

Social Problem Solving Scale
Kindergarten / Year 1
Fast Track Project Technical Report
Prepared by Ernie Valente (615)343-8858
January 16, 1995

1. The Social Problem Solving Scale

1.1. *Scale Description*

The Social Problem Solving Measure (SPS) has been administered as part of the Summer child interview in all years of the FAST Track Project to date. It is available for children in all Cohorts, but there have been changes in the way the SPS has been administered from year to year.

The child interviewer shows the respondent a series of eight drawings that depict two types of social situations, and reads a brief description of the situations. In four of the situations the actor is a solitary child who would like to join in a social activity with other children (e.g. solitary child would like to play with other children on jungle gym, solitary child would like to join in a game of tag). In these situations the child's task is to **solicit social involvement**. The other four situations involve a child who is being teased or frustrated by another child (e.g., other child won't share a swing, other child pushes the child out of line). In these situations the child's task is to **resolve a social conflict**. The social situation drawings show stick" figures so that the stimulus children are race and gender neutral.

The interviewer reads a script describing each social situation, and asks the respondent to take the role of one of the actors. An example of the solicit social involvement situation:

[The interviewer shows a picture of several children playing on a jungle gym. Another child stands apart from those playing on the gym, looking at those playing] "Pretend that this is YOU..." [interviewer points to solitary child] "...and that this is COLLEEN/JOSH" [interviewer points to one of the children on the jungle gym]. "COLLEEN/JOSH and some other kids are playing on the jungle gym at school. YOU would like to play with COLLEEN/JOSH and the other kids, but they haven't asked you. What could YOU say or do to get to play with COLLEEN/JOSH and the other kids?"

The interviewer elicits one response from the child, then prompts for two others:

PROMPT 1: "What's another thing you could do or say so that YOU could get to play with COLLEEN/JOSH and the other kids?"

PROMPT 2: "Can you think of anything else that YOU could do or say that YOU could get to play with COLLEEN/JOSH and the other kids?"

The interviewer followed a similar procedure for three pictures showing social conflict situations. For example:

[The interviewer shows a picture of several children playing] "Pretend that this is YOU..." [interviewer points to one of the playing children] "...and that this is TRACY/TREVOR." [Interviewer points to another child] "YOU and TRACY/TREVOR are playing a game and you realize that TRACY/TREVOR has taken your turn. What could YOU do or say so that YOU could get to take your turn?"

PROMPT 1: "What's another thing YOU could do or say so that YOU could get your turn. PROMPT 2: "Can you think of anything else that YOU could do or say so that YOU could get your turn."

In earlier years, the child was prompted for up to six responses to each situation, but the number of responses was reduced to three in 1993 and subsequent years.¹

The interviewer is given instructions about what to do if the respondent gives fewer than three responses or repeats the same response. If the respondent gives fewer than three responses, the interviewer repeats prompt 2. If the respondent again fails to give a valid response, the interviewer enters a code indicating that the respondent is unable to provide further novel responses and moves on to the next picture in the series. Similarly, if the respondent repeats verbatim a previous response, the interviewer does not record the verbatim response and repeats prompt 1. If the respondent again repeats a previous response, the interviewer proceeds to the next picture in the series without recording any repeated responses.

A similar situation occurs when the respondent gives the same response worded in different ways. For example, the respondent may say "I'd be nice," then give "I'd be really nice" as the next response. After two such responses, the interviewer is instructed to say something like: "Well, being nice is one thing you could do, what's something different you could say or do?" If the child gives a different response, the interviewer records it. If the child again gives the same response the interviewer

¹Results from the 1991 Summer interview showed that most children gave three or fewer responses even when prompted for more.

codes the multiple, repeated responses as one response, and proceeds to the next picture in the series.

Interviewers are instructed to probe for more information in cases where the respondent says he would seek out an adult. These responses must be clarified because they may be coded into several categories (see below). Interviewers are also instructed to probe for clarification when the response is vague (e.g., "I'd be nice," "I'd be mad").

1.2. *Subscale Scores*

The SPS is scored in a very complicated way involving a series of steps:

- (1) The child's verbal response describing what he would do in the situations represented in the eight pictures is coded by the interviewer into one of six categories (aggressive, competent, authority-punish, authority-intervene, passive/inept, irrelevant/other). These categories are described in Table 1. Please note that if a response was judged to fit into more than one category, the interviewers were instructed to code it into the appropriate category with the lowest code number. The category codes were assigned so that categories with greater theoretical importance were assigned lower code numbers.
- (2) The total number of responses to each picture in each category (e.g., aggressive) is calculated.
- (3) The total number of valid responses to each picture is calculated. (Valid responses are those coded into any category. Inept or irrelevant responses are valid by the SPS coding scheme. Invalid responses aren't coded -repetitive responses are an example of invalid responses).
- (4) The percentage of responses in each category to each picture is calculated by dividing the number of responses in each category by the total number of valid responses to each picture. These scores are called Picture response *percentages* below.
- (5) The mean of each response category percentage over the eight pictures is calculated, and this score becomes the subject of most subsequent analyses, although the scores at all steps in this process are available in FAST Track SPS datasets. These scores are called *Mean-percentages across pictures* below.

Table 1. *SPS Response Categories*

| Code | Category | Description |
|------|-------------------------------------|---|
| 0 | Aggressive | All forms of physical or verbal aggression are scored in this category, also threats of any kind. |
| 1 | Competent | All responses that suggest a direct, socially appropriate way of handling the situation are scored in this category. Some forms competent responses can take are asking, making deals, bribing, assertively telling, sharing, trading, joining in, etc. |
| 2 | Authority-Punish | Responses that are scored in this category are appeals to some authority figure to punish the provocateur in some way. Authority figures include parents, teachers, other school personnel, and older relatives. |
| | Authority-Intervene | Responses coded into this category are appeals to authority figures to intervene on the child's behalf, rather than punishing the provocateur |
| | Passive/Inept | Responses that indicate a passive or nonassertive response to the depicted situation. |
| | Irrelevant/Other | This category is for nonsense and other responses that don't fit into the other categories |
| | Unable to Provide Further Responses | Used after the child has been issued both prompts and cannot come up with a novel response. |

In summary, respondents' open verbal responses to the eight SPS pictures are categorized into six response categories. The percent of responses in each of these categories to each of the eight pictures is calculated (picture response percentages). Finally, the mean of the eight percentages for each response category is taken. These mean-percentages across pictures are typically analyzed.

2. SPS Reliability and Validity

2.1. *SPS Reliability*

This reliability analysis is based on the 355 complete SPSs from the Cohort 1 normative sample collected in the Summer of 1993.

Because there are several steps in the calculation of SPS mean-percentage across pictures scores, there are two choices for the subject of reliability analyses: the raw counts of responses of each type to each picture (these result from scoring step 3), and the picture response percentages (from scoring step 4). It seems reasonable to focus on picture response percentages, as they are frequencies that have been corrected for total number of responses, and are thus comparable across pictures. There are eight picture response percentages for each of the six response types, one for each stimulus picture. In the scoring procedure, an index is formed by taking the mean of these eight percentages (scoring step 5), so the eight picture response percentages are considered items in a subscale measuring the construct represented by that response category.

Percentages, the subject of this item analysis, have non-normal distributions that can affect the accuracy of estimates of covariability, such as Chronbach's alpha. An inspection of the distributions of the SPS picture response percentages showed all but those for competent responses to be severely negatively skewed. Spearman correlations, which are more appropriate for non-normal data than Pearson correlations, are reported below.

Table 2. *SPS Subscale Inter-Item Reliability Coefficients*

| <u>Response Type/Subscale</u> | <u>Reliability Coefficient</u> |
|-------------------------------|--------------------------------|
| Aggressive Competent | .71 |
| Authority-Punish | .70 |
| Authority-Intervene | .40 |
| Passive-Inept | .55 |
| Irrelevant/Other | .57 |
| | .40 |

Cronbach's alphas for the six subscale scores are shown in Table 2. Spearman SPS inter-item correlations were compared with Pearson SPS inter-item correlations to gauge any bias in the reliability coefficients due to the non-normal distributions of the items. The Spearman correlations were nominally smaller than the Pearson correlations, suggesting that the reliability coefficients in Table 2 may be mildly inflated. Cronbach's alphas for the Aggressive and Competent response types are acceptably high, about .70. Coefficients for the other four subscales (Authority-Punish, Authority-Intervene, Passive/Inept, Irrelevant/Other) indicate that these subscales are unacceptably unreliable.

Simple Statistics and inter-item Spearman correlations for the SPS subscales are presented in Appendix A. Some relevant information taken from Appendix A is shown in Table 3. Table 3 shows the ranges of the means of the eight picture response percentages for each of the six responses categories. These means are based on the picture response percentages of 356 children in the Normative sample. Table 3 also shows the distribution of inter-item Spearman correlations for the eight picture response percentages of each response category.

Table 3. *Ranges of Means and Distribution of Inter-Item Spearman Correlations for Picture Response Percentages in Six SPS Response Categories.*

| Category | Mean Range | Dist. of Inter-Item Correlations | | | | |
|----------------|------------|----------------------------------|-------|-------|-------|-----|
| | | <.1 | .1-.2 | .2-.3 | .3-.4 | >.4 |
| Aggressive | .02- .23 | 0 | 5 | 16 | | |
| Competent | .45- .95 | 3 | 9 | 9 | | |
| Auth-Punish | .00- .12 | 12 | 11 | 4 | | |
| Auth-Intervene | .01- .16 | 8 | 12 | 3 | | |
| Passive/Inept | .02- .10 | 9 | 13 | 4 | | |
| Irrelevant | .00- .01 | 16 | 0 | 3 | | |

The distribution of Spearman inter-item correlations in Table 3 shows that the most of the items comprised by the Aggressive and Competent subscales are at least somewhat intercorrelated, while the items comprised by the other subscales tend not to be intercorrelated. This is consistent with the pattern of

reliability coefficients shown in Table 2.

There is an apparent cause for the low inter-item correlations. The mean-percentages in Table 3 show that Aggressive and Competent responses are given with a reasonably high frequency, but that responses in other categories are rarely given. The low frequency of responding in the Authority-Punish, Authority-Intervene, Passive/Inept and Irrelevant categories is likely the ccj.se of the unreliability of the associated subscales. An examination of the post-kindergarten (Summer, 1991) mean-percentages show similarly small means for these response types, suggesting that the low frequencies aren't simply due to the respondents "outgrowing" the scale. While the group mean of the mean-percentages across pictures of these subscales is small, inspection of the distribution of the individual mean-percentages revealed that most children rarely use the categories, but about 25% of the children use the categories fairly frequently (5%-40% of the time) .

In summary, the Aggressive and Competent subscales of the SPS appear to be adequately reliable. The other four subscales of the SPS are unreliable by classical criteria. Further item analysis suggests that the unreliability of the Authority-Punish, Authority-Intervene, Passive/Inept, and Irrelevant/Other SPS subscales may be due to the fact that these are, on average, rarely used categories. While most children rarely use these categories, a small group of children uses them fairly frequently.

The distribution of the Aggressive subscale scores is highly positively skewed, and the Competent subscale distribution is somewhat negatively skewed. The skewness in these variables suggests the distributions would be improved by square-root or logarithmic data transformations.² Information about the distributions of the raw and transformed variables is shown in Appendix B. It appears that the square-root transformation of the Aggressive subscale scores improves their distribution somewhat, but they remain negatively skewed. The raw Competent subscale scores appear to have the best-conditioned distribution, as neither transformation provides much improvement.

2.2. SPS Validity Analysis

This validity analysis is based on the mean-percentages across pictures for the SPS Aggressive and Competent subscales collected from 355 Cohort 1 normative children in the Summer, 1993. The other subscales are excluded from much of this validity analysis because of their unreliability.

²These scores may have a value of zero, so one was added to the raw scores before square-roots or logarithms were taken.

The SPS Aggressive mean-percentage across pictures is highly-correlated with the Competent score (Spearman $r = -0.61$, $p < .001$). The SPS Aggressive score mean is .10, and its median is .08. The Competent score mean is .73 and its median is .75. These results suggest that, on average, the respondents gave either Aggressive or Competent SPS responses, and that the Aggressive and Competent mean-percentage scores contain much redundant information.

Table 4 shows the correlations among the SPS Aggressive and Competent mean-percentage across pictures subscale scores and the SHP mean subscale scores (correlations for the other subscales are shown in Table 7 below). The Aggressive score is positively and significantly correlated with all of the SHP scores except the Social Contact score. The Competent subscale score is negatively correlated with all but the Social Contact score (remember that the SHP is scored so that a higher score indicates more frequent behavior problems).

Table 4. *SPS Aggressive and Competent Subscale Criterion Validity Correlations with Spring, 1993 Social Health Profile Subscale Mean Scores.*

| | SC | TOCA | SHP | SHP Subscale AA | CC |
|----------------|-------|-------|------|-----------------|-------|
| SPS Aggressive | .15" | .19" | .04 | .17" | .18" |
| SPS Competent | -.15" | -.22" | -.05 | -.18" | -.19" |

NOTE: AA= Authority Acceptance, CC= Cognitive Concentration, SC= Social Contact, TOCA= TOCA 14 Item Subset, SHP= All SHP Items. **= $p < .01$. These are Pearson correlations using data of Cohort 1 Normative Children. Correlations based on square-root of SPS Aggressive score.

The pattern of correlations in Table 4 provides evidence for the criterion validity of the SPS Aggressive and Competent subscales. Children who give, on average, a higher proportion of aggressive responses on the SPS tend to get higher teacher-reported problem behavior scores on the SHP. Children who give competent responses on the SPS tend to get lower problem behavior scores on the SHP.

An Intervention vs. Control comparison of 1993 Competent scores shows a significant difference after controlling for site, $F(1,278) = 7.33$, $p < .01$. Intervention children give a higher proportion of Competent responses at all sites but Nashville.

3. Longitudinal Analysis

The previous analysis concerned data from the Summer, 1993 administration of the SPS. These data were chosen because they were the most current at the time of this report. Data from 1991 and 1992 were also available to investigate the longitudinal characteristics of the SPS.

3.1. Longitudinal Means and SDs

Appendix C shows the means and standard deviations of the six SPS mean-percentage subscale scores for 1991, 1992 and 1993 normative sample. The means for the Competent subscale tend to nominally increase each year (1991= .64, 1992= .69, 1993= .73), while the means for the Aggressive (1991= .16, 1992= .14, 1993= .10) and Authority-Punish (1991= .06, 1992= .04, 1993= .03) subscales decrease. The means for the other subscales stay about the same. This suggests that in earlier years, respondents use somewhat more Aggressive and Punish responses that tend to be replaced by Competent responses in later years. These are not large differences.

3.2 Longitudinal Inter-Item Reliabilities

Table 5 shows the inter-item reliability coefficients (Chronbach's alpha) for each of the SPS subscales in 1991, 1992, and 1993.

Table 5. *Comparison of 1991, 1992 and 1993 SPS Subscale Inter-Item Reliability Coefficients*

| | Chronbach' s Alpha | | |
|------------------------|--------------------|------|------|
| Response Type/Subscale | 1991 | 1992 | 1993 |
| Aggressive | .75 | .76 | .71 |
| Competent | .76 | .71 | .70 |
| Authority-Punish | .69 | .59 | .40 |
| Authority-Intervene | .61 | .58 | .55 |
| Passive-Inept | .74 | .68 | .57 |
| Irrelevant/Other | .71 | .59 | .40 |

The coefficients for the Aggressive and Competent subscales are about the same in all three years, while those for other scales decrease each year. This pattern is perhaps explained by the fact that the four subscales that decrease tend to be less frequently used in later years, leading to less variability and lower coefficients.

3.3. SPS Stability Correlations

Table 6 shows the stability correlations of the six SPS subscales over 1991, 1992, and 1993. The Aggressive and Competent subscales of 1991 have small stability correlations with data from other years, but the scales become more stable across 1992-1993. The other four subscales start with small stability correlations and get less stable in later years.

Table 6. *SPS Subscale 1991, 1992 and 1993 Stability Correlations*

| Response Type/Subscale | 91-92 | 91-93 | 92-93 |
|------------------------|-------|-------|-------|
| Aggressive | .21** | .21** | .30** |
| Competent | .17** | .22** | .27** |
| Authority-Punish | .22** | .12* | .27** |
| Authority-Intervene | .22** | .12* | .12* |
| Passive-Inept | .11* | .08 | .06 |
| Irrelevant/Other | .19** | .14* | .02 |

Note: These are Spearman correlations using data of Cohort 1 Normative Children. Correlations are based on untransformed SPS subscale scores. * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

One possible explanation for the increased stability of Aggressive and Competent scores may be due to decreased measurement error in these data for older children. The reduction in stability of the other subscales may be explained by their poor reliability in later years.

3.4. Longitudinal Criterion Validity Correlations

Table 7 shows the 1991, 1992, and 1993 criterion correlations of the six SPS subscales with the subscales of the SHP administered about the same time.

Table 7. *1991, 1992 and 1993 SPS Criterion Validity Correlations with Social Health Profile Subscale Mean Scores.*

| SPS Subscale | SHP Subscale | | | | |
|--------------|--------------|---------|------|-------|-------|
| | AA | CC | SC | TOCA | SHP |
| Agg 1991 | .03 | .07 | .05 | .07 | .08 |
| 1992 | -.00 | .01 | -.02 | -.01 | -.00 |
| 1993 | .15" | .16" | .02 | .15" | .16" |
| Comp 1991 | -.14" | -.13" | -.08 | -.16" | -.16" |
| 1992 | -.09 | -.07 | -.02 | -.09 | -.09 |
| 1993 | -.17" | -.21*** | -.03 | -.17" | -.19" |
| Pun 1991 | .11* | .05 | .03 | .08 | .09 |
| 1992 | .09 | .17" | .03 | .14" | .14" |
| 1993 | .20** | .16" | .07 | .20" | .18" |
| Int 1991 | .04 | -.05 | .01 | .00 | -.02 |
| 1992 | -.02 | -.09 | -.05 | -.05 | -.05 |
| 1993 | .00 | -.02 | -.05 | -.01 | -.02 |
| Pas 1991 | .06 | .02 | -.03 | .05 | .04 |
| 1992 | .08 | .02 | .00 | .07 | .04 |
| 1993 | -.02 | .09 | .02 | .01 | .05 |
| Irr 1991 | -.03 | .08 | .02 | .02 | .05 |
| 1992 | .13* | .12* | .11* | .14" | .14" |
| 1993 | .02 | .04 | .13* | .02 | .04 |

NOTE: SPS - Agg = Aggressive, Com = Competent, Pun = Authority-Punish, Int = Authority-Intervene, Pas = Passive, Irr = Irrelevant SHP - AA= Authority Acceptance, CC= Cognitive Concentration, SC= Social Contact, TOCA= TOCA 14 Item Subset, SHP= All SHP Items.
 These are Spearman correlations using data of Cohort 1 Normative Children, Correlations based on untransformed SPS subscale scores. 1991 SPS was correlated with Fall, 1991 SHP; 1992 SPS with Spring, 1992 SHP; and 1993 SPS with Fall, 1993 SHP. *= p<.05; **= p<.01; ***= p<.001.

The correlations in Table 7 suggest small but stable criterion validity correlations for the Competent subscale. The Aggressive and Authority-Punish subscales do not correlate well with the criterion measure except in the 1993 data.

3.5. Longitudinal Intervention - Control Comparisons

We examined preliminary longitudinal Intervention-Control differences in SPS scores by running separate univariate ANOVAs of Aggressive and Competent subscale scores for each cohort for each year of available data. The square-root of the Aggressive scores and the raw Competent scores were the dependent variables, and site effects were statistically eliminated, in the models.

Cohort 1 SPS Aggressive scores of Intervention and Control children did not significantly differ after kindergarten (Summer, 1991), $F(1,298) = 0.46$, n.s. There were significant differences after grade 1 (Summer, 1992), $F(1,290) = 8.84$, $p < .01$, and grade 2 (Summer, 1993), $F(1,278) = 3.52$, $p < .06$; Intervention children gave, on average, a lower proportion of Aggressive responses than Control children.

We found a similar pattern for Competent scores; this wasn't surprising given the high negative intercorrelation of Aggressive and Competent scores. There weren't differences in Competent scores after kindergarten, $F(1,298) = 0.29$, n.s., but there were significant Intervention-Control differences in 1992, $F(1,290) = 10.38$, $p < .001$, and 1993, $F(1,278) = 7.33$, $p < .01$.

Analysis of the data for Cohort 2 did not show any differences for either score for either year.

Appendix D shows graphs of the Aggressive and Competent score means for each cohort over several years.

4. Conclusions and Recommendations

The SPS generates several different types of subscale scores, but the most appropriate for end analysis are the *mean-percentages across pictures*. Two mean-percentage subscale scores, Aggressive and Competent, have stable psychometric characteristics and are recommended for use in any analysis, especially large models containing constructs measured by the SPS. The other four subscale scores have unstable psychometric characteristics and should only be used in careful, limited analyses when other, better measures of their constructs are not available.

4.1. *Using the SPS Aggressive and Competent Scores*

The SPS has two subscale scores - Aggressive and Competent - that are reliable and valid by classical criteria. The two scores are highly negatively correlated (-.61), suggesting they contain redundant information, and probably should not be used in the same analysis. They are not so highly correlated that less competent responding implies more aggressive responding, so the Aggressive score cannot be replaced by the Competent score in analyses focussing on aggressive responding.

The distribution of Competent scores is positively skewed, but closer to normal than that of Aggressive scores. The raw Competent scores can be used directly in OLS analyses. A square-root transformation of the Aggressive scores improves their distributional characteristics, but the transformed data are still far from normally distributed.

4.2. *Using Other SPS Scores*

Other SPS subscale scores (Authority-Punish, Authority-Intervene, Passive/Inept, Irrelevant/Other) are not appropriate for analysis using OLS procedures. They can, however, be dichotomized and included in nonlinear (e.g., logit) models. For example, I assigned a binary variable to each child that took a value of one if their mean-percentage Authority-Intervene subscale score was above .1 (meaning they used the category ten percent of the time or more, on average), zero otherwise. I then used PROC CATMOD in SAS to test whether the proportion of Intervention children who tended to use this category differed from the proportion of Control children who did so.³

These subscale score variables must be examined more closely to determine optimum cut-off points for the dichotomization before binary versions can be included in further analyses. Psychometric characteristics of the binary variables cannot be presented here because this has not yet been done.

4.3. *Summary*

Table 5 summarizes our recommendations about how to use the six SPS subscales. The table shows the subscale name, the SAS variable name, the "best" form of the variable (e.g., the distribution with the most information), and the recommended use of the subscale (Any or Limited analysis, Ordinary Least-Squares or Logistic models).

I repeated this procedure for the other subscales and found no significant differences for any of the four subscales.

Table 5 *Summary of Recommendations about SPS Mean-Percentage Across Pictures Subscale Scores*

| Subscale | SAS Var Name | Best Form | Rec. Use |
|----------------|--------------|-----------|------------------|
| Aggressive | MCxBAGP | Sq. Rt. | Any-OLS |
| Competent | MCxBCOP | Raw | Any-OLS |
| Auth-Punish | MCxBPUP | Binary | Limited-Logistic |
| Auth-Intervene | MCxBINP | Binary | Limited-Logistic |
| Passive-Inept | MCxBPAP | Binary | Limited-Logistic |
| Irrelevant | MCxBIRP | None | Do not use |

In summary, the Aggressive and Competent subscales of the SPS appear to be reliable, valid, and sensitive measures of child social cognition. The utility of the other four subscales has not yet been demonstrated. In the future, it may be reasonable to reduce the number of response categories because of the low frequency of responding in categories other than Aggressive and Competent.