DOCUMENT RESUME

ED 268 128 TM 850 673

AUTHOR	Dorans, Neil J.; Kulick, Edward
TITLE	The Two-Predictor Validity Curve. Resarch Report.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-85-23
PUB DATE	Jun 85
NOTE	22p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Aptitude Tests; Correlation; *Display Aids; Graphs; Higher Education; Mathematics Tests; *Predictive Measurement; *Predictive Validity; *Predictor Variables; Secondary Education; Statistical Analysis; Statistical Data; *Statistics; Verbal Tests
IDENTIFIERS	*Linear Measurement; Linear Models; Predictive Models; Scholastic Aptitude Test; *Two Predictor Validity Curve

ABSTRACT

A visual display is introduced that enables one to compare the predictive validities of all possible non-negatively weighted sums of two predictors. This two-predictor validity curve assesses the level of predictability of a criterion from composites of two predictors. Derivations are presented for the general case. For illustrative purposes, the two predictors are given as the SAT-Verbal (V) and the SAT-Mathematics (M), and the criterion of interest is the first year college grade point average (FYA). The two-predictor validity curve allows one to assess the relative importance of SAT-V in relation to SAT-M for predicting FYA. Since each display takes up only a small amount of space, it is possible to pack nine on a page, which allows the user to examine these questions in the context of consistency over time, gender or other relevant classification variables. Illustrations are presented using SAT data from 1977 through 1981. (PN)

RR-85-23

RESEARCH REPORT

THE TWO-PREDICTOR VALIDITY CURVE

Neil J. Dorans
Edward Kulick

Educational Testing Service
Princeton, New Jersey
June 1985

The Two-Predictor Validity Curve

Neil J. Dorans
Edward Kulick
Educational Testing Service

June 1985

Abstract

A visual display known as the two-predictor validity curve is introduced. Derivations are presented for the general case. Then, for illustrative purposes the two predictors are taken to be SAT-V (V) and SAT-M (M), and the criterion of interest is the first year college grade point average (FYA). The two-predictor validity curve, in this case SAT V&M, allows one to assess the relative importance of SAT-V in relation to SAT-M for predicting the criterion of interest. When several V&M curves are placed on a single page, comparisons across gender can be made and trends across time can also be studied. Several illustrations using SAT data are presented.

THE TWO-PREDICTOR VALIDITY CURVE

The two-predictor validity curve is a visual display that enables one to compare the predictive validities of all possible non-negatively weighted sums of the two predictors. For instance, assume the two predictors, X and Y, are both positively related to the criterion Z. This display allows one to assess the relative importance of X in relation to Y for predicting the criterion of interest Z. When several displays are placed on a single page, comparisons across genders can be made and trends across time can also be studied.

Mathematics

The mathematics underlying the two-predictor validity curve are straightforward and rest on the theory of linear composites. Consider a composite score (CS) which can be any combination of predictors X and Y, i.e.,

(1) $$CS = W_X X + W_Y Y,$$

where W_X and W_Y, are arbitrary weights applied to X and Y, respectively. Suppose we limit W_X and W_Y to be non-negative because in practice it is unlikely that either X or Y would be assigned a negative weight in a prediction equation. If it is likely, then the direction of the predictor can be reversed via a sign change. For any third variable Z, the correlation of CS with that variable can be computed if certain basic statistics are known, namely, the correlation of this third variable with both X and Y, i.e. R(X,Z) and R(Y,Z), and the standard deviations (or variances) of all three variables, i.e. SD(Z), SD(X), SD(Y) (or VAR(Z), VAR(X), VAR(Y)). The correlation of CS with Z is

(2) $$R(Z,CS) = \frac{COV(Z,CS)}{SD(Z)\ SD(CS)} ,$$

where COV(Z,CS), the covariance of Z with CS, is

$$(3)\quad COV(Z,CS) = COV(Z,W_X X) + COV(Z,W_Y Y)$$
$$= W_X SD(X)R(Z,X)SD(Z) + W_Y SD(Y)R(Z,Y)SD(Z),$$

and

$$(4)\quad SD(CS) = (W_X{}^2 VAR(X) + W_Y{}^2 VAR(Y) + 2W_X W_Y COV(X,Y))^{\frac{1}{2}}$$
$$= (W_X{}^2 VAR(X) + W_Y{}^2 VAR(Y) + 2W_X W_Y SD(X)SD(Y)R(X,Y))^{\frac{1}{2}}.$$

Substituting (3) and (4) into (2) yields

$$(5)\quad R(Z,CS) = \frac{[W_X SD(X)R(Z,X) + W_Y SD(Y)R(Z,Y)]}{[W_X{}^2 VAR(X) + W_Y{}^2 VAR(Y) + 2W_X W_Y SD(X)SD(Y)R(X,Y)]^{\frac{1}{2}}}.$$

This expression demonstrates that R(Z,CS) depends on only W_X, W_Y, the correlations among X, Y and Z and the standard deviations (or variances) of X and Y.

For non-zero weights, it can be shown that the correlation of Z with CS is the same as the correlation of Z with CS/W_X, which is the same as the correlation of Z with CS/W_Y, because CS, CS/W_X and CS/W_Y are perfectly correlated with each other. This means that for all positive weights W_X and W_Y, the composite $CS = W_X X + W_Y Y$ is correlated perfectly with $\ddot{CS} = X + \ddot{W}_Y Y$, where $\ddot{CS} = CS/W_X$ and $\ddot{W}_Y = W_Y/W_X$. Likewise, both CS and $\ddot{CS}$ are perfectly correlated with $\tilde{CS} = \tilde{W}_X X + Y$, where $\tilde{CS} = CS/W_Y = W_X \ddot{CS}/W_Y$ and $\tilde{W}_X = W_X/W_Y$. In short, for any pair of X and Y weights, a simple transformation of variables can be used to express the composite score as a sum of X (or Y) and some weight between 0 and 1 for Y (or X) without changing the correlation of that composite score with a third variable, such as Z.

A straightforward algebraic proof will verify the points made in the preceding paragraphs. It can be shown that $R(Z,CS) = R(Z,\ddot{CS}) = R(Z,\tilde{CS})$, where $CS = W_X X + W_Y Y$, $\ddot{CS} = X + \ddot{W}_Y Y$, $\tilde{CS} = \tilde{W}_X X + Y$, $\ddot{W}_Y = W_Y/W_X$ and $\tilde{W}_X = W_X/W_Y$.

For $\ddot{CS} = CS/W_X$, (5) becomes

$$(6) \qquad R(Z,\ddot{CS}) = \frac{[SD(X)R(Z,X) + \ddot{W}_Y SD(Y)R(Z,Y)]}{[VAR(X) + \ddot{W}_Y^2 VAR(Y) + 2\ddot{W}_Y SD(X)SD(Y)R(X,Y)]^{1/2}} .$$

Substitution of W_Y/W_X for $\ddot{W}_Y$ in (7) yields

$$(7) \qquad R(Z,\ddot{CS}) = \frac{[SD(X)R(Z,X) + (W_Y/W_X)SD(Y)R(Z,V)]}{[VAR(X) + (W_Y/W_X)^2 VAR(Y) + 2(W_Y/W_X)SD(X)SD(Y)R(X,Y)]^{1/2}} .$$

Multiplying (7) by W_X/W_X yields,

$$(8) \qquad R(Z,\ddot{CS}) = \frac{[W_X SD(X)R(Z,X) + W_Y SD(Y)R(Z,Y)]}{[W_X^2 VAR(X) + W_Y^2 VAR(Y) + 2W_X W_Y SD(X)SD(Y)R(X,Y)]^{1/2}} ,$$

which is $R(Z,CS)$. The reader can follow a similar procedure to prove that $R(Z,\tilde{CS})$ also equals $R(Z,CS)$ and $R(Z,\ddot{CS})$.

The two-predictor validity curves utilize the fact that $R(Z,CS)=R(Z,W^*CS)$, where W^* is some positive weight, to portray the correlation of Z with every composite of X and Y that uses non-negative weights for X and Y. Such a display using SAT-V(V), SAT-M(M) and first-year grade point average (FYA) is shown in Figure 1. In Figure 1, the y-axis ranges from 0.0 to 0.4 and represents the correlation of first-year grade point average (FYA) with CS, a composite of SAT-V(V) and SAT-M(M). The x-axis ranges from 0.0 to 1.0 and represents a weight W, which is assigned to either V or M. In the figure are two curves, one short-dashed and one long-dashed. The long-dashed curve (— — —) traces the correlation of FYA with CS=M+WV, where W ranges from 0 to 1. For W equal to zero, this curve crosses the y-axis on the left at R(FYA,M) the simple correlation of FYA with SAT-M. For W equal to one, this long-dashed curve crosses the y-axis on the right at R(FYA,M+V), the correlation of FYA with M+V or V+M.

Insert Figure 1 about here

The short-dashed curve (---) traces the correlation of FYA with CS=V+WM, where W ranges from 0 to 1. For W equal to zero, this short-dashed curve crosses the y-axis on the left at R(FYA,V), the simple correlation of FYA with SAT-V. For W equal to one, this short-dashed curve crosses the y-axis on the right at R(FYA,V+M), the correlation of FYA with V+M.

Interpretation of the Validity Curves

Note that at W=1, the two curves intersect at R(FYA,V+M), which is a point of interest. Other points of interest are W=0, which yields R(FYA,M) for the long-dashed curve and R(FYA,V) for the short-dashed curve, and W=.5, which yields R(FYA,M+.5V) = R(FYA,2M+V) for the long-dashed curve and R(FYA,V+.5M) = R(FYA,2V+M) for the short-dashed curve.

The changes in the slopes of these curves and the direction of their slopes as W increases provide information about the relative importance of SAT-V and SAT-M and information about their exchangability. If the slope is positive, increasing W adds to the validity of the composite. If the slope is negative, increasing W subtracts from the validity of the composite. For example, a positive slope for CS=M+WV (— — —) implies that changing the relative weight from 0 to 1 for SAT-V yields improved prediction over using SAT-M only. The steepness of the slope matters also. To the extent that both curves are flat, SAT-V and SAT-M are exchangeable in the sense that it doesn't matter much whether the ratio of the verbal to math weights approaches infinity or zero. When one curve is steeper than the other, the weighted variable in the steep curve (e.g. V in CS=M+WV) has the most effect on predicting FYA; the direction of the steep slope, which is usually positive, indicates whether the effect on prediction is positive or negative.

Illustrative SAT Data

Figure 1 contains SAT V&M validity curves for a selective college (college A) for the years 1977, 1978 and 1979. The top three panels are based on data for females only, the bottom three panels are based on data for males only, and the middle three panels are based on data combined across gender.

In Figure 1, the simple correlations of SAT-M with FYA range from a low near .20 in the female 1977 data to a high near .40 in the male 1979 data. The SAT-V correlations with FYA range from a low just under [illegible] in the female 1978 data to a high just over .35 in all three 1979 data sets. In contrast to Figure 1, the level of the correlations with FYA in Figure 2 is high, and the patterns are more consistent.

Across all three years of data for the male data, SAT-M is slightly more important than SAT-V, and consequently V+M is slightly superior to 2V+M. Across all three years for the combined data, SAT-V is slightly more important than SAT-M and V+M is also slightly superior to 2V+M. Across all three years of female data, SAT-V is more important than SAT-M. In 1977, 2V+M is slightly superior to V+M, but in 1978 and 1979 V+M is as good as 2V+M if not slightly superior. If one were interested in consistency across time and gender and one had to select a single set of weights to use for SAT-V and SAT-M, the safest bet would be to use a 1 to 1 ratio of weights, i.e., V+M.

Figure 2 depicts the data from a less selective college (college B) for the three years 1979, 1980 and 1981. Note that the scale for Figure 2 validities ranges from .35 to .65 in contrast to the validity scale in Figure 1, which ranges from .15 to .45. The lower magnitudes of validity in Figure 1 are consistent with the attenuating effects of selection.

The simple correlations presented in Figure 2 range from approximately .55 down to a low of near .35, correlations generally higher than those seen in Figure 1. The female data are consistent in that SAT-V and SAT-M appear comparable in predictive power of FYA. The same situation applies to the combined data for 1980. Even more consistent, however, is the finding that V+M appears to be the optimal (or nearly so) composite weighting in each of the 9 panels. Note that in 1980 SAT-M was slightly superior to SAT-V, and that the reverse occurred in 1981.

Insert Figure 2 about here

Figure 3 contains data for another college for 1979, 1980 and 1981. Correlations range from .12 to .18 for the SAT-V and from .18 to .29 for the SAT-M in the female group. Correlations in the male group range from .20 to .28 and .16 to .25 for SAT-V and SAT-M, respectively. Hence, this figure provides a nice contrast to Figure 1 where SAT-V was more valid for females and SAT-M was more valid for males.

Insert Figure 3 about here

Except for the female data in 1979 and 1980, the panels are similar in that, for the most part, as W increases, so do the composite score validities. Thus for all seven of these panels, weighting both verbal and math scores equally (i.e., V+M) is one of the best formations of the composite score, in terms of correlation with FYA. In the male group the short-dashed curve consistently falls above the long-dashed curve. This and the gradual slope of the short-dashed curve mean that a rather broad range of possible weights are nearly equal in producing the optimal correlation between FYA and composite score. For the three male data panels, and the 1980 and 1981 combined group panels and the 1981 female group panel the consistently best ratio of SAT-V and SAT-M in forming the composite score is between 2:1 and 1:1. The two female data panels in 1979 and 1980 suggest weighting the math score more heavily than the verbal score, perhaps by as much as 3:1 or 6:1, but this weighting would be injudicious for the other panels. All panels considered, the validities of V+M and 2V+M appear to be the most consistently high.

Figure 4 presents the data for college D. Simple correlations of SAT-V and FYA range from approximately .29 to .44. Simple correlations of SAT-M with FYA are between .31 and .44.

Insert Figure 4 about here

An important feature of Figure 4 is the similarity, in many panels, of the two curves. For all three years of female data, as well as for the 1981 combined group data and the 1980 male data, the two validity curves are essentially exchangeable. Even in the remaining four panels, where in three instances the short-dashed curve (CS=V+WM) lies above the long-dashed curve (CS=M+WV), the curves are fairly close.

The shapes of the curves are very consistent across all nine panels. Each of the curves increases with W, perhaps leveling off slightly as W approaches 1.0. Since the slopes of the curves are not very steep, it suggests that a broad range of composite scores are comparably correlated with FYA. Across gender in 1979, the relative weighting of verbal to math is optimally between 3 to 1 and 1 to 1. In 1980, also across gender, weighting the verbal scores to the math scores by ratios between 2 to 1 and 1 to 1 seem best. For 1981 male data, the highest correlations with FYA are obtained from composite scores weighting SAT-M by as much as 3 times as SAT-V. The other two panels of 1981 data suggest either score could be weighted more heavily than the other, but by no more than 1.5 times as much.

Considering all the data presented in Figure 4, it seems that all composite scores ranging from equal weighting of SAT-V and SAT-M (i.e., V+M) to weighting the SAT-V by as much as twice SAT-M (i.e., 2V+M) provide consistently high correlations with FYA.

The data for college E are contained in Figure 5. The three panels of data for females in the years 1978, 1979, and 1981 appear most consistent. Simple correlations with FYA range between .3 and .4, with SAT-M in each case correlating more highly. The optimal weighting scheme across these three panels of data is one that weights the math score at least equally, and possibly by twice as much as the verbal score.

Insert Figure 5 about here

For the male data, simple correlations range from .18 to .32 for math and between .19 and .28 for verbal. The relative importance of the two scores is not consistent across years in the male data. The data from 1978 and 1979 suggest that the composite score is best formed by weighting math scores between 1 and 2 times as much as verbal scores. The panel for 1981, however, suggests an optimal composite score would be obtained by weighting the verbal score at least equally and perhaps by 6 times as much as the math score.

In the panels presenting the combined group data, the simple correlation of math with FYA drops with each succeeding year. While the math correlations drop from approximately .31 to .21, the verbal correlations drop only from .29 and .26. The relative importance of SAT-V and SAT-M is not consistent across years for the combined data, but the differences are not very great. The composite scores that would work best in any of these 3 panels are those formed so that neither score is outweighted by a ratio of greater than 1.5 to 1.

Some of the panels for this college suggest the SAT-V ought to be weighted more heavily while other panels suggest the opposite. In every panel, however, equal weighting of the two scores is an acceptable solution.

Summary and Conclusions

Examination of the 45 SAT V&M validity curves produced for this study uncovered some interesting consistencies. In general, equal weighting of SAT-V and SAT-M produces a composite, V+M, with a predictive validity that is very close to the highest attainable by optimal weighting of SAT-V and SAT-M. The most common patterns of results were these: (1) V+M and 2M+V both close to the maximum (in predictive power), with 2V+M somewhat lower; (2) V+M and 2V+M both close to the maximum, with 2M+V somewhat lower. The data also suggested that using V-only or M-only, i.e., giving zero weight to SAT-M or SAT-V, will often yield a noticeably suboptimal predictive validity. In sum, when predicting FYA from SAT-V and SAT-M, it is better to give weight to both SAT-V and SAT-M than to either one exclusive of the other, and if consistency across time and comparability across gender are sought, equal weights for SAT-V and SAT-M should produce a composite that attains a predictive validity that is close to the maximum attainable for these colleges.

These results are consistent with findings reported in the literature on alternative weighting schemes (Dorans and Drasgow, 1978; Einhorn and Hogarth (1975); Wainer (1976). This literature suggests, via theoretical argument (Einhorn, Wainer) and empirical evidence (Dorans and Drasgow), that equal weighting of predictors will yield a composite score with nearly optimal validity that is stable across time and subgroups whenever the predictors are highly correlated with each other and approximately equally correlated with the criterion. Such is typically the case for SAT-V and SAT-M as predictors of first year college grade point average.

These empirical results were used as a vehicle for introducing the two-predictor validity curve. The two-predictor validity curve is a powerful visual tool for assessing the level of predictability of a criterion, such as FYA, from composites of two predictors, such as SAT-V and SAT-M, the relative importance of the two predictors (X and Y) to prediction of the criterion (Z), and the exchangability of X and Y in the predictive context. Since each display takes up only a small amount of space, it is possible to easily pack nine on a page, which allows the user to examine these questions in the context of consistency over time, gender or other relevant classification variables. Generalization of the methodology to the multiple predictors case is straightforward.

References

Dorans, N. J., & Drasgow, F. Alternative weighting schemes for linear prediction. Organizational Behavior and Human Performance, 1978, 21, 316-345.

Einhorn, H. J. & Hogarth, R. M. Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 1975, 13, 171-192.

Wainer, H. Estimating coefficients in linear models: It don't make no never mind. Psychological Bulletin, 1976, 83, 213-217.

SAT V&M VALIDITY CURVES FOR COLLEGE A

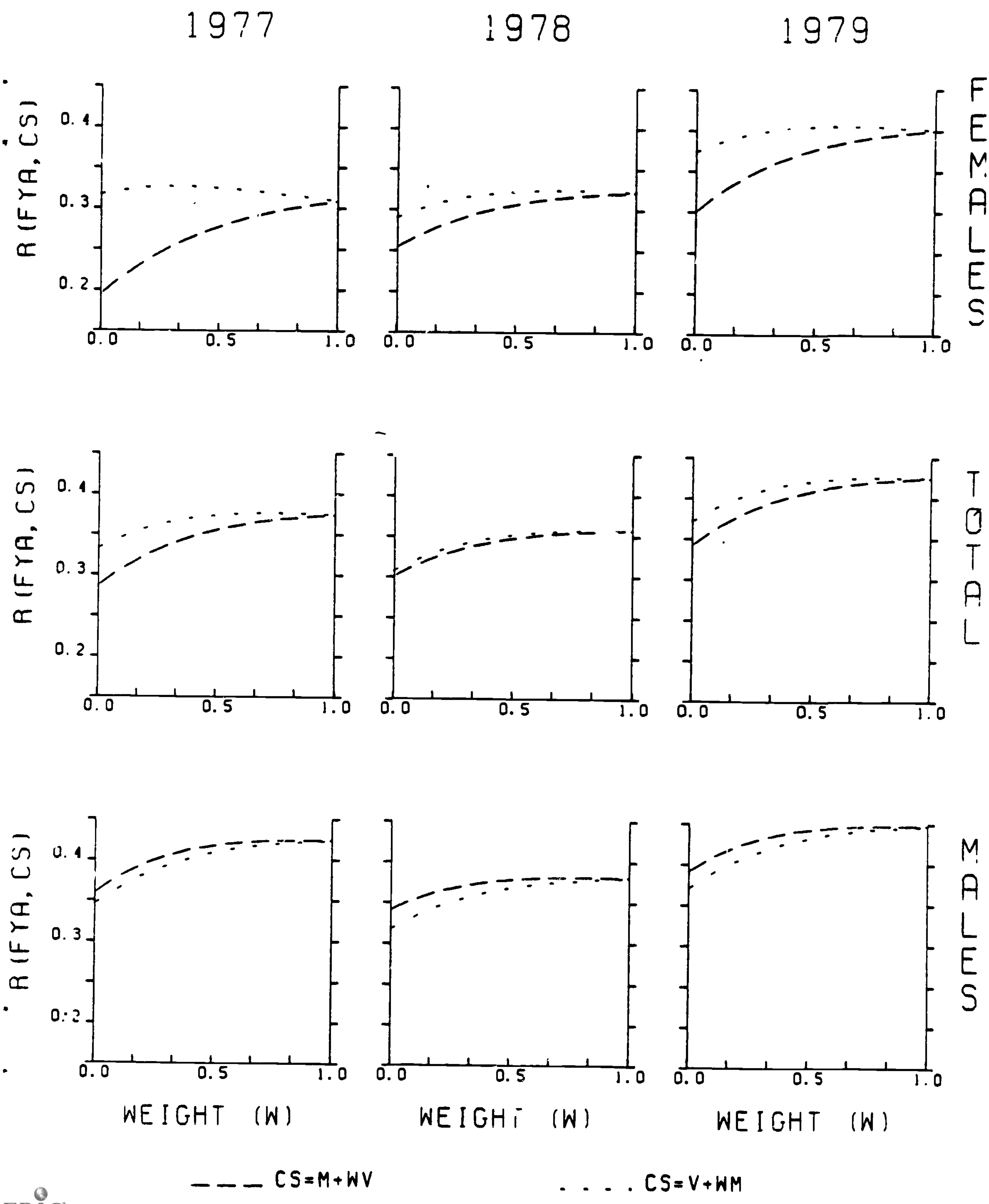

FIGURE 1

SAT V&M VALIDITY CURVES FOR COLLEGE B

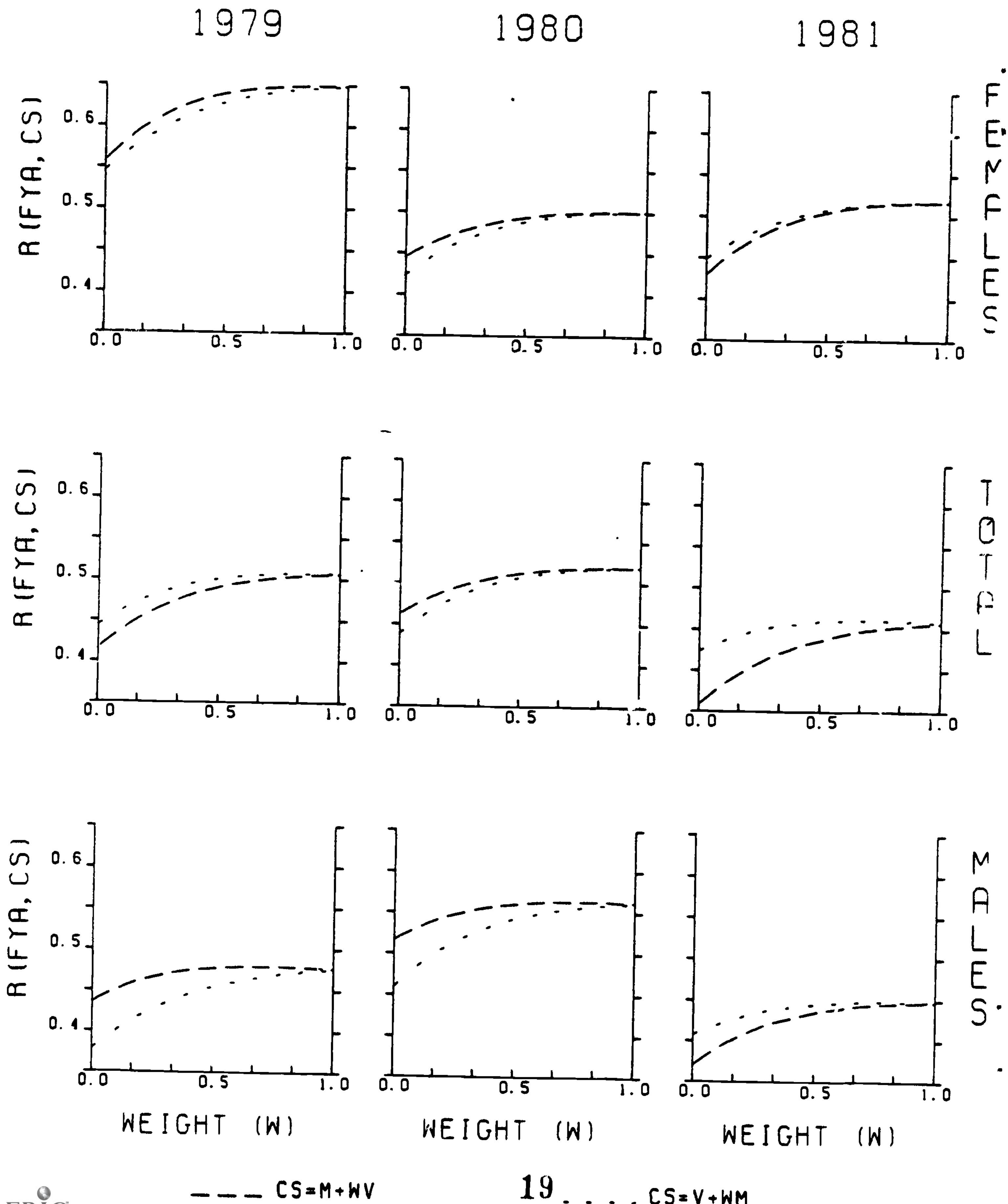

FIGURE 2

SAT V&M VALIDITY CURVES FOR COLLEGE C

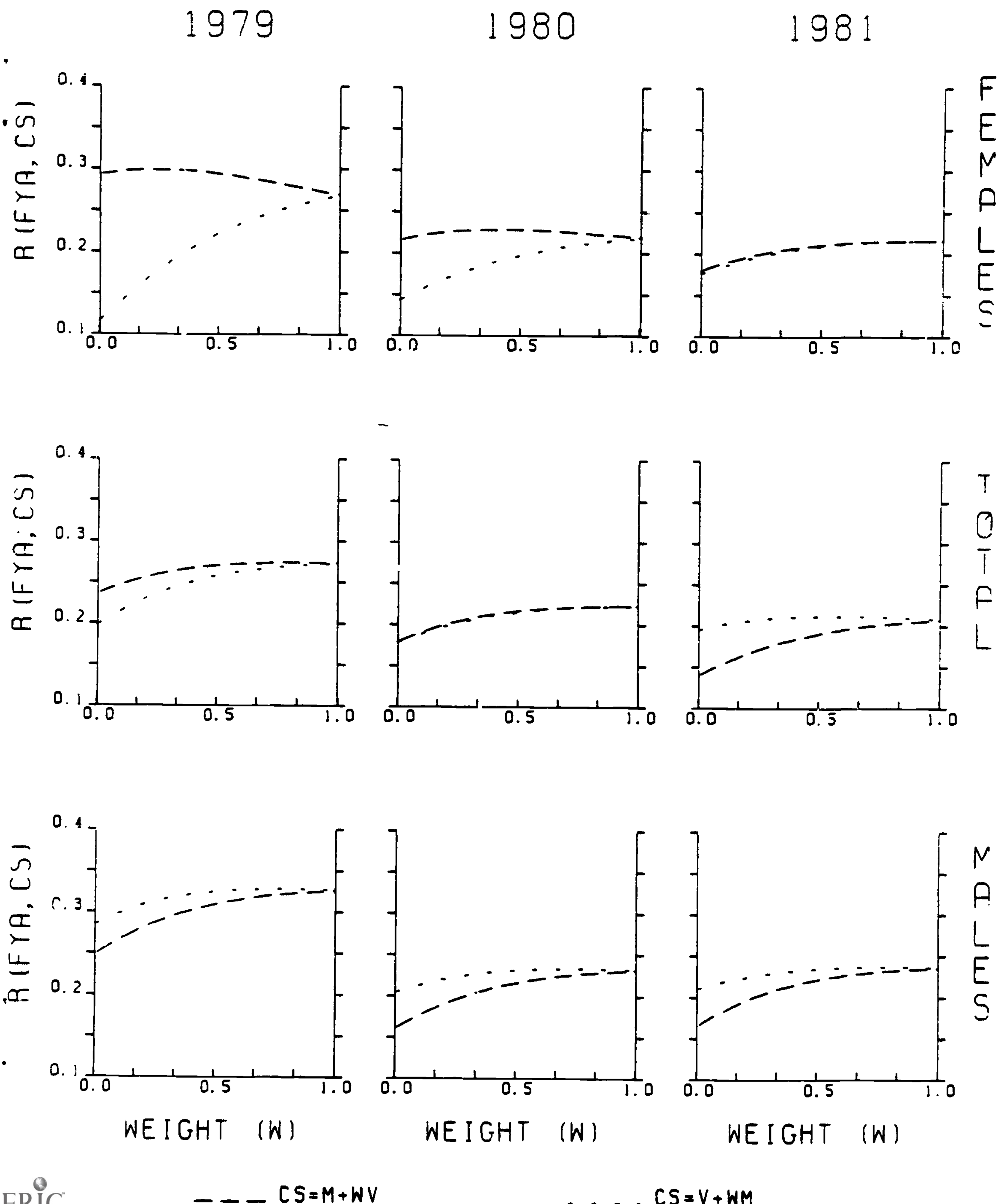

FIGURE 3

SAT V&M VALIDITY CURVES FOR COLLEGE D

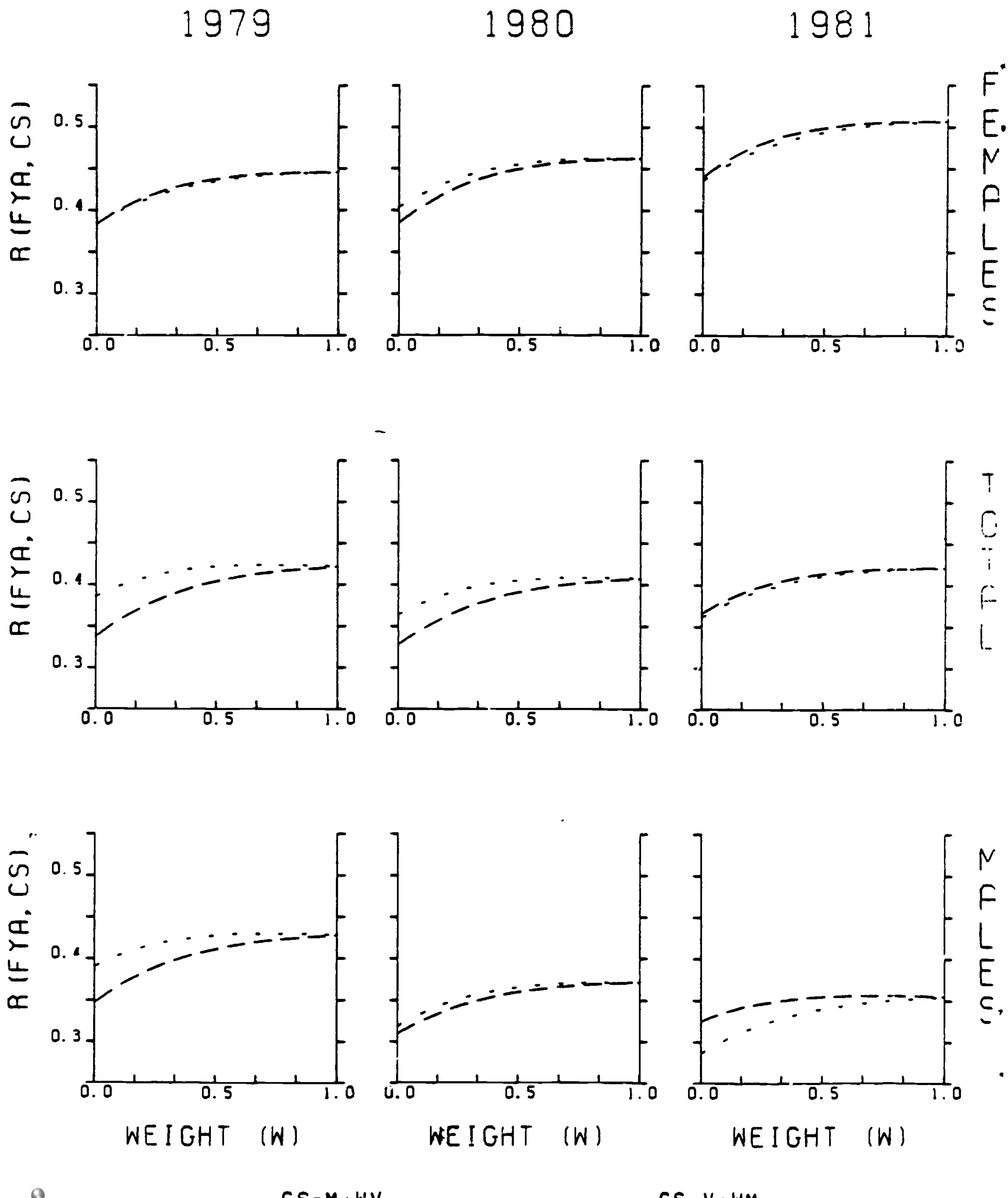

FIGURE 4

SAT V&M VALIDITY CURVES FOR COLLEGE E

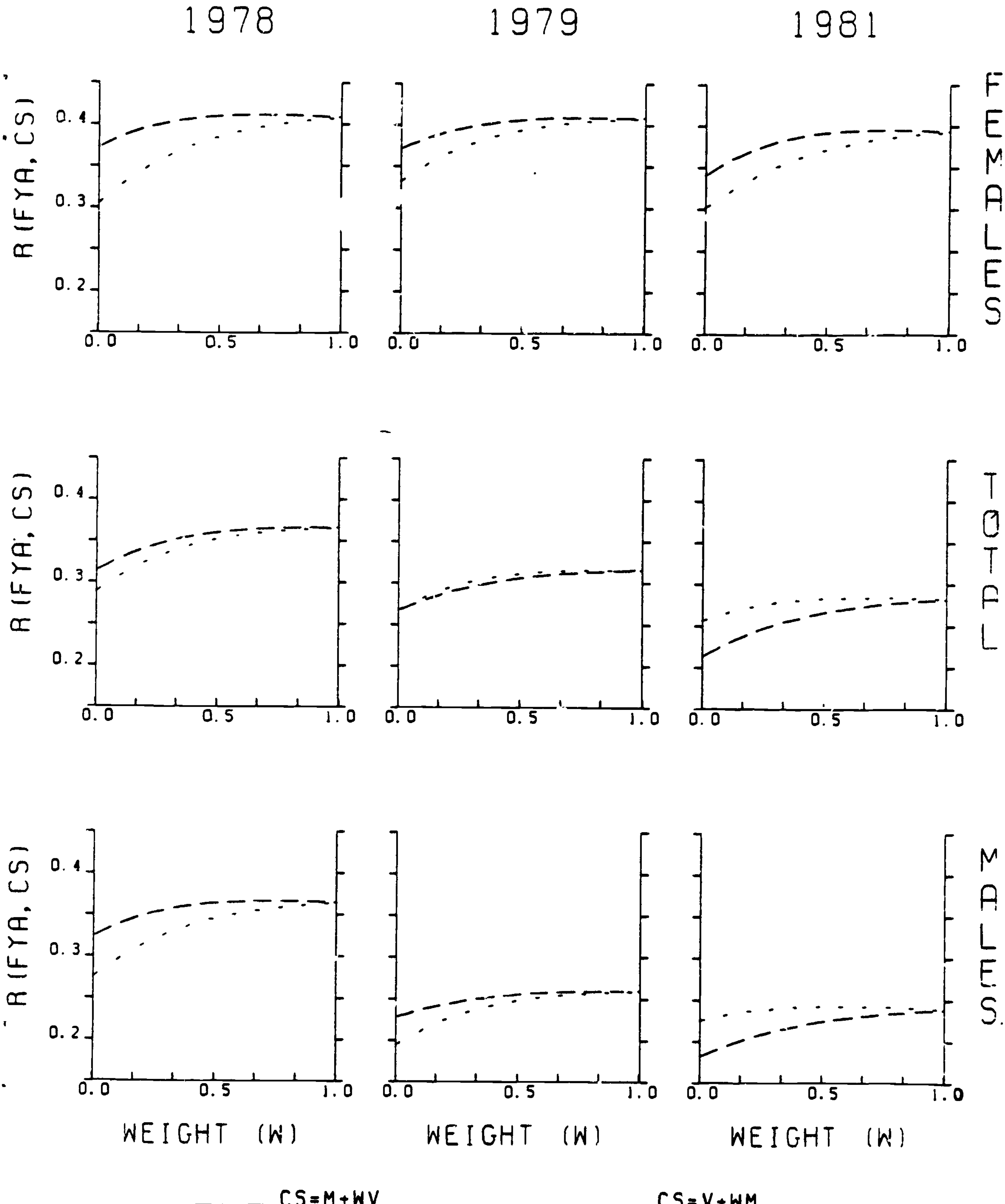

FIGURE 5