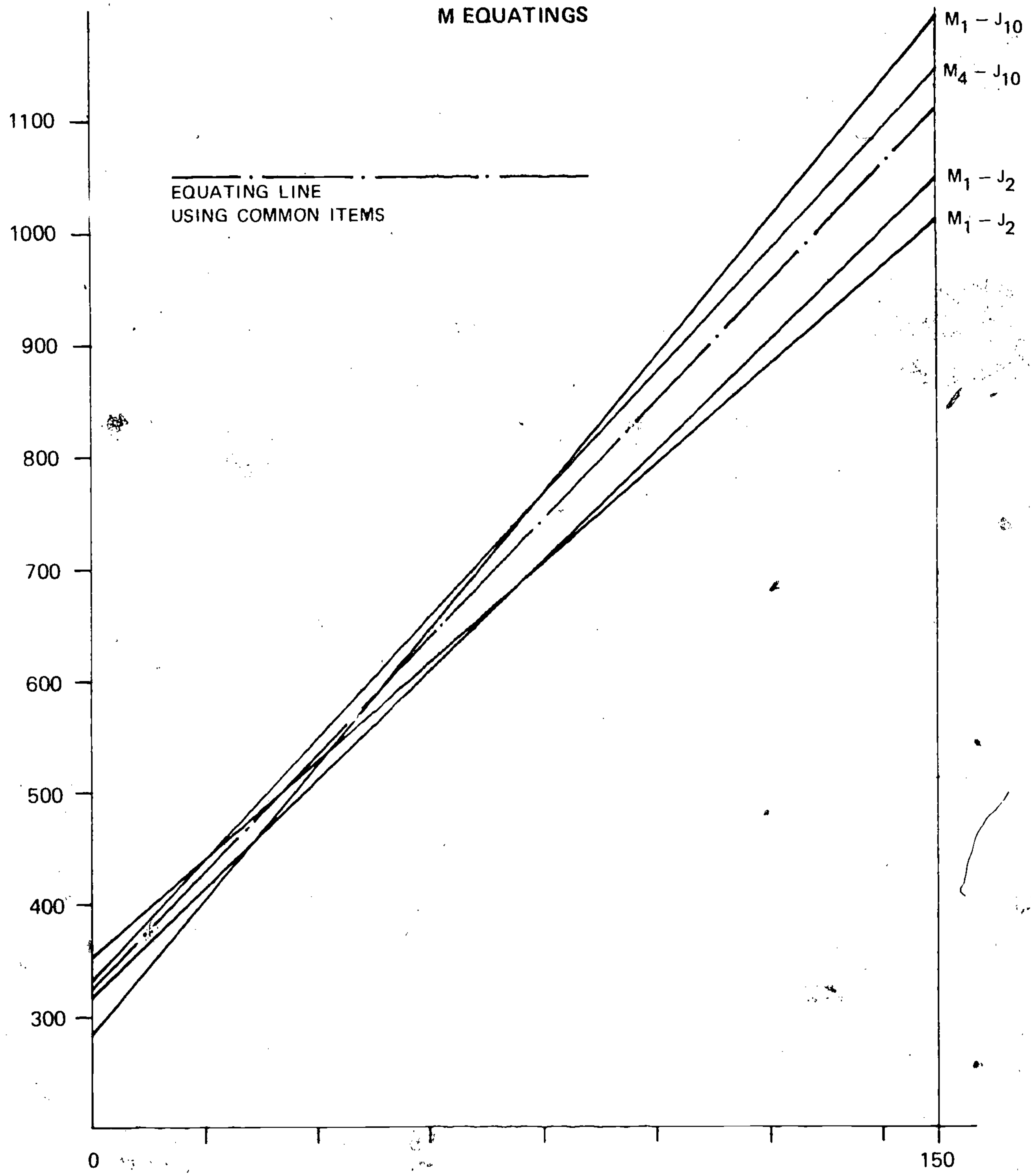
MUSIC


O_{12} EQUATINGS
EQUATING LINE
USING COMMON ITEMS
$O_{12} - M_7$
$O_{12} - M_{10}$
$O_{12} - O_2$
$O_{12} - M_1$
1100
1000
900
800
700
600
500
400
300
0
100
200

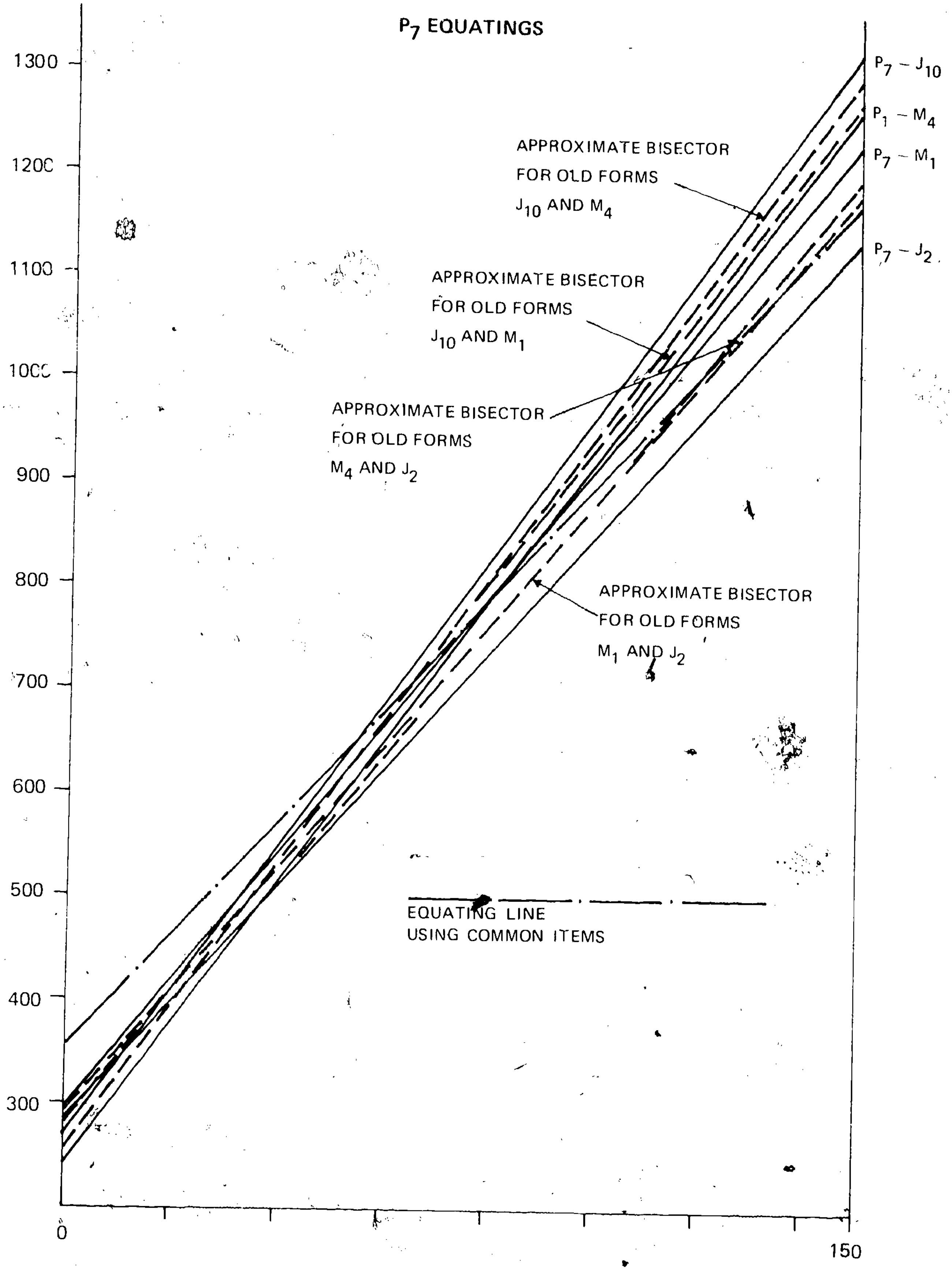
MUSIC


Q EQUATINGS
EQUATING LINE
USING COMMON ITEMS
APPROXIMATE BISECTOR USING
OLD FORMS M_1 and O_{12}
$Q_4 - M_{10}$
$Q_4 - M_1$
$Q_4 - O_2$
$Q_4 - M_7$
$Q_4 - O_{12}$
1000
900
800
700
600
500
400
300
0
100
200

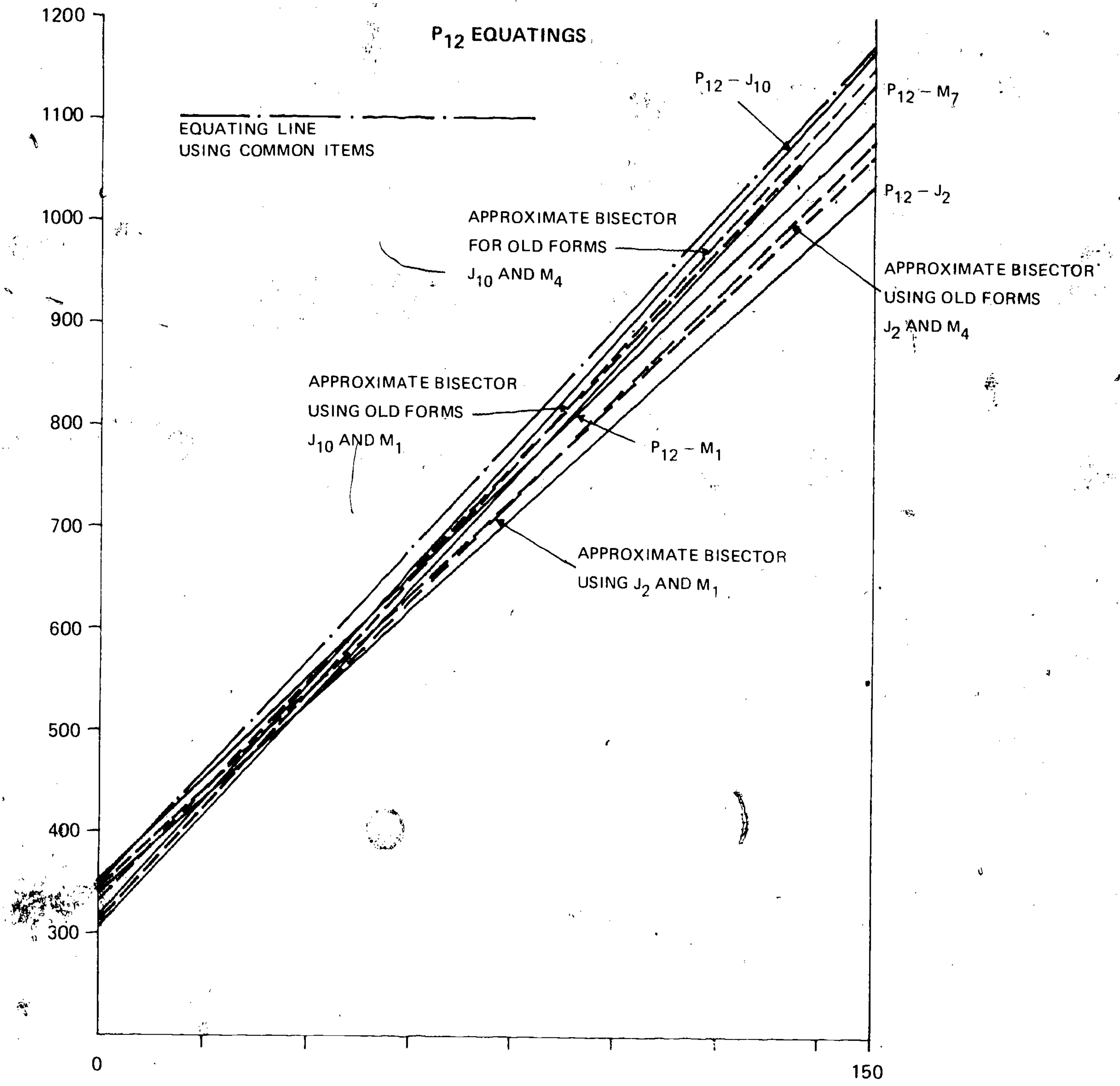
PHILOSOPHY


M EQUATINGS
EQUATING LINE
USING COMMON ITEMS
$M_1 - J_{10}$
$M_4 - J_{10}$
$M_1 - J_2$
$M_1 - J_2$
1100
1000
900
800
700
600
500
400
300
0
150

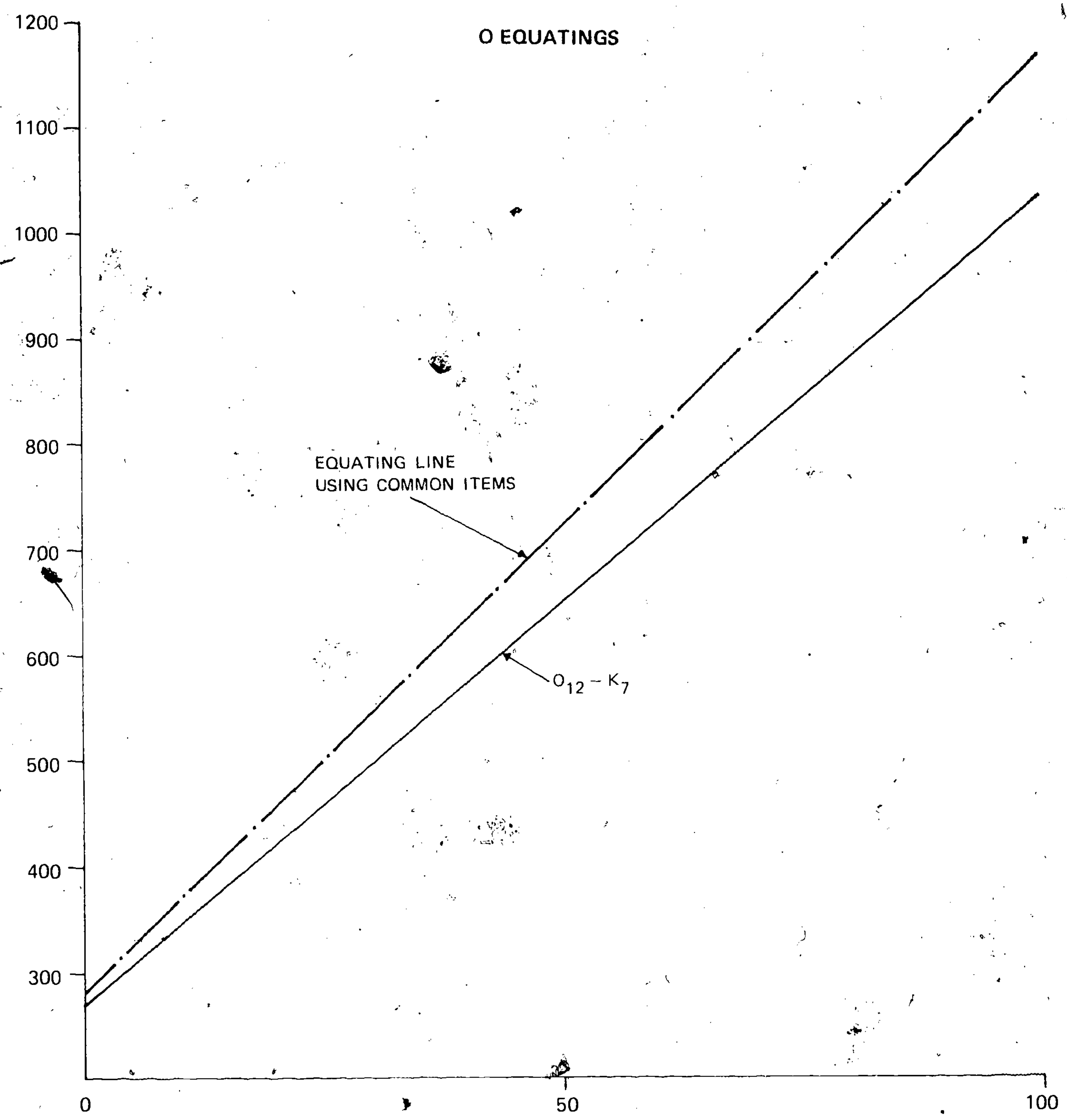
PHILOSOPHY


P7 EQUATINGS
1300
1200
1100
1000
900
800
700
600
500
400
300
0
150
P7 – J10
P1 – M4
P7 – M1
P7 – J2
APPROXIMATE BISECTOR FOR OLD FORMS J10 AND M4
APPROXIMATE BISECTOR FOR OLD FORMS J10 AND M1
APPROXIMATE BISECTOR FOR OLD FORMS M4 AND J2
APPROXIMATE BISECTOR FOR OLD FORMS M1 AND J2
EQUATING LINE USING COMMON ITEMS

PHILOSOPHY

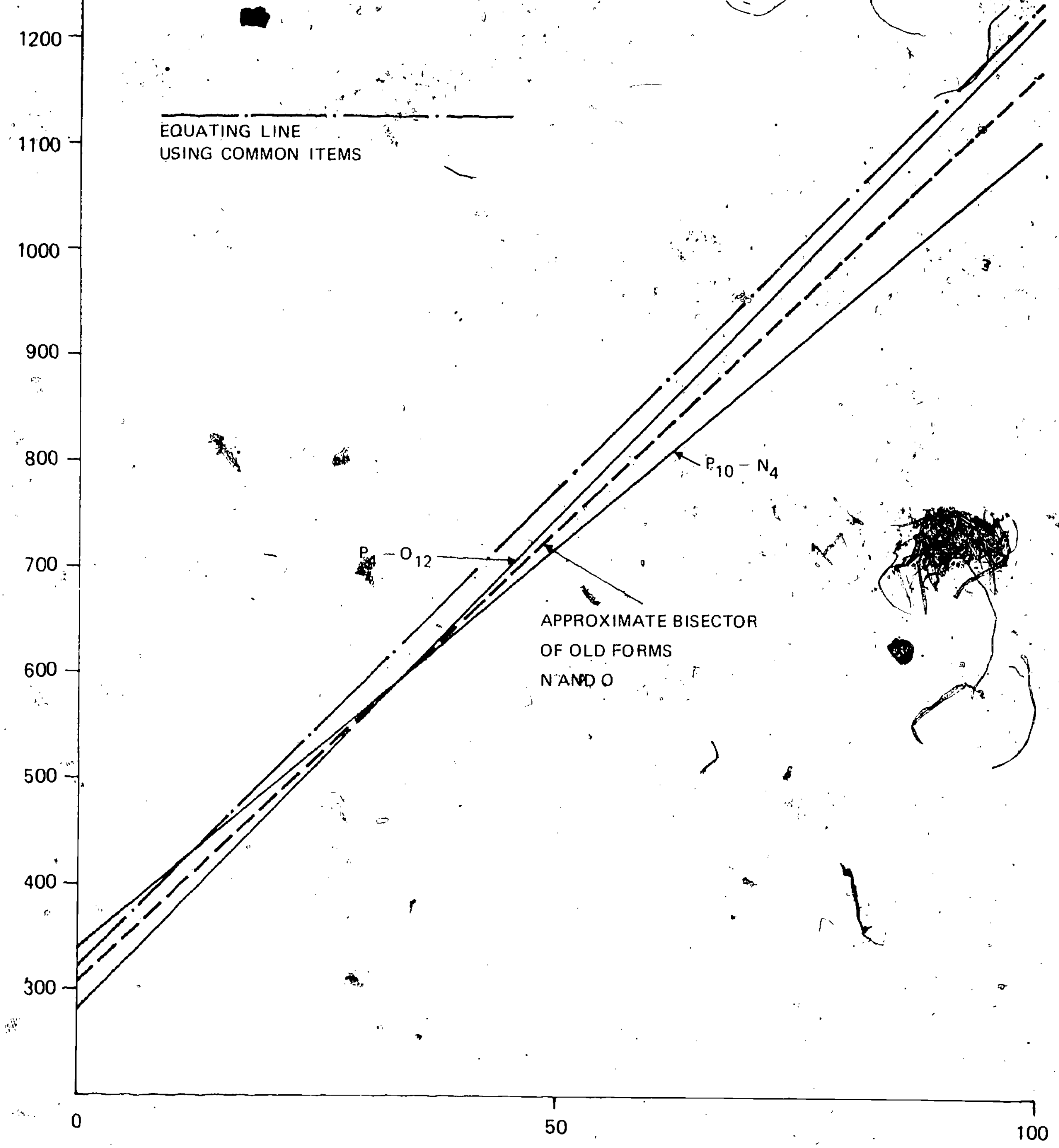

P12 EQUATINGS
EQUATING LINE
USING COMMON ITEMS
P12 – J10
P12 – M7
P12 – J2
APPROXIMATE BISECTOR
FOR OLD FORMS
J10 AND M4
APPROXIMATE BISECTOR
USING OLD FORMS
J2 AND M4
APPROXIMATE BISECTOR
USING OLD FORMS
J10 AND M1
P12 – M1
APPROXIMATE BISECTOR
USING J2 AND M1
1200
1100
1000
900
800
700
600
500
400
300
0
150

PHYSICS

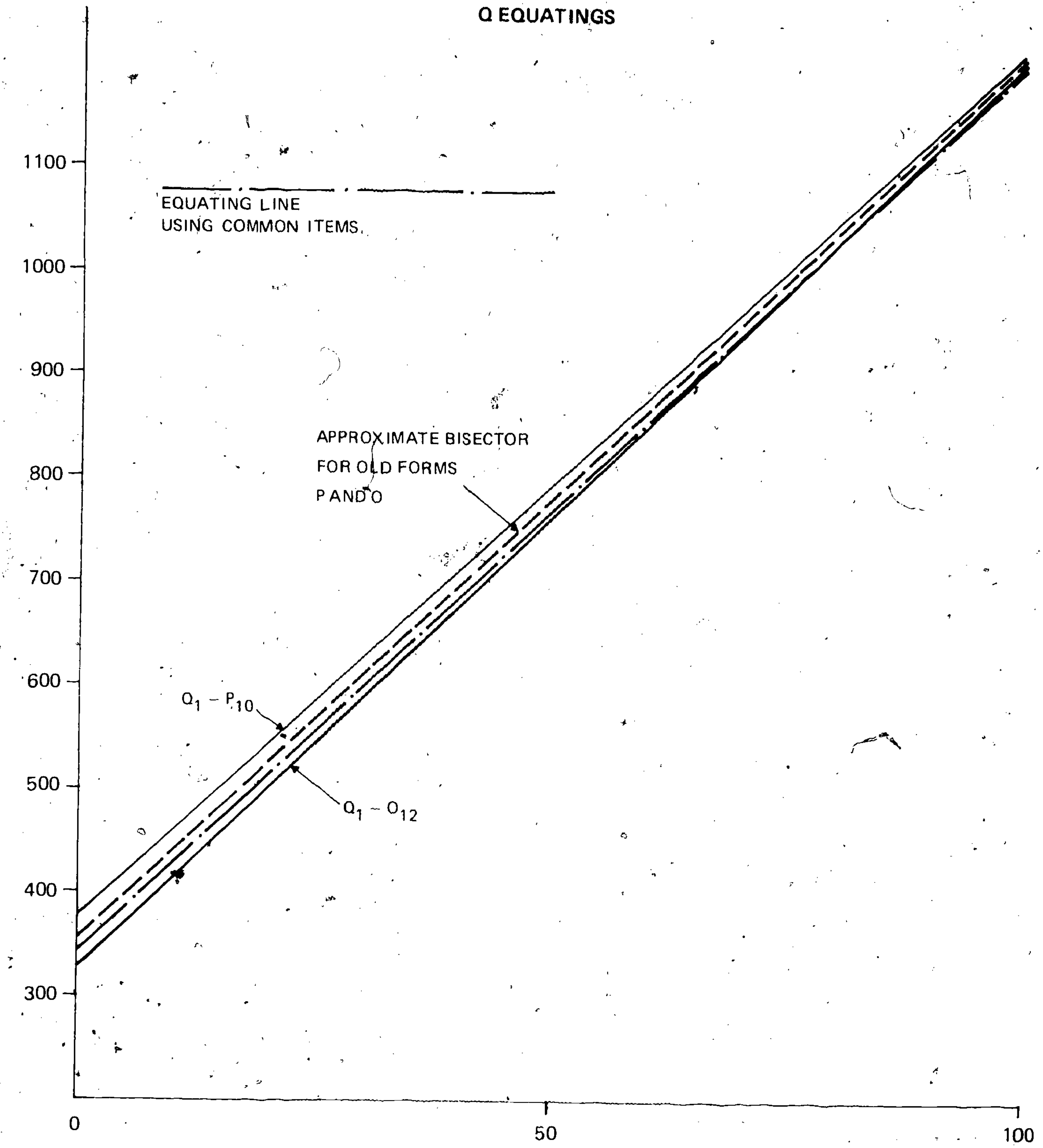

O EQUATINGS
1200
1100
1000
900
800
700
600
500
400
300
0
50
100
EQUATING LINE
USING COMMON ITEMS
$O_{12} - K_7$

PHYSICS

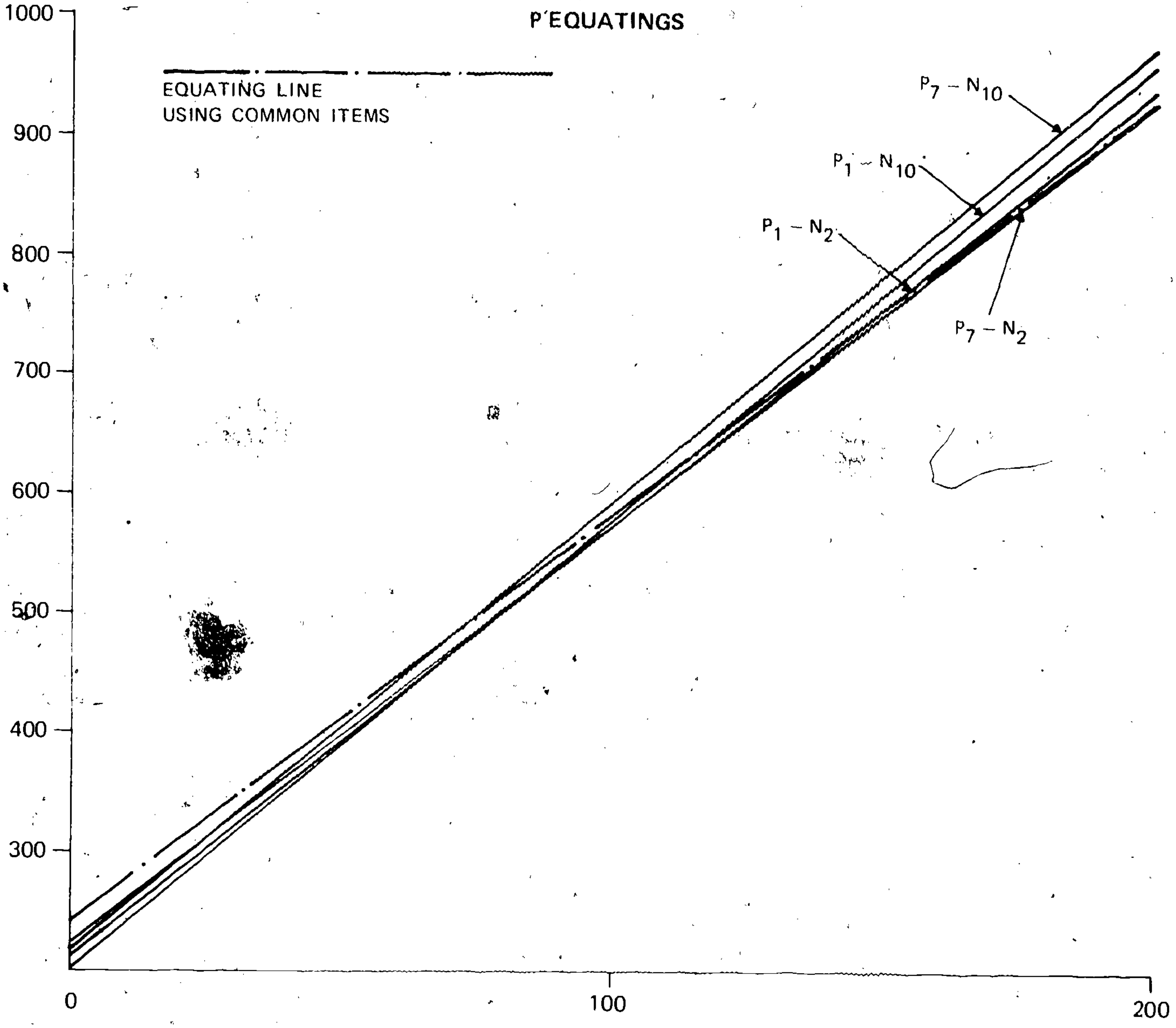
P EQUATINGS



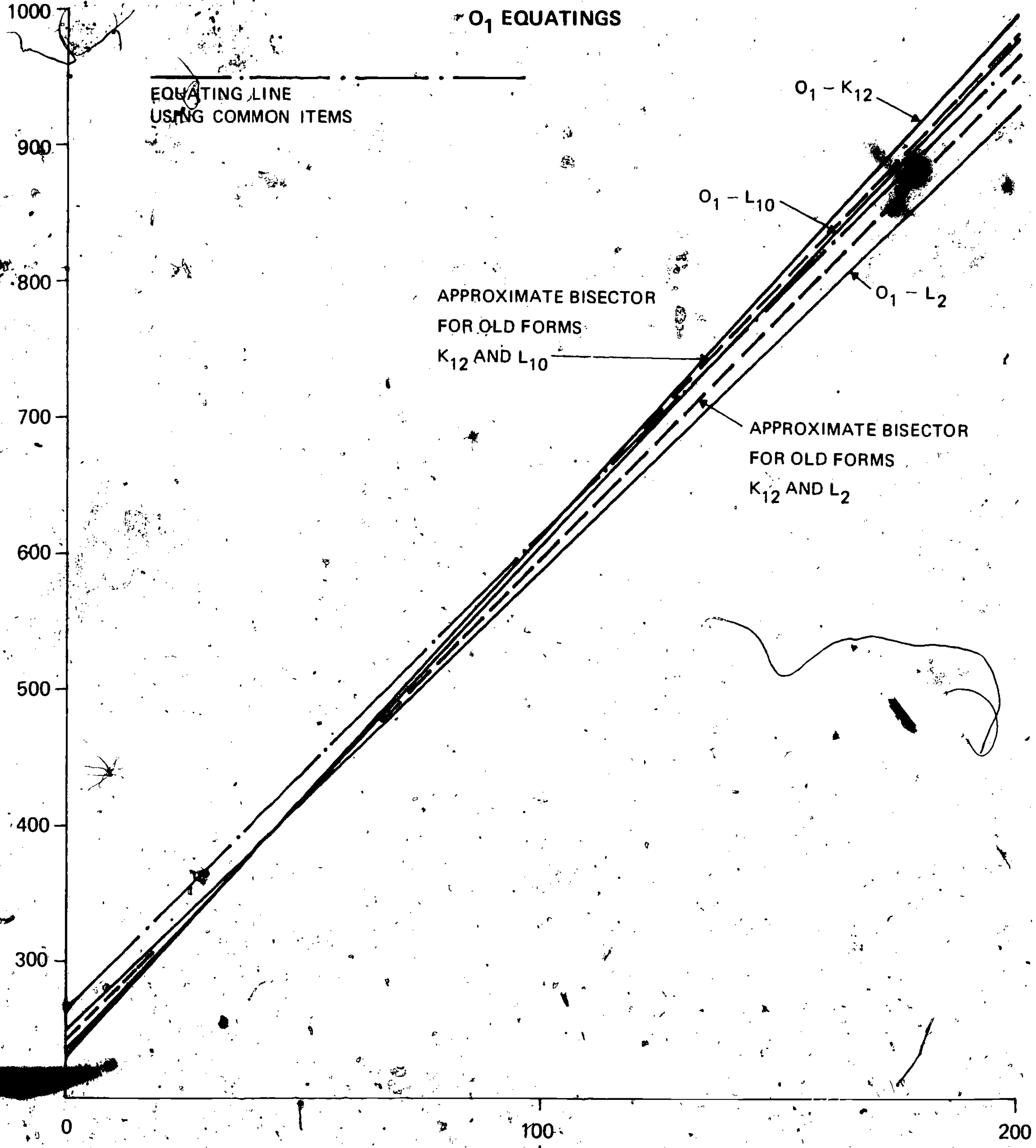
PHYSICS


Q EQUATINGS
EQUATING LINE
USING COMMON ITEMS.
APPROXIMATE BISECTOR
FOR OLD FORMS
P AND O
$Q_1 - P_{10}$
$Q_1 - O_{12}$
1100
1000
900
800
700
600
500
400
300
0
50
100

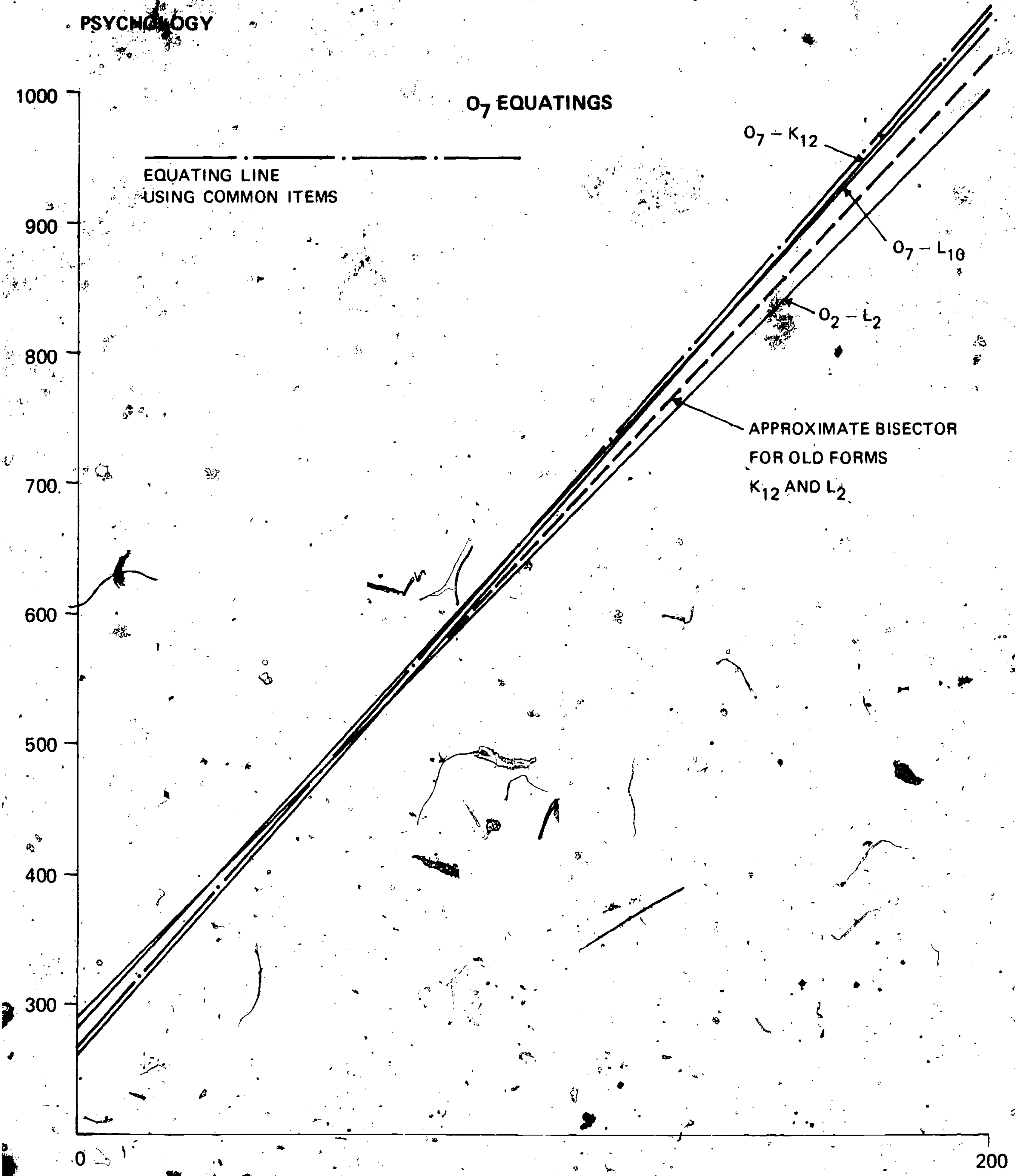
POLITICAL SCIENCE

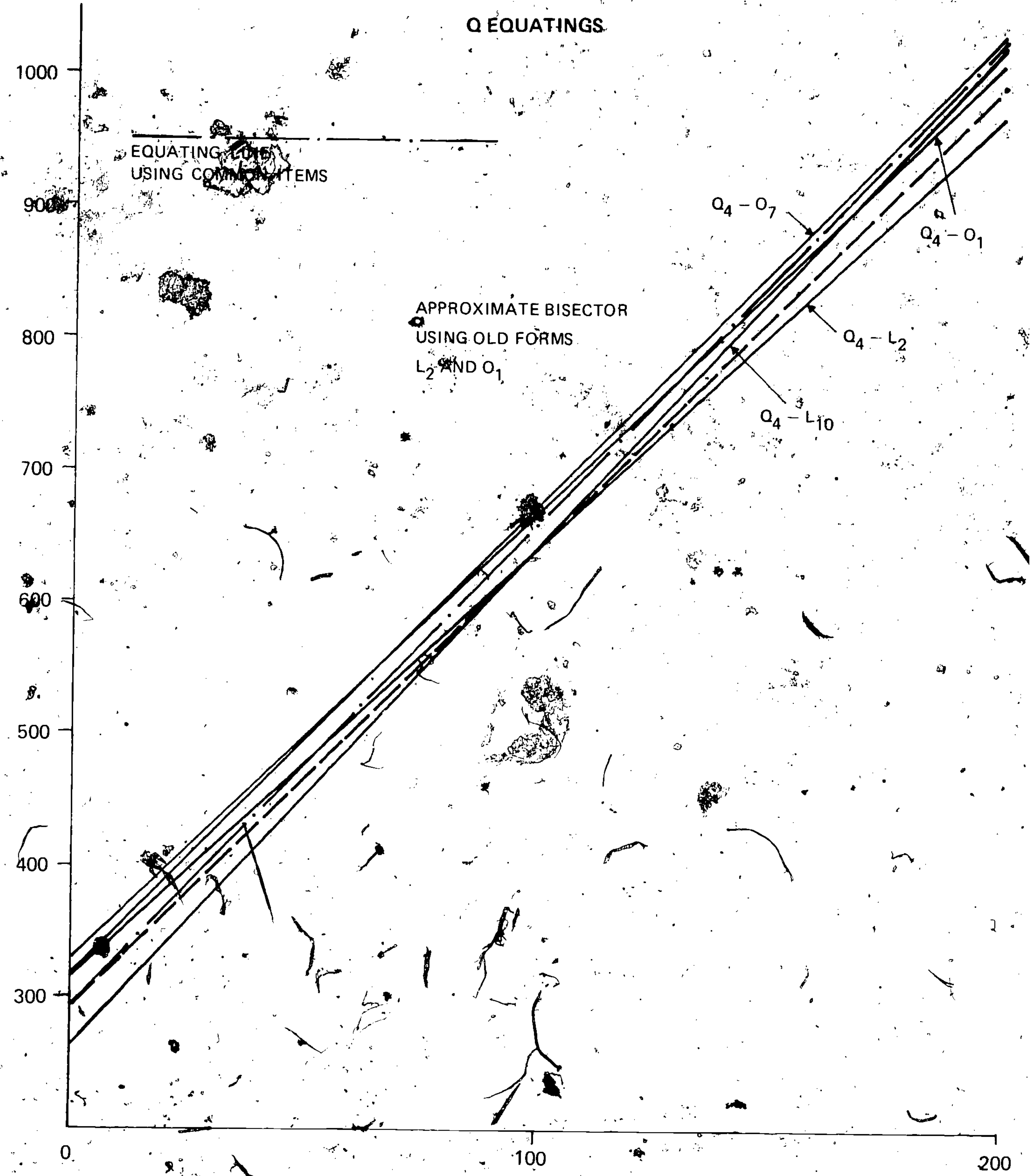


PSYCHOLOGY

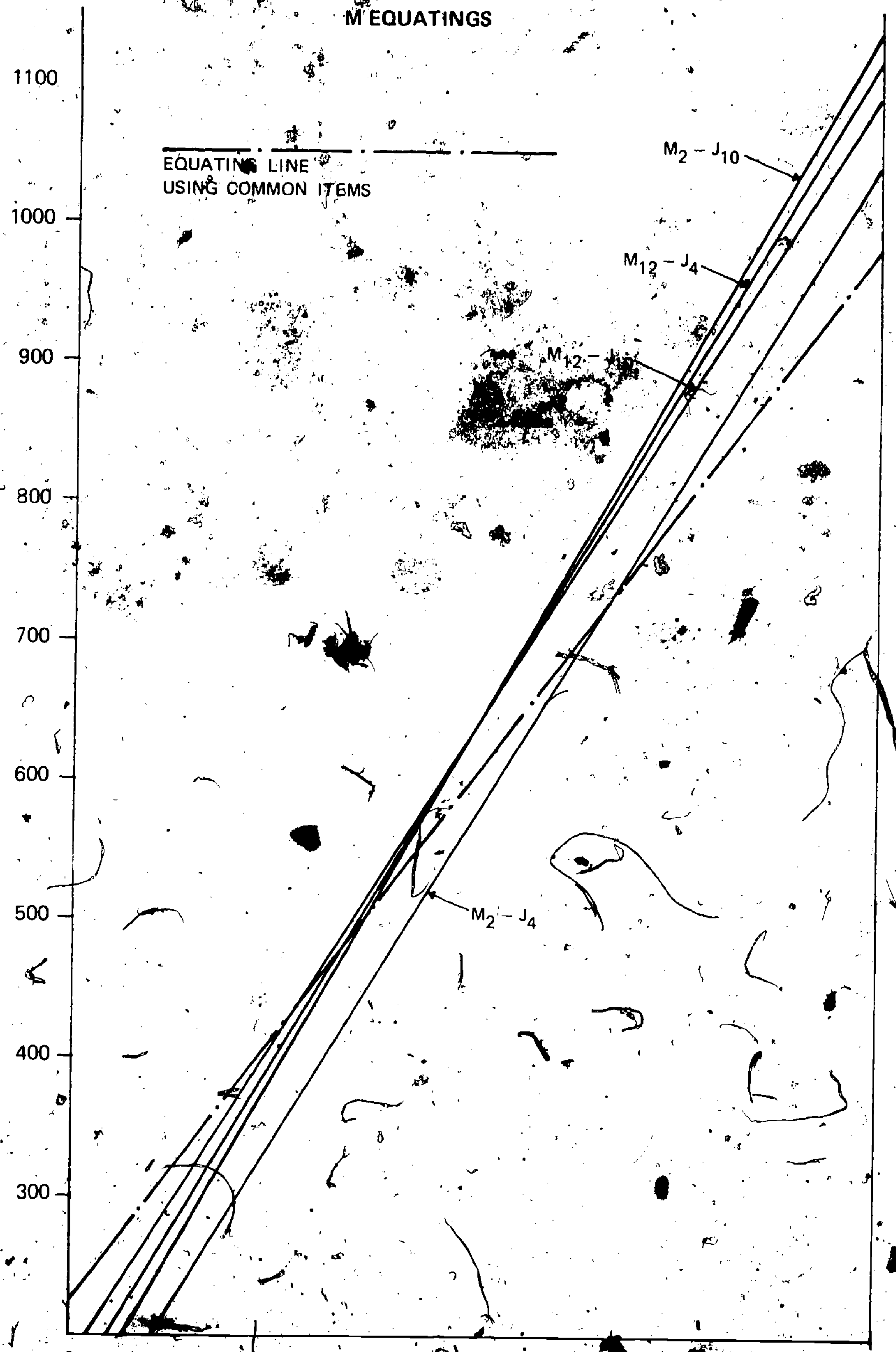

O$_1$ EQUATINGS
EQUATING LINE USING COMMON ITEMS
O$_1$ – K$_{12}$
O$_1$ – L$_{10}$
O$_1$ – L$_2$
APPROXIMATE BISECTOR FOR OLD FORMS K$_{12}$ AND L$_{10}$
APPROXIMATE BISECTOR FOR OLD FORMS K$_{12}$ AND L$_2$
1000
900
800
700
600
500
400
300
0
100
200

PSYCHOLOGY

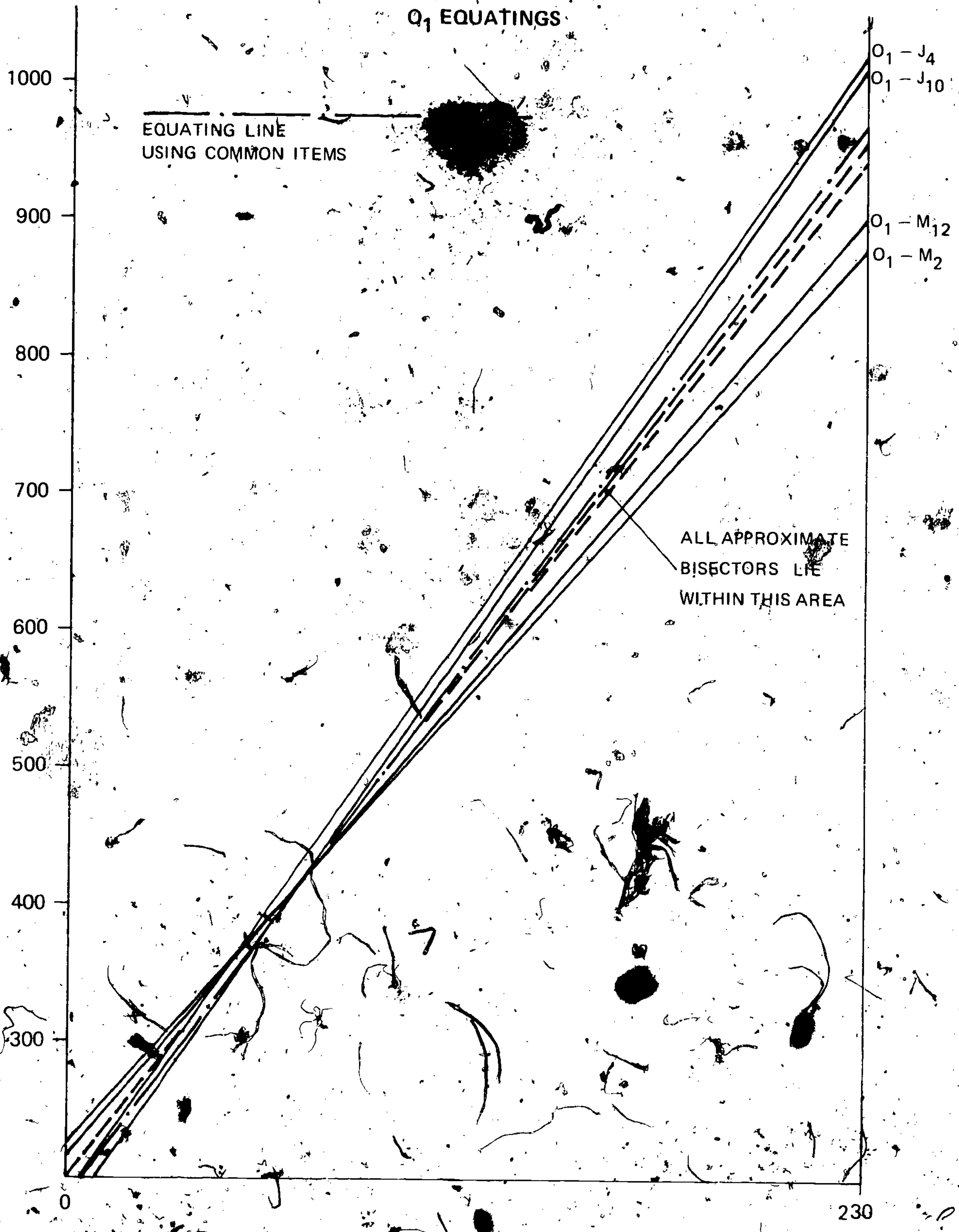

O7 EQUATINGS
EQUATING LINE
USING COMMON ITEMS
O7 – K12
O7 – L10
O2 – L2
APPROXIMATE BISECTOR
FOR OLD FORMS
K12 AND L2
1000
900
800
700
600
500
400
300
0
200


Q EQUATINGS
EQUATING LINE
USING COMMON ITEMS
APPROXIMATE BISECTOR
USING OLD FORMS
L2 AND O1
Q4 – O7
Q4 – O1
Q4 – L2
Q4 – L10
1000
900
800
700
600
500
400
300
0
100
200

SOCIOLOGY

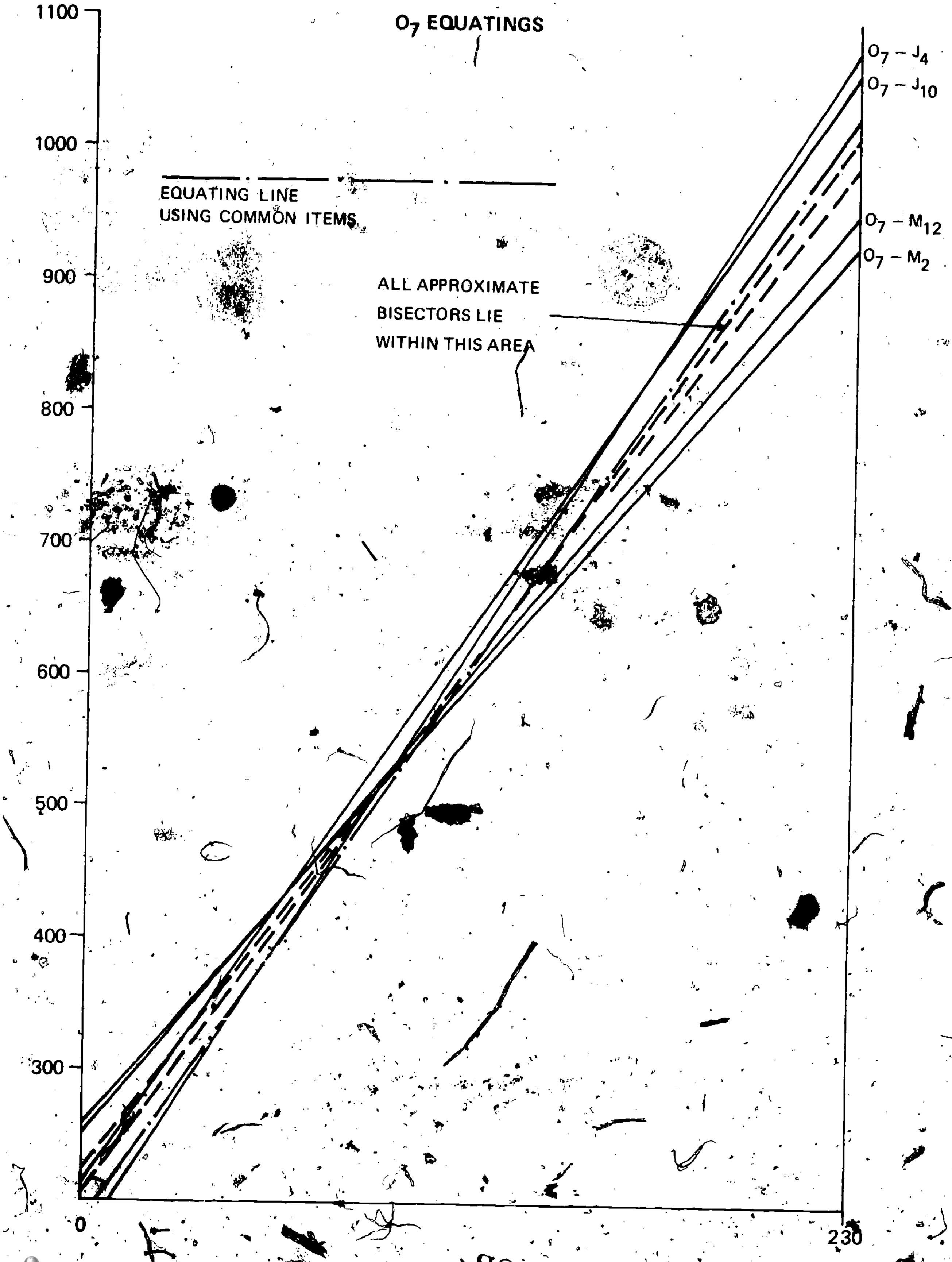

M EQUATINGS
EQUATING LINE
USING COMMON ITEMS
1100
1000
900
800
700
600
500
400
300
M2 – J10
M12 – J4
M12 – J10
M2 – J4

SOCIOLOGY



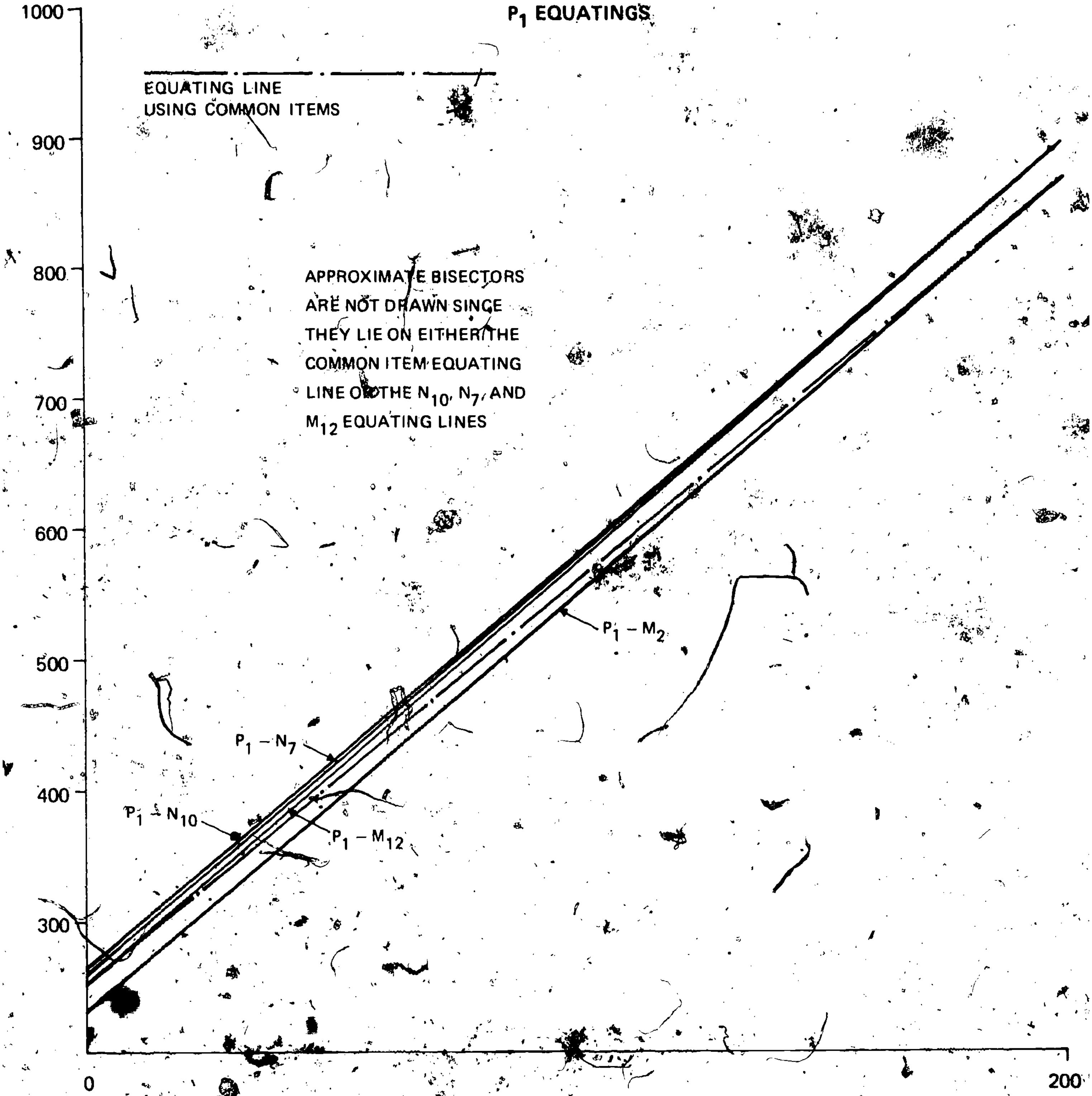
Q1 EQUATINGS
1000
900
800
700
600
500
400
300
0
230
EQUATING LINE USING COMMON ITEMS
O1 – J4
O1 – J10
O1 – M12
O1 – M2
ALL APPROXIMATE BISECTORS LIE WITHIN THIS AREA

SOCIOLOGY



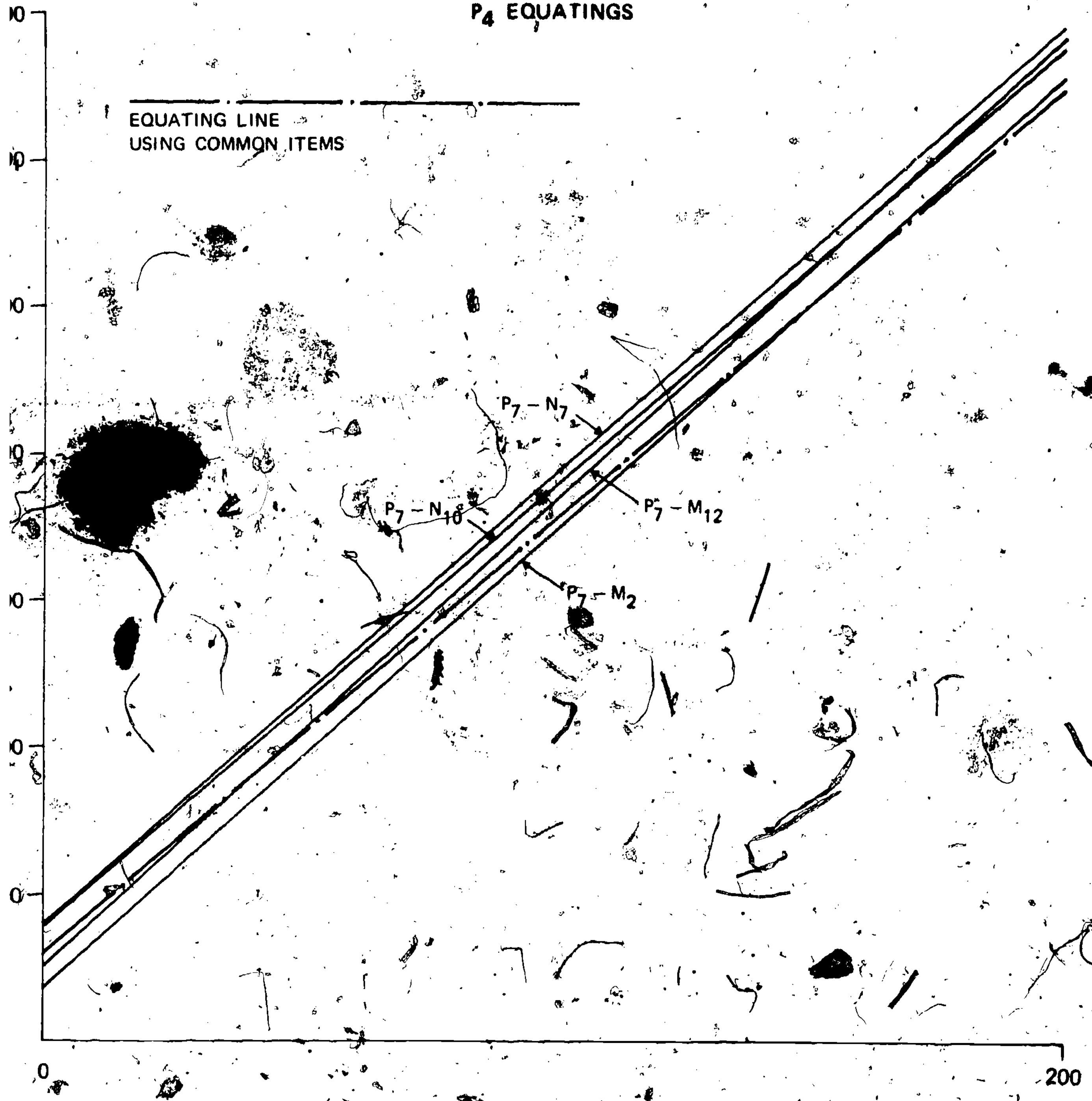
O₇ EQUATINGS
1100
1000
900
800
700
600
500
400
300
0
230
EQUATING LINE USING COMMON ITEMS
ALL APPROXIMATE BISECTORS LIE WITHIN THIS AREA
$O_7 - J_4$
$O_7 - J_{10}$
$O_7 - M_{12}$
$O_7 - M_2$

SPANISH


P1 EQUATINGS
1000
900
800
700
600
500
400
300
0
200
EQUATING LINE USING COMMON ITEMS
APPROXIMATE BISECTORS ARE NOT DRAWN SINCE THEY LIE ON EITHER THE COMMON ITEM EQUATING LINE OR THE N10, N7, AND M12 EQUATING LINES
P1 – M2
P1 – N7
P1 – N10
P1 – M12

SPANISH

P4 EQUATINGS


EQUATING LINE
USING COMMON ITEMS
$P_7 - N_7$
$P_7 - N_{10}$
$P_7 - M_{12}$
$P_7 - M_2$