DOCUMENT RESUME

ED 331 872 TM 016 419

AUTHOR Dancer, L. Suzanne; And Others
TITLE Examination of Differential Item Functioning in Likert-Type Items Using Log-Linear Models.
SPONS AGENCY Wisconsin Univ., Milwaukee.
PUB DATE Apr 91
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adults; *Background; Blacks; Coping; Culture Fair Tests; *Ethnicity; Hispanic Americans; Individual Characteristics; *Item Bias; Item Response Theory; *Likert Scales; Mental Health; Minority Groups; Models; *Sex Differences; Testing Problems; *Test Items, Whites
IDENTIFIERS Contingency Tables; *Log Linear Models; Self Report Measures; Suicide Probability Scale

ABSTRACT

The use of log-linear models for investigating differential item functioning (DIF) associated with examinee/respondent background characteristics was examined. The Likert-type items used in this study were drawn from a 36-item self-report measure--the Suicide Probability Scale. Specifically, log-linear models were used to investigate whether contingency tables for ethnicity (55 African Americans, 186 Anglo Americans, and 189 Hispanic Americans) or gender (332 males and 627 females) by item response by mental health status suggested evidence of an interaction between the background variable and item response. The investigation focused on a set of 35 Likert-type items that measure subjective well- being and coping behavior. Several log-linear models were fit to the data, and rationale for the composition of the various models is discussed. Among tables where a statistically significant ethnicity by item response interaction or a gender by item response interaction was found, the technique of proportional standardization to unity was used to plot response rates according to ethnic and gender subgroups. Plots show that most of the interaction comes from respondents whose mental health status is diminished. In general, log-linear models were found useful for investigating DIF. Two tables and 24 graphs present study data. (Author/SLD)

EXAMINATION OF DIFFERENTIAL ITEM FUNCTIONING IN LIKERT-TYPE ITEMS USING LOG-LINEAR MODELS

L. Suzanne Dancer
Arthur J. Anderson
Roberta L. Derlin
University of Wisconsin--Milwaukee

ABSTRACT

The use of log-linear models for investigating differential item functioning (DIF) associated with examinee/respondent background characteristics was examined. Specifically, log-linear models were used to investigate whether contingency tables for ethnicity [African American (n=55); Anglo (n=186); Hispanic (n=189)] or gender [females (n=627); males (n=332)] by item response by mental health status suggested evidence of an interaction between the background variable and item response. The investigation focused on a set of 35 Likert-type items that measure subjective wellbeing and coping behavior. Several log-linear models were fit to the data, and rationale for the composition of the various models is discussed. Among tables where a statistically significant ethnicity by item response interaction or a gender by item response interaction was found, the technique of proportional standardization to unity was used to plot response rates according to ethnic and gender subgroup. Plots show that most of the interaction comes from respondents whose mental health status is diminished. In general, log-linear models were found useful for investigating DIF.

EXAMINATION OF DIFFERENTIAL ITEM FUNCTIONING IN LIKERT-TYPE ITEMS USING LOG-LINEAR MODELS*

L. Suzanne Dancer
Arthur J. Anderson
Roberta L. Derlin
University of Wisconsin--Milwaukee

The problem of differential item functioning (DIF) associated with examinee/respondent background characteristics such as ethnicity and gender continues to be an important issue in the measurement of cognitive and attitudinal behavior. With academic and psychological tests playing an ever prominent role in activities such as applicant screening, personnel management, and identification of targeted populations, the need for an understanding of how background characteristics interact with the measurement process cannot be understated. Concern over DIF--that is, the extent to which the "ruler" underlying a test or questionnaire item takes on one set of properties when used to measure Hispanics and another set when measuring African Americans, for example--has been shown to be well-founded (e.g., Becker, 1990; Scheuneman & Gerritz, 1990; Schmitt & Dorans, 1990). While studies of DIF in the cognitive/academic domain have established a solid base of information, less is known about the effect of background variables in measuring attitudinal and psychological functioning. Recognizing the influence of sources of heterogeneity that originate from immersion in diverse ethnic/racial, biological, and cultural backgrounds and understanding the influence of these factors on the process of measurement is fundamental to theory construction in the behavioral sciences.

To date most methods for investigating DIF require assumptions that stretch much data characteristic of the social sciences beyond their limits, and violations of these assumptions seriously jeopardize the integrity of results. For example, item response theory (IRT), a framework with seemingly high potential for exposing DIF, assumes that items measure but a single dimension of behavior. The presence of multidimensionality--an all too common occurrence in the measurement of complex behavioral domains--seriously compromises findings derived from this theoretical framework, thereby limiting its usefulness.

Log-linear methods have been shown to be useful in studies of DIF in cognitive and achievement domains (e.g., Greene, Crone, & Folk, 1989; Mellenbergh, 1982); however, use of these models for examining DIF in attitude measurement has been far less common. The lack of basic research in this area as well as the need for applied research concerning the effects of respondents'

---

* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, April 1991. This research was funded in part by a grant to the first author from the University of Wisconsin-Milwaukee Graduate School Research Committee. The authors are grateful to John G. Cull and Wayne S. Gill for the use of their data.

background characteristics on attitudinal and psychological measurement gave rise to the present investigation. The purpose of our study was to examine the usefulness of methods of log-linear analysis (Goodman, 1968) for investigating DIF associated with ethnicity (African American, Anglo, Hispanic) and with gender in a set of multidimensional Likert-type items. In addition to the expectation that these methods would shed light on certain psychometric properties of the items, it was expected that the log-linear models would provide useful information as to possible differences in psychological functioning across the ethnic and gender subgroups incorporated into our study.

The items examined are polychotomous, having a Likert-type format, and are used to measure subjective wellbeing and coping behavior among persons from the general population as well as from clinical populations. As a precaution against the possible confounding effects of differences in level of psychological functioning among subjects, information as to subjects' mental health status (MHS) was used to hold constant respondents' level of functioning while examining differences in item response across the subgroups. To this end, our study focused on two sets of three-way contingency tables, one set representing the cross-classification of respondents according to item response, MHS, and ethnicity subgroup and the other set crossing item response and MHS with gender. A brief overview of the theoretical framework underlying log-linear models and a detailed description of the tables examined in our analyses follow.

THEORETICAL FRAMEWORK

Log-Linear Models

Only in recent years have general log-linear models been developed sufficiently to make them amenable and useful for applied social research. Historically, examination of cross-classification tables was limited to analysis of two-dimensional tables, and the traditional Pearsonian chi-square goodness-of-fit statistic, χ^2, was used to test the null hypothesis of no association between the row and column variables. In cases where more than two variables were of interest, analyses were more difficult to conduct, generally being carried out on all possible pairs of variables in a way that gave rise to piecemeal investigations of numerous two-way tables. In the 1970's methods for simultaneous analysis of several tables originating from three or more variables become sufficiently refined for practical use. While two-dimensional tables can be analyzed equally well with both traditional χ^2 and the newer log-linear methods, analysis of tables in more than two dimensions is carried out more efficiently using log-linear methods.

In essence, fitting log-linear models to cross-classified data involves assessing the goodness-of-fit of estimated expected cell frequencies to observed counts. Expected frequencies reflect the effects, as specified by a particular model, of various combinations of main effects and interactions stemming from the variables under consideration. The goodness-of-fit of the expected to the observed frequencies is assessed using either the Pearson χ^2 statistic or the likelihood-ratio statistic, L^2. Both statistics have approximate chi-square distributions when total sample size is large and when the log-

linear model under consideration provides a good fit to the observed data. Although the two indices bear several similarities, properties of the L^2 statistic generally make it preferable to the χ^2 statistic. One of these is the fact that the total L^2 value associated with a model can be partitioned into orthogonal, additive components, making it possible to test the statistical significance of individual terms of the model as well as of the model as a whole. Decomposing total L^2 in this way is analogous to the familiar practice in analysis of variance of decomposing systematic variance into additive components that correspond to main effects and interactions.

Specification of Log-Linear Models for Examining DIF

Our analysis was designed to test a number of hypotheses by fitting models to the two sets of tables outlined above. Our principal hypothesis was that neither ethnicity nor gender interacts with responses to items sampled from the domain of subjective wellbeing and coping behavior. This notion corresponded to the statistical hypothesis that terms denoting the interaction of item response and ethnic or gender subgroup were not needed to adequately model the pattern of frequencies in the two sets of tables. Our second hypothesis reflected the notion that we fully expected the need for a term denoting the interaction of item response with MHS in modeling the observed data. More specifically, we expected that, relative to other possible interaction terms, the item response by MHS term would account for the largest proportion of the structure of a table. The rationale for this expectation was based on the fact that the items examined in our study, described in detail below, have been shown to discriminate effectively between normative populations and clinical populations whose level of psychological functioning is diminished. This hypothesis corresponded to the statistical notion that any log-linear model which adequately described the structure of a table would contain a statistically significant term for the interaction of MHS and item response and that when compared to a null model, the model containing this interaction term would give a substantially better fit to the observed data than a model lacking this term.

METHOD

Instrument

The Likert-type items used in this research were drawn from a 36-item self-report measure, the Suicide Probability Scale (SPS; Cull & Gill, 1982). These items clearly are not unidimensional (Dancer, 1990), measuring several aspects of subjective wellbeing (e.g., "Things seem to go well for me.") and coping behavior (e.g., "I have trouble finding and keeping a job I like."). All items have four response categories depicting frequency of some behavior: none or a little of the time, some of the time, a good part of the time, much or all of the time. For some items, the less frequent a particular behavior, the more psychologically healthy the respondent is thought to be (e.g., a response of "none or a little of the time" to Item 12 which reads "I feel so lonely I cannot stand it."). In other cases, the more frequent the behavior, the more healthy the respondent (e.g., a response of "much of the time" to Item 6 which states "I feel there is much I can do that is worthwhile."). Because one of the items (Item

11) is ambiguous and can be interpreted in either of two contradictory ways, it was excluded, leaving 35 items for analysis.

Subjects

Data for this investigation were drawn from the responses of 1158 adults living in the southwestern U.S. who served as the standardization sample for the SPS. These respondents represented three categories of MHS: a normative sample comprised of individuals with no psychiatric history and no history of suicidal behavior; psychiatric inpatients under psychiatric care at the time they took the SPS but having no history of suicidal behavior; and suicidal persons who were administered a series of psychological measures, including the SPS, no more than 48 hours after a potentially fatal suicide attempt.

In addition to completing the SPS, subjects were asked to indicate ethnicity and gender. Because a large number of respondents failed to provide information on one or both of these background variables, the construction of four-way tables for ethnicity by gender by item response by MHS, which would have allowed for examination of the simultaneous effects of ethnicity and gender on item response, was considered unfeasible. The amount of missing information would have substantially reduced the usable sample size. To minimize the loss of data, the decision was made to construct two separate sets of three-way tables, one set for ethnicity by item response by MHS and the other set for gender by item response by MHS. Although this approach maximized overall sample size, it precluded investigation of the three-way interaction for ethnicity, gender, and item response.

Approximately one-third of the 1158 respondents provided information as to ethnicity, with three ethnic subgroups being large enough in number to permit analysis. Of these respondents, 189 were Hispanic, 186 were Anglo, and 55 were African American, for a total of 430 respondents. When categorized according to MHS, 152 were from the normative sample, 138 were psychiatric inpatients, and 140 were suicidal. Responses of these subjects were used to construct separate cross-classification tables for each SPS item, yielding 35 tables whose dimensions were 4x3x3 corresponding to the four item response categories, the three MHS categories, and the three categories of ethnicity. From table to table, the total number of observations varied somewhat owing to occasional missing data on one item or another. However, in no case were the fluctuations substantial.

Nine hundred and sixty-five respondents indicated their gender: females numbered 627 and males numbered 332. When classified according the MHS, 405 of these were from the normative sample, 250 were psychiatric inpatients, and 310 were suicidal. Cross-classifying these respondents led to 35 tables whose dimensions were 4x3x2, corresponding to the number of categories in the item response, MHS, and gender variables. Again, missing data on some items resulted in minor, inconsequential fluctuations of sample size from table to table.

ANALYSIS

As noted earlier, the hypotheses under consideration in our study, in part, dictated specific components of our log-linear models, while other components were necessitated by the sampling design that led to the data for this research. Because the distribution of respondents in the ethnic, gender and MHS categories was fixed by the manner in which the data were collected [i.e., the marginal distributions of these two variables reflect an attempt to get large sample sizes for minority (Hispanic and African American) and low base-rate (suicidal) populations], these analyses were viewed as methods of asymmetric inquiry. This is in contrast to symmetric modes of inquiry where samples are randomly drawn with no constraints imposed on the marginal distributions on any of the variables.

As is customary in asymmetric inquiries, the variables whose marginals were fixed by the sampling design--ethnicity, gender, and MHS--were conceptualized as independent, explanatory variables while item response, a variable whose marginals were free to vary, was considered a dependent variable. This perspective required that all models fit to the tables based on ethnicity contain a term for the MHS by ethnicity interaction, denoted [ME], to insure that marginal distributions of the fixed variables maintained values determined by the sampling scheme. Likewise, in analyzing tables based on gender, the asymmetric perspective required that models include a term for the interaction of MHS and gender, denoted [MG]. However, because the [ME] and [MG] terms were a reflection of the sampling scheme, their effects hold no substantive value and interpretation is inappropriate.

The two models--one containing only an [ME] term and one containing only the term [MG]--were viewed as null models in these analyses. The L^2 values arising from the fit of these models to the observed data provided a baseline measure of the degree to which other interactions, specifically interactions between background characteristics and item response and between MHS and item response, were not present in the data. Small L^2-values and their associated large p-values were interpreted as meaning that the probability of observing an L^2-value of this magnitude by chance, given that a [ME] or [MG] interaction was sufficient for modeling the data, was high, and hence the model was considered a good fit. On the other hand, large L^2-value and the associated low p-values indicated that the probability of observing a statistic of this magnitude given the truth of the model, was quite small and, hence, the model was considered inadequate for describing the pattern of observations in a table.

Building on the baseline models, other models fit to the data contained one or more additional interaction terms in accordance with our two hypotheses. To test the hypothesis that a MHS by item response interaction is present in the data, the term [MR] denoting this interaction was added to the baseline models. To test the hypothesis that an interaction between ethnicity or gender and item response was not present in the data, yet another term was added to the model; models fitted to the ethnicity tables included an [ER] term for the interaction of ethnicity and item response while models fitted to the gender tables included the term [GR] for the interaction of gender and item response.

Hence, a series of three hierarchical log-linear models were fitted to each of the 35 tables for ethnicity and to each of the 35 tables for gender. The first, and least restrictive, model in the hierarchy was a null model that contained either the single term [ME] or [MG], depending on whether the tables being modeled were based on ethnicity or gender, respectively. The second model, either [ME][MR] or [MG][MR], contained but one additional term. In accordance with our hypothesis, this model was expected, at a minimum, to provide a substantial improvement in fit over the null model and, at best, to fully describe the observed data. The third and final model, denoted either [ME][MR][ER] or [MG][MR][GR] again depending on whether tables for ethnicity or gender were being modeled, differed from its predecessor by a single term, and this term denoted the interaction between ethnicity or gender and item response. In accordance with our principal hypothesis, the third model was not expected to substantially improve on the fit of the second model. A component L^2 value resulting from the difference between the L^2 values and associated degrees of freedom for the second and third models was used to test the statistical significance of the component in the third model depicting the interaction between ethnicity or gender and item response. A component L^2 value that exceeded the critical value at an α-level of .05 was taken an evidence of a statistically significant interaction component.

RESULTS AND DISCUSSION

Values of L^2 , along with the associated degrees of freedom and p-values, corresponding to the three models fit to the ethnicity tables, as well as the component L^2 value corresponding to the [ER] term, are shown in Table 1. Similarly, L^2 values from analyses of tables based on gender are shown in Table 2. As can be seen in both tables, in no case did the null model fit the tables. Consistent with our hypothesis, in every case, the second model whose only additional term represented the [MR] interaction greatly improved on the fit and, in most cases, was sufficient for modeling the data.

Tables 1 and 2 about here

As further shown in Table 1, in only four of the 35 tables for ethnicity--tables corresponding to SPS items 6, 8, 26, and 36--were the three terms [ME][MR][ER] needed to model the data. The need for the [ER] term was taken as evidence of DIF associated with ethnicity in these four items. In two additional tables, those for SPS items 10 and 20, the model [ME][MR] provided a good fit to the data, but the component L^2 value indicated a statistically significant [ER] interaction term. Though it was not imperative that this term be included when modeling the data, the statistical significance of this interaction nonetheless was taken as evidence that ethnicity has some bearing on responses to these two items. For the remaining 29 items, log-linear analyses did not provide any evidence of DIF.

To examine the ethnicity by item response interaction in more detail and to contrast items showing evidence of DIF with items where no such interaction was evident, graphs were constructed

for a subset of items so that, holding MHS constant, the relative proportion of respondents from each ethnic subgroup who selected each response category could be seen. Observed frequencies across response categories of each item were converted to proportions that summed to 1.0--a technique known as standardization to proportional unity--so that across an item's categories and within each level of ethnicity and each level of MHS the frequencies were normalized to sum to 1.0. These proportions were then plotted. Plots showing evidence of DIF are those in which response rates differ markedly across ethnic subgroups. Figure 1 includes plots for two items--Items 6 and 36--showing evidence of DIF and two items--Items 5 and 12--for which DIF is not evident.

Figure 1 about here

The plots for Items 6 and 36 show the [ER] interaction. For every MHS category, the proportions in each ethnic group who endorse each item response category vary to some degree, with the greatest differences in response rates being observed for psychiatric inpatients and suicidal respondents. Response rates among normal respondents from each ethnic category come close to being uniform. For Items 5 and 12, on the other hand, the little interaction that is evident occurs for the suicidal group only and even then it is not statistically significant.

As shown in Table 2, for only one of the 35 tables for gender--the table for Item 12--was a model with the three terms [MG][MR][GR] needed to fit the observed data. The need for the [GR] term was taken as evidence of DIF associated with gender for this item. In six additional tables, those for Items 5, 19, 26, 28, 33, and 34, the model [MG][MR] provided a good fit to the data, but the component L^2 value indicated a statistically significant [GR] interaction term. Again, the statistical significance of this interaction term was taken as evidence that responses to these two items was dependent in part on gender, even though the model [MG][MR] sufficed for describing the data. For the remaining 28 items, no evidence of DIF was found.

Figure 2 shows plots of item response rates based on gender. As in Figure 1, these plots include two items--Items 5 and 12--for which DIF associated with gender was evident and two items -- Items 6 and 36--whose responses seemingly are not affected by gender. Even though the [GR] term was statistically significant for Items 5 and 12, the plots show that for each MHS level, response rates for males and females differ by no more than 0.10, and for most of the response categories, women have a higher response rate than men. The statistical significance of the interaction observed in these two plots could well be a reflection, at least in part, of the large sample size on which the analyses were based. Plots for Items 6 and 36 show that response rates for women and men are all but uniform, an observation that is consistent with the fact that a [GR] interaction term was not needed to model tables for these items.

Figure 2 about here

Examination of the content of the six items for which evidence of DIF associated with ethnicity was found and a similar examination of the seven items evidencing DIF associated with gender provided no clues as to characteristics of the items which might be thought of as engendering DIF. The content of the items varied widely, from issues of social relationships to pressures associated with responsibilities, and the format of the items varied with some, but not all, requiring reverse scoring. Thus, there were no immediately obvious commonalities among the items, either in terms of content or format, that could be linked to DIF. Despite the fact that an explanation for the DIF is not evident, results of these analyses suggest that log-linear models are indeed a useful tool for investigating DIF in the sense of detecting an interaction between item response and some background variable and of providing detailed information as to the contribution of subgroup differences to that interaction.

REFERENCES

Becker, B. J. (1990). Item characteristics and gender differences on the SAT-M for mathematically able youths. American Educational Research Journal, 27(1), 65-87.

Cull, J. G., & Gill, W. S. (1982). Suicide probability scale. Los Angeles: Western Psychological Services.

Dancer, L. S. (1990). Suicide prediction and the partial order scalogram analysis of psychological adjustment. Applied Psychology: An International Review, 39(4), 479-497.

Goodman, L. A. (1968). The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries. Journal of the American Statistical Association, 69, 1091-1131.

Greene, B. F., Crone, C. R., & Folk, V. G. (1989). A method for studying differential distractor functioning. Journal of Educational Measurement, 26(2), 147-160.

Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. Journal of Educational Statistics, 7(2), 105-118.

Scheuneman, J. D., & Gerritz, K. (1990). Using differential item functioning procedures to explore sources of item difficulty and group performance characteristics. Journal of Educational Measurement, 27(2), 109-131.

Schmitt, A. P., & Dorans, N. J. (1990). Differential item functioning for minority examinees on the SAT. Journal of Educational Measurement, 27(1), 67-81.

TABLE 1

Log-Linear Analysis of Tables Cross-Classifying Ethnicity, Mental Health Status, and Response to Polychotomous Items Measuring Subjective Wellbeing and Coping Behavior

	LOG-LINEAR MODELS[1]									
	[ME]		[ME][MR]			[ME][MR][ER]			Test of [ER] Component	
ITEM	L^2	df	L^2	df	p	L^2	df	p	L^2	df
1. When...mad I throw things.	458.20	27	20.36	15	.16	9.65	9	.38	10.71	6
2. ...people care for me deeply.	155.04	27	16.20	18	.58	13.36	12	.34	2.84	6
3. ...tend to be impulsive.	169.02	27	25.37	18	.12	15.48	12	.22	9.89	6
4. ...think bad things.	288.88	27	25.04	18	.12	15.09	12	.24	9.95	6
5. ...too much responsibility.	162.62	27	20.94	18	.28	10.07	12	.61	10.87	6
6. ...much I can do worthwhile.	187.35	27	29.28	18	.05	16.39	12	.17	12.89	6
7. ...think suicide to punish others.	592.57	27	26.30	18	.09	16.15	12	.19	10.15	6
8. ...feel hostile toward others.	399.61	27	30.56	18	.03	14.94	9	.09	15.62	9
9. ...feel isolated from others.	278.39	27	24.38	18	.14	16.12	12	.19	8.26	6
10. ...people appreciate me.	146.21	27	27.21	18	.08	12.29	12	.42	14.92	6
12 ...feel so lonely..cannot stand it.	337.78	27	14.60	15	.48	10.77	9	.29	3.83	6
13. Others...hostile to me.	378.43	27	21.08	18	.28	17.21	12	.14	3.87	6
14. ...make many changes in...life.	147.43	27	23.94	18	.16	13.43	12	.34	10.51	6
15. ...not able to do things well.	261.31	27	24.49	18	.14	15.09	12	.24	9.40	6
16. ...trouble finding and keep job.	375.47	27	16.27	18	.57	13.35	12	.34	2.92	6
17. ...no one will miss me.	199.36	27	17.13	18	.51	8.91	12	.71	8.22	6
18. Things...go well for me.	143.37	27	16.81	18	.54	13.56	12	.33	3.25	6
19. ...people expect too much.	147.63	27	27.53	18	.07	15.55	12	.21	11.98	6
20. ...I need to punish myself.	344.65	27	26.95	18	.08	12.03	12	.44	14.92	6
21. ...not worth continuing to live.	478.51	27	15.38	18	.64	10.56	12	.57	4.82	6
22. I plan for the future....	70.48	27	17.50	18	.49	9.63	12	.65	7.87	6
23. ...no friends to count on.	199.82	27	19.65	18	.35	15.62	12	.21	4.03	6
24. ...people...better if I were dead.	518.67	27	18.07	15	.26	14.88	9	.09	3.19	6
25. ... less painful to die....	449.77	27	11.68	15	.70	9.02	9	.44	2.66	6
26. I feel ...close to my mother.	162.63	27	30.27	18	.04	14.71	12	.26	15.56	6
27. I feel ...close to my mate.	105.95	27	20.32	18	.32	15.47	12	.22	4.85	6
28. I feel hopeless	281.48	27	23.94	18	.16	13.12	12	.36	10.82	6
29. ...people do not approve of me.	288.72	27	20.42	18	.31	13.62	12	.33	6.80	6
30. ...thought of how to do self in.	438.76	27	27.88	18	.06	18.39	12	.10	9.49	6
31. ...worry about money.	106.00	27	14.24	18	.71	9.51	12	.66	4.73	6
32. ...think of suicide.	502.89	27	20.99	18	.28	13.57	12	.33	7.42	6
33. ...feel tired and listless.	256.73	27	17.56	18	.49	6.24	12	.90	11.32	6
34. When...mad I break things.	535.73	27	18.53	18	.42	11.18	12	.26	7.35	6
35. I feel ...close to my father.	61.81	27	12.30	18	.83	10.36	12	.58	1.94	6
36. I can't be happy....	425.86	27	30.54	15	.01	16.46	9	.06	14.08	6

[1] [ME] denotes Mental Health Status x Ethnicity interaction; [MR] denotes Mental Health Status x Item Response interaction; [ER] denotes Ethnicity x Item Response interaction

TABLE 2

Log-Linear Analysis of Tables Cross-Classifying Gender, Mental Health Status, and Response to Polychotomous Items Measuring Subjective Wellbeing and Coping Behavior

	LOG-LINEAR MODELS[1]									
	[MG]		[MG][MR]			[MG][MR][GR]			Test of [GR] Component	
ITEM	L^2	df	L^2	df	p	L^2	df	p	L^2	df
1. When...mad I throw things.	156.47	15	16.25	9	.06	11.66	6	.07	4.59	3
2. ...people care for me deeply.	157.63	15	12.09	9	.21	9.51	6	.15	2.58	3
3. ...tend to be impulsive.	56.59	15	9.81	9	.37	4.85	6	.56	4.96	3
4. ...think bad things.	150.74	15	4.00	9	.91	2.83	6	.83	1.17	3
5. ...too much responsibility.	153.11	15	13.42	9	.15	1.82	6	.94	11.60	3
6. ...much I can do worthwhile.	91.05	15	7.83	9	.55	2.33	6	.89	5.5	3
7. ...think suicide to punish others.	180.23	15	18.19	9	.03	15.16	6	.02	3.03	3
8. ...feel hostile toward others.	142.94	15	26.58	9	.00	19.87	6	.00	6.71	3
9. ...feel isolated from others.	215.17	15	10.70	9	.30	6.23	6	.40	4.47	3
10. ...people appreciate me.	127.30	15	15.91	9	.07	7.13	6	.31	8.78	3
12 ...feel so lonely..cannot stand it.	393.28	15	20.12	9	.02	11.24	6	.08	8.88	3
13. Others...hostile to me.	150.21	15	10.46	9	.31	5.86	6	.44	4.60	3
14. ...make many changes in...life.	235.27	15	1.75	9	.99	1.39	6	.97	0.36	3
15. ...not able to do things well.	228.61	15	19.92	9	.02	18.19	6	.01	1.73	3
16 ...trouble finding and keep job.	250.90	15	7.06	9	.63	5.76	6	.45	1.30	3
17. ...no one will miss me.	213.99	15	5.85	9	.76	4.53	6	.61	1.32	3
18. Things...go well for me.	263.97	15	14.19	9	.12	9.47	6	.15	4.72	3
19. ...people expect too much.	142.47	15	11.70	9	.23	2.59	6	.86	9.11	3
20. ...I need to punish myself.	197.80	15	14.21	9	.12	11.44	6	.08	2.77	3
21. ...not worth continuing to live.	308.09	15	10.03	9	.35	6.16	6	.41	3.87	3
22. I plan for the future....	107.16	15	6.89	9	.65	4.12	6	.66	2.77	3
23. ...no friends to count on.	184.18	15	8.75	9	.46	3.34	6	.77	5.41	3
24. ...people...better if I were dead.	340.75	.5	20.21	7	.01	13.43	4	.01	6.78	3
25. ... less painful to die....	344.55	15	10.40	9	.32	8.44	6	.21	1.96	3
26. I feel ...close to my mother.	75.91	15	16.18	9	.06	7.50	6	.28	8.68	3
27. I feel ...close to my mate.	98.73	15	27.25	9	.00	15.17	6	.02	12.08	3
28. I feel hopeless	296.95	15	10.60	9	.30	1.97	6	.92	8.63	3
29. ...people do not approve of me.	213.94	15	6.34	9	.71	5.08	6	.53	1.26	3
30. ...thought of how to do self in.	278.50	15	11.54	9	.24	9.71	6	.14	1.83	3
31. ...worry about money.	106.30	15	10.30	9	.33	6.88	6	.33	3.42	3
32. ...think of suicide.	238.64	15	11.32	9	.26	10.77	6	.10	0.55	3
33. ...feel tired and listless.	298.68	15	14.01	9	.12	5.54	6	.48	8.47	3
34. When...mad I break things.	144.29	15	16.91	9	.05	5.91	6	.43	11.00	3
35. I feel ...close to my father.	42.90	15	7.05	9	.63	1.91	6	.93	5.14	3
36. I can't be happy....	264.40	15	7.11	9	.63	5.97	6	.43	1.14	3

[1] [MG] denotes Mental Health Status x Gender interaction; [MR] denotes Mental Health Status x Item Response interaction; [GR] denotes Gender x Item Response interaction

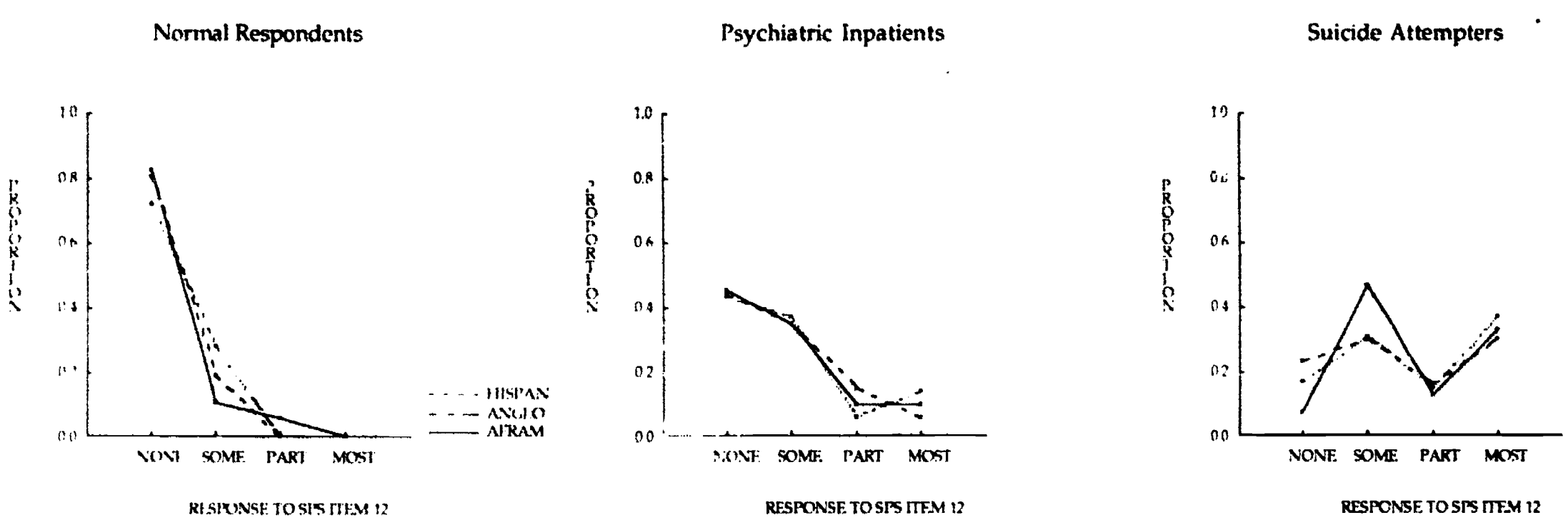
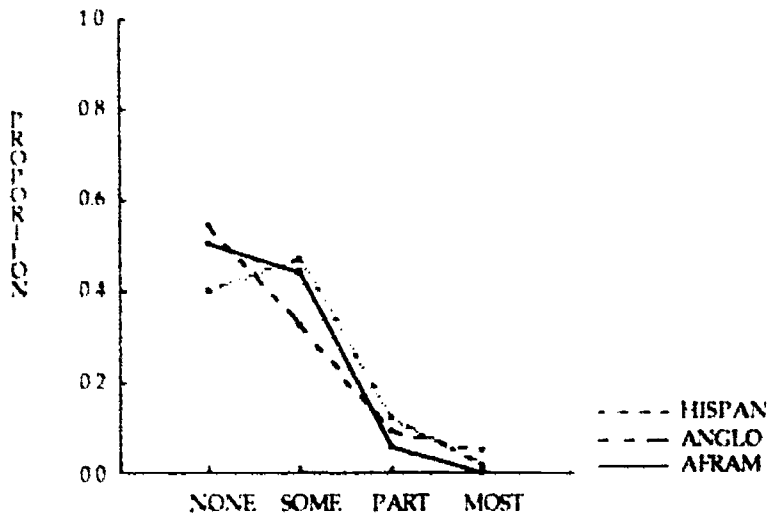
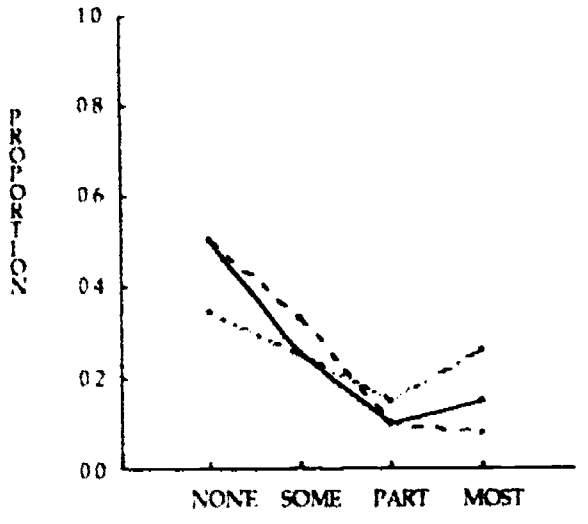
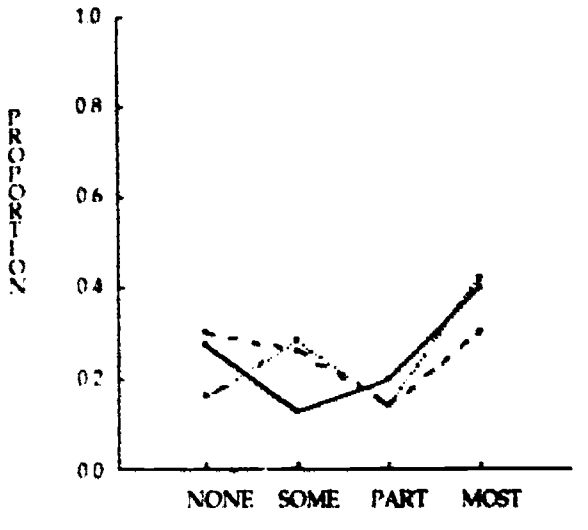
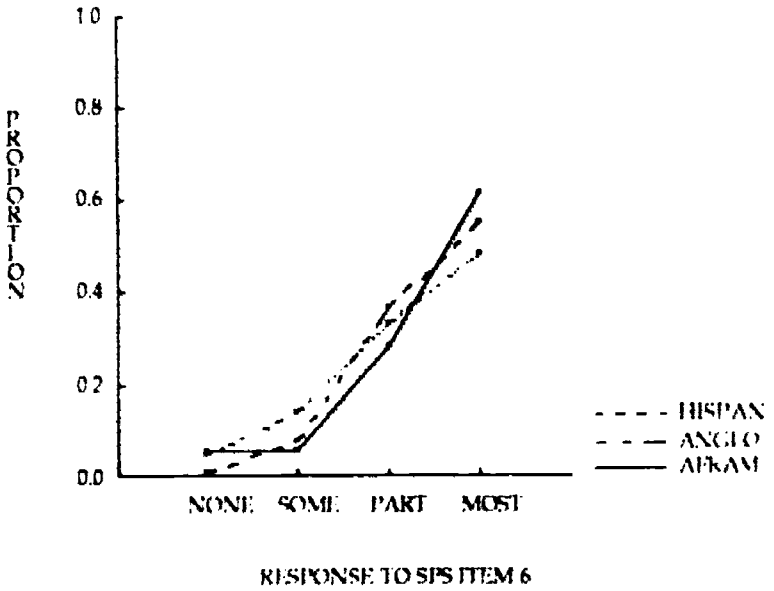


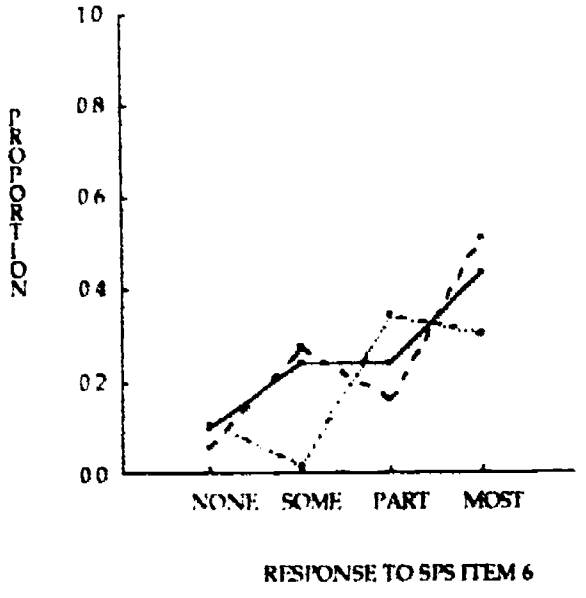
Figure 1. Proportion of three ethnic groups (HISPANIC, ANGLO, AFRICAN AMERICAN), at three mental health levels (NORMAL, PSYCHIATRIC INPATIENT, SUICIDE ATTEMPTER), selecting each response alternative of SPS Items 5, 6, 12, and 36.











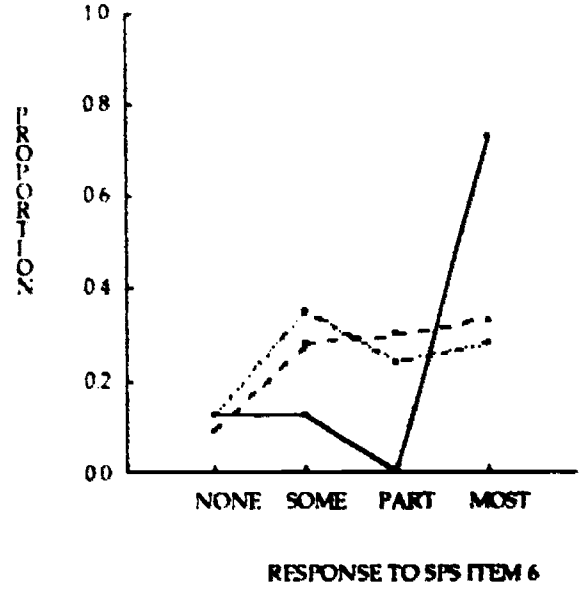


Figure 1 continued

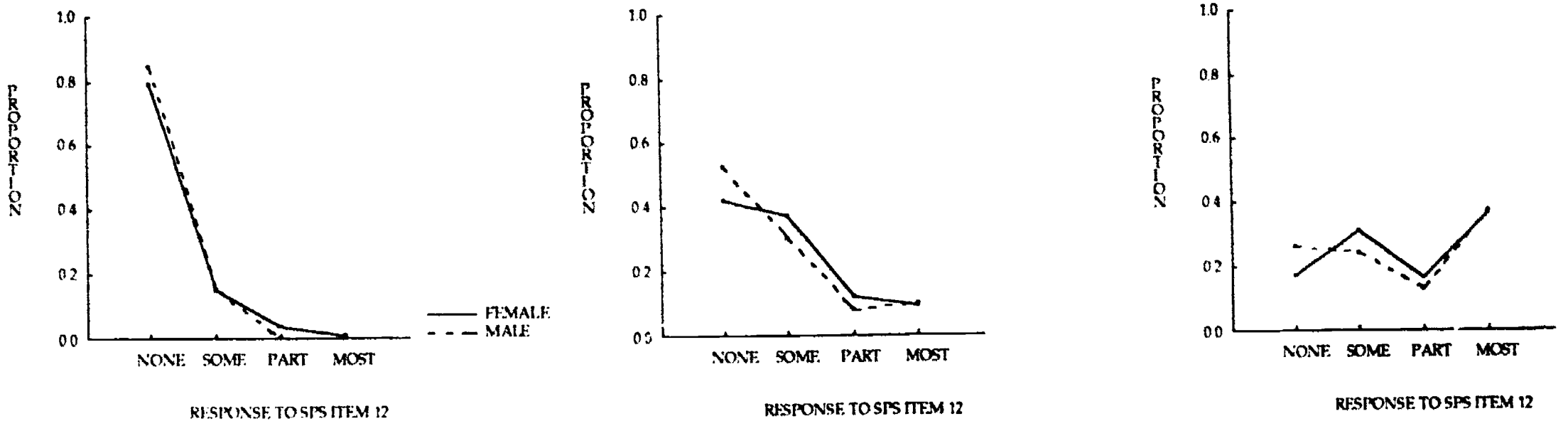
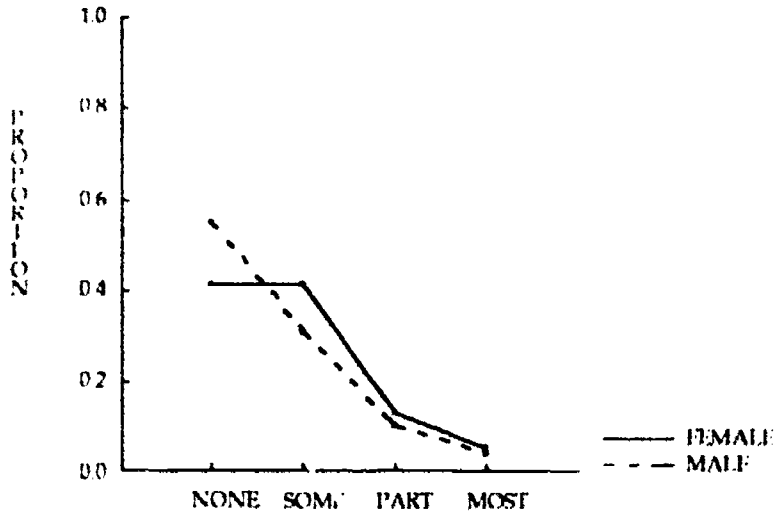
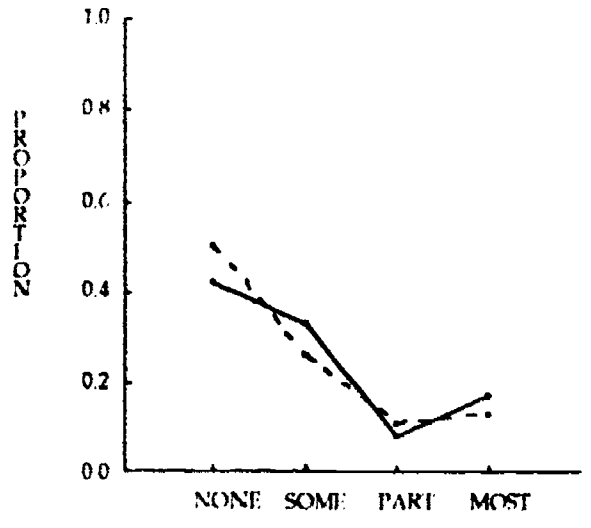
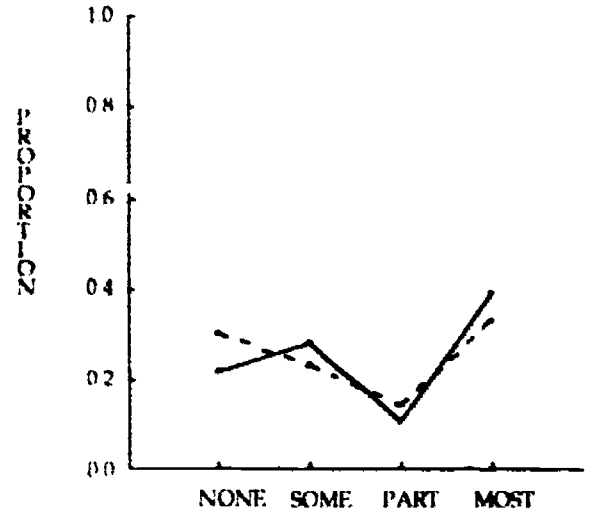
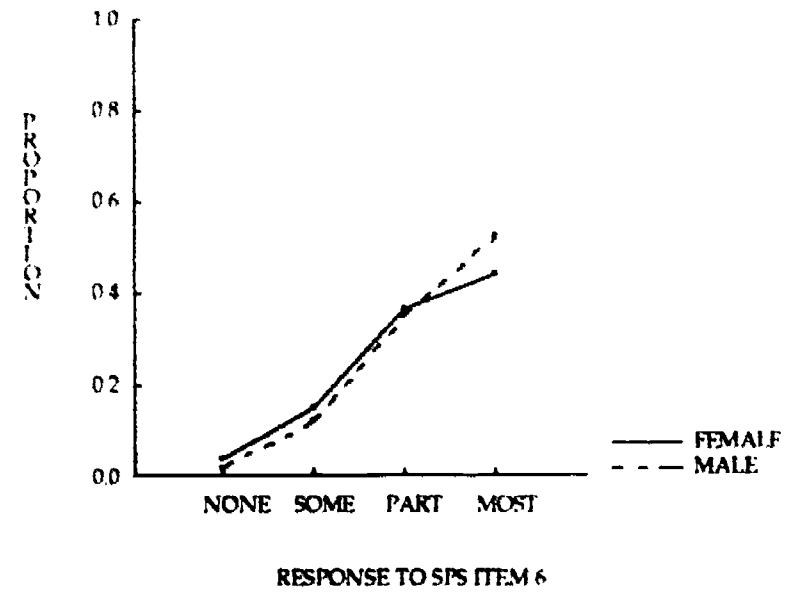


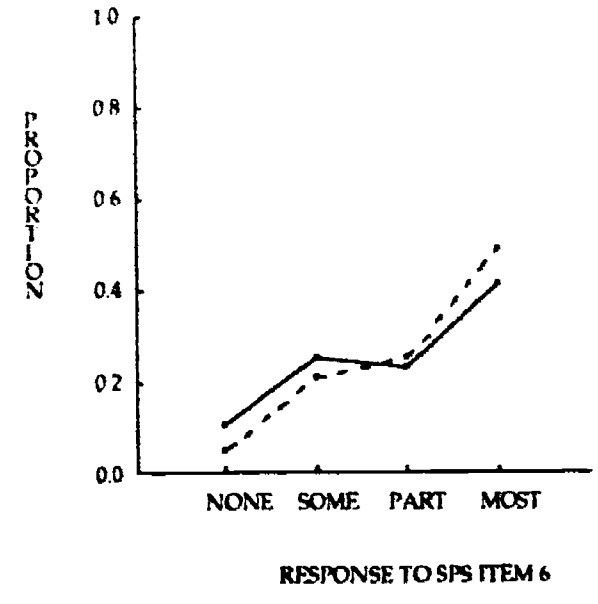
Figure 2. Proportion of females and males, at three mental health levels (NORMAL, PSYCHIATRIC INPATIENT, SUICIDE ATTEMPTER), selecting each response alternative of SPS Items 5, 6, 12, and 36.











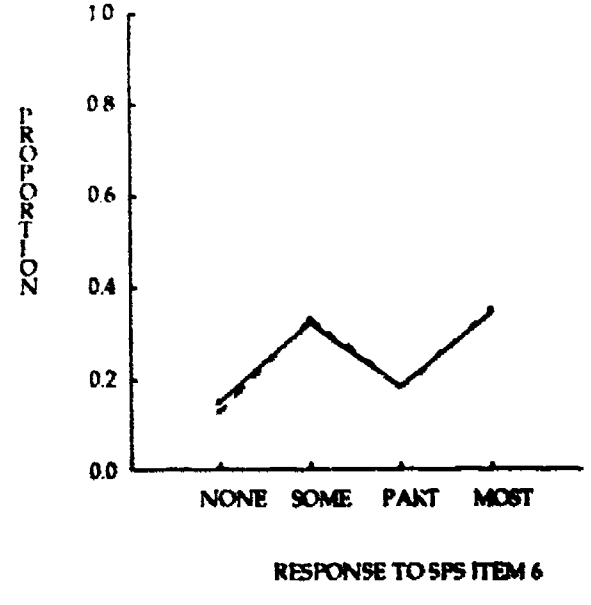


Figure 2 continued