3. METHOD

The present study employs data of the "Self-Regulation and Social Relations" project, a cross-national study on action-control in middle childhood and adolescence. The following sections provide a brief description of the general assessment procedure of the project, the participants, and the measures that were used in the present study. Subsequently, the preparation of rawdata for the analyses and the general statistical procedures of this study are described.

3.1 General Assessment Procedures

The data have been collected as part of more intensive data collections within a cross-national study assessing children and adolescents (grades 3 - 10) in Berlin, Moscow, and New Haven. The cross-sectional study was conducted within the "Self-Regulation and Social Relations" project (Principal Investigators: Todd D. Little and Lothar Krappmann). The project represented a subproject of the Action Control and Child Development Project at the Max Planck Institute in Berlin (Principal Investigators: T. D. Little and P. B. Baltes). The project is continued as the "Agency in Development Project" (Principal Investigator: Todd D. Little) at the Yale University.

The data were collected in three inner-city primary schools in the city of Berlin in fall 1996, using a cluster sampling approach. The participants represent almost the entire child population of these age groups living in the neighborhood around the three schools, because in Berlin all children living in a neighborhood usually attend the same primary school. The three schools were located in an eastern district which received the rank number of 20 in the overall ranking of the 23 districts of the city of Berlin. Generally, the district with the highest rank (i.e., 23) has the highest level of social problems (cf., Hermann, Imme, Meinlschmidt, 1998). On a more fine-grained level, the districts of the city of Berlin can be decomposed into smaller subdistricts or areas which can be further characterized by rankings according to social problems. Thereby, the highest rank number of 7 indicates the highest level of social problems. All three schools were located in sub-districts that received a rank number of 6

(Hermann, et al., 1998) and, thus, were comparable with regard to the social structure[39]. Only 2 % of the children belonged to ethnic-minority groups (mainly Turkish).

3.2 Participants

Overall, 882 children represented the total population of students in the investigated 39 classrooms of the three schools. 681 (77.1 %) parents provided permission for participation of their children in the study. Thus, the overall participation rate can be considered to be of acceptable size. In total, 39 children (5.7%) of the 681 children (7.9 - 13.5 years, M age = 10.3 years, SD = 1.3) had to be removed from the analyses for various reasons. First, eight children (1.2%) of a classroom missed the first part of the testing session due to school assignments and, therefore, their data were incomplete. Second, four (0.6%) children refused to participate because they were new in school. Third, two (0.3%) children's responses were unreliable because they were handicapped; two (0.3%) children's responses were unreliable because they had serious problems to understand the German language because they belonged to one of the ethnic minoriy groups. Moreover, four (0.6%) children were excluded from the analyses because their responses were unreliable for reasons such as lacking motivation to participate in the study. Finally, children had to be excluded from the analyses if there were no valid responses of a single nominated friend available (see also Points 3.4.2 and 3.4.3). Specifically, five (0.7%) children had to be excluded from the study because they had not nominated a single friend who also attended the same school[40]. Moreover, 14 children (i.e., 11 boys and 3 girls) nominated friends who attended the same school but their friends did not

[39] The reported rankings of the social structure of the city of Berlin refer to an index of social structure that is based on various indicators of demographical, educational, professional, income, and health information of the population of the specific districts. The planning of data collection was based on information about the social structure of the city of Berlin assessed in year 1994 and which was restricted to the overall rankings of the 23 larger districts of Berlin according to social structures (Hermann & Meinlschmidt, 1995; Statistisches Landesamt Berlin, 1995). However, the overall rank ordering of the 23 districts remained rather stable between the years 1994 and 1997 (r = .95, cf. Hermann, et al., 1998). In the year 1994 (two years before the data were collected) as well as in the year 1997 (one year after the data were collected), the district received the same rank number (i.e., 20). The rankings of the three sprecific subareas of the district where the schools were located were only available for the year 1997.

[40] In terms of sociometric status one child was neglected, two children were unclassified, one child was rejected, and one child was average.

participate in the study[41]. The in total 39 children who were removed from the analyses did not significantly differ from the final sample with regard to mean age (8.2 - 13.5 years, M age = 10.2 years, SD = 1.4); $t_{(679)} = -0.33$, $\eta^2 = .00$, $p = .48$).

The final sample was comprised of 642 students (7.9 - 13.2 years, M age = 10.3 years, SD = 1.2): 183 from Grade 3 (7.9 - 11.0 years, M age = 8.9 years, SD = 0.4); 152 from Grade 4 (8.8 - 10.8 years, M age = 9.8 years, SD = 0.4); 157 from Grade 5 (9.8 - 12.7 years, M age = 10.9 years, SD = 0.5), and 150 Grade 6 (11.1 - 13.3 years, M age = 11.9 years, SD = 0.4). A summary of the sample sizes by grade levels and gender for the overall sample and by school is presented in Appendix A, Table A1.

3.2.1 Testing the Effects of the Different School Contexts

The implemented sampling procedures randomly selected the schools as units of observations. In contrast, the children represented the units of the conducted analyses. The children within the schools can't be considered to represent independent observations (see Cronbach, 1976). Hence, assessment of systematic differences in children's responses related to the different school contexts were recommended. Appendix B provides information about the comparisons of (a) psychometric comparability, (b) correlational patterns, (c) mean levels, and (d) variances of the assessed constructs across the three school contexts. The results of these analyses did not yield any systematic differences in any of the parameters across the three schools. The single exception was that in one of the school contexts the number of correlations which differed significantly from the correlations evinced in the remaining school contexts exceeded the number of significant differences that would be expected by chance. In sum, the results of the present study, generally, can be considered to be unbiased by effects that are related to differences in socialization practices across the three school contexts.

[41] In terms of sociometric peer status three children were unclassified, one child was neglected, four average, and six children were rejected. Children who had to be excluded because their friends did not participate were not equally distributed across sociometric groups; $\mathbf{X^2_{(5)}} = 13.89$, $p = .02$. Specifically, 5% of the rejected and 5% of the neglected children had to be excluded from the analyses because their nominated friends did not participate. Four of these rejected children nominated only one single friend while the other two rejected children and the neglected child nominated three friends who did not participate in the study. In contrast, only 1% of the average children and 1% of the unclassified children, and none of the popular and controversial children nominated friends who did not participate.

3.2.2 Assessment Procedures

At three subsequent sessions the children responded to packages comprised of two questionnaires each. There was about one week between each time of assessment. Each of the sessions was divided into two parts. The two parts of each session were separated by a break of twenty minutes. In each part of a specific session one of the two questionnaires was administered. The ordering of the two questionnaires was counterbalanced within the sessions.

At the first two sessions alternately children's perceived control about school and perceived control about friendship were assessed. The ordering of assessment of perceived control about school and perceived control about friendship was counterbalanced. Classrooms were assigned randomly to order of assessment but with the constraint that, to the extent possible, equal numbers of classrooms within school and grade be assigned to each order. Children's perceived control about school was assessed with the Multi-CAM for School (Little & Wanner, 1997). Moreover, in this session a speeded short version of the RAVEN's Standard progressive Matrices (SPM, Raven, 1971) developed by Little and Wanner (1997) measuring children's fluid intelligence was administered. Children's perceived control about friendship and their strategies to act on friendship problems were assessed with the Multi-CAM for Friendship (Little & Wanner, 1997). Furthermore, in this session children's perceptions of friendship quality and quantity were assessed with the Friendship Inventory (Little, Krappmann, Brendgen, & Wanner, 1997). This questionnaire also included the instructions for the sociometric procedures assessing social acceptance and aggression. The ordering of the assessment of the Multi-CAM for Friendship and the Friendship Inventory was counterbalanced. Again, as has been done when counterbalancing the ordering of assessment of perceived control about school and friendship, classrooms were assigned randomly to order of assessment but with the constraint that, to the extent possible, equal numbers of classrooms within school and grade be assigned to each order. In the last session, (a) the Big-Five Personality Questionnaire for Children (B5P-C; Little & Wanner, 1997), (b) a social desirability scale (Rost & Hartmann, 1993), and (c) a questionnaire assessing well-being at school (I-Feel Questionnaire, Little, 1997) were administered.

The measures of perceived control about friendship, action strategies, friendship quality, sociometric measures, social desirability, and intelligence are used in the present study and are described below. The measures of perceived control about school, personality, and well-being were not included in the present study.

3.2.2.1 Testing the Effects of the Ordering of Administering the Questionnaires

Appendix C presents the results of testing whether the ordering of the sessions (academic vs. friendship) and the ordering of the questionnaires assessed in the "Friendship Session" (Multi-CAM for Friendship vs. Friendship Inventory) affected children's responses. The conducted comparisons addressing the psychometric comparability of the investigated constructs as well as invariance of the factor variances, means, and correlations did not yield any systematic differences across the assessment conditions. Hence, the results of the present study can be considered to be unbiased by effects that are related to differences in the ordering of administering the questionnaires.

3.2.2.2 Administration Procedures

The questionnaires were group administered in the classrooms during class periods and in the teachers' absence. The presentation time of the Multi-CAM for Friends varied between 30 minutes for younger children and 15 minutes for older children. The presentation time of the Friendship Inventory varied between 50 minutes for younger children and 30 minutes for older children. All participants had parental permission. In a separate room of the school, a trained proctor of the assessment teams took care for the children whose parental permission was missing.

The trained assessment teams consisted of an interviewer and one or two proctors. The teams were comprised of two proctors when assessing younger children (grades 3 and 4) and one proctor when assessing older children (grades 5 and 6). The proctors' task was to assist the children with questions or problems when answering the questionnaires. After a short introduction by the interviewer the instrument was read aloud while the children followed silently along. The instructions emphasized that the children's answers were confidential (no one they knew would be allowed to see them). In relation to confidentiality, children were instructed to work on their own and not to look at the responses of others. In addition, the instructions stressed that there were no true or wrong answers because we were interested in

the children's opinions. Children were trained in the scale usage by means of four example items and a large example scale fixed to a clipboard. The interviewer demonstrated how to cross out the answers on a large example scale fixed on the board and explained how the response alternatives of the scales correspond to the children's possible feelings about the item. At the end of the session the children received a chocolate bar as a reward for their assistance.

3.3 Measures

3.3.1 Children's Agency and Means-ends Beliefs, Goal Difficulty, Action Strategies, and Goal Importance

Children's agency and means-ends beliefs, perceptions of goal difficulty, and action strategies were assessed with the Multi-dimensional Control, Agency, and Means-ends Beliefs Inventory for Friendship (Multi-CAM Questionnaire for Friendships, Little & Wanner, 1997). Each of these indicators of self-regulatory processes are defined for the subgoals of "making a friend", "keeping a friend", and "getting a friend to play" which are assumed to represent rather representative facets of the middle level goal of "having good friendships" (see Table D1 in Appendix D, for an overview of the items). This measurement approach is based on assumptions from a domain sampling point of view (Little, Lindenberger, & Nesselroade, 1998). Agency and means-ends beliefs can be differently operationalized with regard to the constituents of action (i.e., goal, means, and agent), time (i.e., present, past, future), and specific wordings (e.g., smart or bright for ability). In the Multi-CAM instrument both types of beliefs are formulated in present tense. Such formulations assess time-neutral mental processes and cognitive constructions (cf., Skinner, 1995)[42].

Agency beliefs[43] are worded in first person and, thus, refer to the individual or the self as actor. Means-ends beliefs are worded in third person (e.g., other children of your age) and,

[42] With reference to control perceived in the present, I will use the time-neutral terms estimates, judgments, representations, and evaluations as well as beliefs, convictions, understanding, and sense of control interachangeable. Note, the previous set of terms reflects mental processes while the latter set of terms reflects cognitive constructions.

[43] Agency beliefs are also termed capacity beliefs (e.g., Skinner et al., 1990; Skinner et al., 1998). Generally, the individual person represents only one unit of analysis (i.e., agent) of the possible ways to conceptualize actions. Dyads, groups, societies, etc. represent other possible agents (e.g., Austin & Vancouver, 1996; Bandura, 1995, 1997; Freund & Baltes, 2000) defining the system level of analyses.

consequently, represent the individuals perceptions of action-outcome relationships as they pertain to a specific reference group (e.g., Little, 1998)[44].

Means-ends beliefs are operationalized for success outcomes[45] (i.e., the agent attained the goal or end *because* the means was available). In a similar vein, agency beliefs are operationalized for success outcomes (i.e., when it comes to attain the goal the agent has the means available). Moreover, agency beliefs are operationalized for success outcomes in which tough odds are acknowledged[46] (i.e., when it comes to attain the goal, even if it is difficult, the agent has the means available). Note, in the Multi-CAM only agency and means-ends beliefs defined for success outcomes are employed. The operationalization of means-ends beliefs for failure outcomes affords the logical complexity of negating negative outcomes (i.e., the agent failed to attain the goal or end *because* the means was not available). Eliminating beliefs about failure outcomes may be justified because both agency and means-ends beliefs defined for success outcomes have been found to be moderately to highly correlated with their counterparts defined for failure outcomes and, consequently, their calculated unit-weight composites usually are entered into the analyses (e.g., Little, et al., 1995).

Note, in the present conceptualization beliefs about both the availability and the usefulness of the whole range of action means (i.e., effort, ability, personal attributes, luck, teachers, and parents) are considered. In contrast, in other conceptualizations of control beliefs competence beliefs encompass only personal means such as skills and personal attributes (e.g., Bandura, 1977, 1989; Ford, 1992) while beliefs about the responsiveness of

---

[44] Means-ends beliefs can be worded in first person and, thus, represent attributions of personal outcomes. This operationalization is refered to as Strategy beliefs (see, e.g., Skinner et al., 1990; Skinner et al., 1998). An alternative conceptualization of means-ends beliefs would be to use the wording "children of your age" and, consequently, involving agents in general. However, this more general attribution of outcomes may be confounded with the attribution of personal outcomes. In contrast, the wording "other children of your age" refers to the children's general views of causal relationships and, at the same time, excludes attributions of personal outcomes. Although, the wording "other children" appears to imply social comparison processes, Wood (1996) proposes that when individuals make such comparative ratings, they typically do not even think about the supposed objects of comparison, much less consider them in relation to the self. As a consequence, individuals may ignore the comparative features of such ratings. Finally, positively worded means-ends beliefs are confounded with agency beliefs for the specified reference group. The attribution of an success outcome implies, in addition, to the usefulness of the specific means, that the agents had the means available.

[45] In addition, the outcomes or goals could be operationalized according to the distinction of approach and avoidance goals (e.g., Carver & Scheier, 1998, 2000) or the distinction of prevention and promotion focus (e.g., Higgins & Silberman, 1998).

[46] Schwarzer (e.g., 1994) proposes a similar operationalization for generalized self-efficacy beliefs.

the context encompass only means residing outside of the individual such as powerful others and situation (Ford, 1992).

3.3.1.1 Agency and Means-ends Beliefs, Goal Difficulty, and Goal Importance

Operationally, (a) making, (b) getting a friend to play, and (c) keeping a friend are defined as subgoals of the middle-level goal of having high-quality friendships (see Section 2.2.1.5). Each of the three subgoals is crossed with the operationalizations of each possible goal-relevant means (i.e., effort, ability, personal attributes, luck, teachers and parents as powerful others). Consequently, each belief type (i.e., agency and means-ends beliefs) that is defined for the goal-relevant means is assessed three times. Similarly, both difficulty and importance are defined for these three subgoal and, thus, each of these goal properties is assessed three times. In addition, each dimension of agency beliefs are defined for attainment of the three subgoals under difficult circumstances. Thus, the agency beliefs comprise six item-level responses for each construct. These six items are aggregated into three parcels of two items each according to the guidelines provided in the coding guide of the instrument (Little & Wanner, 1997, see Table D5 in Appendix D).

For the operationalization of difficult situations a hypothetical situations methodology is used to study children's responses to attain the goal of having good friendships (see Table D2 in Appendix D). This widely used methodology (e.g., Dodge, 1980; Erdley & Asher, 1996; Rabiner & Gordon, 1992; Renshaw & Asher, 1983; Slaby & Guerra, 1988) was chosen because it enables researchers to make controlled comparisons between children, inasmuch as each child encounters the same social situations. A concern regarding hypothetical situations methodology has been with whether children's self-reported behavior corresponds to what they actually do. Research findings suggest that children's responses to hypothetical situations do correspond with their behavior (Chung & Asher, 1996; Dodge & Frame, 1982). In addition to the operationalization of agency beliefs for attainment of the three subgoals in difficult circumstances this methodology was used for assessing action strategies.

3.3.1.2 Action strategies in Difficult Friendship Situations

Moreover, the action strategies Direct Action, Seeking Help, Problem Avoidance, and Doing Nothing were assessed with the Behavioral Inventory of Strategic Coping (BISC, Lopez & Little, 1994). Notably, in this instrument the strategies are defined for acting on the

same three subgoals in difficult friendship situations as they are defined for the agency beliefs.

Table D3 in Appendix D gives an overview of the operational characteristics of specific components of the instrument for the English version (Table D4 in Appendix D gives the German operational definitions).

3.3.1.3 Response Format

The response format involves organizing the type of response around (a) the type of belief (agency or means-ends), goal difficulty, goal importance and action strategy, and the (b) the operationalized goal component (e.g. making, keeping, or engaging friends). The items are presented such that the items of a given category and goal are presented on a single page and organized randomly within a block for each outcome; for example, the means-ends belief items for making friends are together, then goal difficulty, goal importance and the agency belief items for making friends are presented together, immediately following the means-ends items. Then the agency beliefs for making friends under difficult circumstances follow. Then, the strategies to act on the subgoals under rather low control conditions follow. Then, the means-ends items for getting a friend to play are grouped together, followed again by the corresponding goal difficulty, goal importance, and agency belief items, and so on.

The instrument has satisfying measurement qualities in terms of internal consistency and both internal and external validity (Lopez & Little, 1996; Wanner, 1995) as well as re-test reliability (Lopez & Little, 1996).

3.3.2 Quality Characteristics of Children's Friendships

Children's friendship characteristics were assessed with the Friendship Inventory which is designed for group assessments (Little, Krappmann, Brendgen, & Wanner, 1997). The Friendship Inventory is based on the Interview about Friends (Krappmann et al., 1991) which represents a comprehensive standardized instrument designed to assess both quantity and quality of the children's entire friendship networks inside and outside of school. The measured quantitative and qualitative features of children's friendships are derived on basis of previous research on children's friendship characteristics (Krappmann & Oswald, 1983; Oswald & Krappmann, 1984; Oswald, Krappmann, & Fricke, 1988). Of the various qualitative friendship features (e.g., fun, and support) the present study focuses on "Intimacy"

and "Conflict". More specifically, in the Friendship Inventory the higher-order construct Intimacy is conceptualized to represent the two highly related factes "intimate friendship actions" and "mutual liking". In the present study, only the facet “intimate friendship actions” is used (for reasons of simplicity, in the following this factet will be refered to as “Intimacy”). As a consequence, the measures of Intimacy and Sociometric Status are unconfounded with regard to the aspect of Liking. Table D6 in Appendix D provides for an overview of the items.

In contrast to its predecessor, the Friendship Inventory (Little et al., 1997) targets on the assessment of only the three best friendships. Limiting the assessment to the three best friendships is based on research showing that only few children (e.g., Brendgen et al., 2000) and adolescents (e.g., Berndt & Keefe, 1995) have three or more reciprocated friendships.

The measurement qualities of the instrument in terms of internal consistencies and internal validity are satisfying (see Appendix E Table E1 showing the internal consistencies by grade level, Table E2 showing them for friended children across sociometric status, and E3 showing them for friendless children across sociometric status; see also Table E5 showing the internal consistencies of children's typical views of friendship quality and children's own and their friends' views of friendship quality; see also Table H5 in Appendix H reporting the unstandardized factor loadings of children's typical view of friendship quality across sociomentric groups of friended and friendless children).

3.3.2.1 Assessing Friendship Reciprocity

By interviewing entire classrooms, many of the children that are nominated by the interviewed child are interviewed as well. Thus, reciprocity of friendship nomination as well as perceptual correspondence with respect to the several characteristics of the friendship can be assessed. Because friendships in the classroom are an especially important part of children’s social experiences during middle childhood (Krappmann et al., 1993), these data reflect crucial friendship relationships (see Section 2.4.2.5).

Generally, the assessment of nominated friends who don’t attend the same school and, who may even live in other towns, is for logistical reasons difficult. Thereby, the issue of confidentiality puts further restrictions on the logistics. Such limitations in generalizability of the results are the rule in studies investigating friendship reciprocity (e.g., Brendgen, et al., 2000). However, because most elementary children of a specific neighborhood attend the

same school there may be a large overlap between friends in school and friends in the neighborhood (see Section 2.4.2.5). Generally, children nominated more friends who attended the same school than friends who did not[47]. (M = 6.19 , Std = 2.82, min. = 1, max. = 14) than friends not attending the same school ((M = 2,32, Std = 2.35, min. = 0, max. = 14); t (641) = 26.26, p < .01. Note, the number of children's friends who do not attend the same school may not exclusively be comprised of children living in the neighborhood. This number may, for example, also include friends the children meet only on holidays because they live somewhere else.

3.3.2.2 Response Format

The Friendship Inventory firstly asks children to nominate the three best friends they have in school. Thereby, the children are supposed to rank the friends according to the closeness of their friendship (i.e., the very best friend is named first, and the least best friend is named as third). If the children have less than three friends in school they are allowed to nominate less friends. The names of the three best friends are written at the top of the questionnaire to ensure that the children can see to whom the following items refer to. That is, a item is printed at the left side of a page and the response scales refering to the three friends are sequentially ordered at the right side of a page. Thereby, each response scale is placed below the name of the friend it refers to.

[47] The children were allowed to nominate friends who attend the same school and were not restricted to nominate only children of the own classroom.

3.3.3 Response Scales

The items of each of the administered instruments (i.e., Multi-CAM for Friendships and Friendship Inventory) are presented with 4-point Likert response scales[48] (i.e., almost never, sometimes, often, almost always; German: sehr selten, ab und zu, häufig, sehr oft).

3.3.4 Children's Sociometric Status

Children's sociometric status was assessed by means of the nomination procedure developed by Coie et al. (1982). Each child is asked to write down the names of three children in the classroom he or she likes most and likes least[49]. The classroom roster is put in front of the child to ensure that all children are equally remembered and, thus, have the same theoretical probability to be nominated. In order to estimate a child's sociometric status, the number of positive and negative votes in the classroom received by each child are separately aggregated. By further treatment of the data, five different sociometric groups can be clearly distinguished to which most children can be assigned: popular, average, controversial, neglected, and rejected children. A confounding of theoretically distinct groups -- for instance among neglected and rejected children -- is avoided. Aside from a rating scale approach developed by Newcomb and Bukowski (1983), this procedure is most frequently used classification system (Rubin et al. 1998). It not only meets the conceptual criteria of popularity measurement as proposed by Bukowski and Hoza (1989) but the nomination procedure also reveals strong relationships with other measures of popularity, such as rating scales (Bukowski & Hoza, 1989), which suggests good validity of measurement. Furthermore, the sociometric status measure yielded by this procedure shows acceptable stability across

[48] The number of scale points (i.e., answer categories) in a measure has a direct effect on the extent to which the measure can reflect the actual variation that exists in a population. Given a reasonably "normal" distribution and an optimal match of how the scale categories match that distribution two-point scales reflect roughly 67% of the underlying variation, three-point scales, roughly 80%, four-point scales, roughly 90%, five-point scales, roughly 95%, seven-point scales, roughly 97%, and 11-point scales, roughly 99% (e.g., Cochran, 1968; Cox, 1980). These figures show that the gain in precision achieved by lenghtening the scale peaks at four-point scales. The percentage of covered variation increases more rapidly at the shorter end of the continuum representing the scale lenghts than at the longer end. Competing criteria such as cognitive processing-demands, however, also should be considered. Generally, children may have less difficulties to respond to four answer categories than to higher numbers of answer categories.

[49] It has been suggested that the administration of negative nomination measures may implicitly sanction negative judgments about peers, and children to view disliked peers even more negatively, and serve to increase the salience of their marginal status within the group. However, sociometric testing has not been found to have adverse effects on subsequent peer relations (e.g., Bell-Dolan, Foster, & Sikora, 1989; Hayvren & Hymel, 1984) or to result in feeling lonely (Bell-Dolan et al., 1989) or the children having hurt feelings (Iverson, Barton, & Iverson, 1997). In addition, potential risks involved in the administration of the negative nomination measure were minimized because the children did not see (or hear) each others nominations.

both time (Coie & Dodge, 1983) and new situations (Coie & Kupersmidt, 1983), indicating good reliability of measurement.

3.3.5 Control Variables

As outlined in Section 2.4.3.2 Aggression, Raven Intelligence, School Achievement, and Social Desirability were assessed as control variables. Aggression was assessed by using a single indicator adapted after the Revised Class Play Method of Peer Assessment (Masten, Morison, & Pellegrini, 1985). The children were asked to nominate three peers for whom the attribute is most true (for the wordings of the English and German instructions, see Table D8 in Appendix D). In order to account for differences in the sizes of the classrooms the sum scores of the nominations were z-standardized within classrooms. Fluid intelligence was measured by a speeded short version of the RAVEN's Standard progressive Matrices (SPM, Raven, 1971) developed by Little, Wanner, & Mauch (1997). The year-end grades of math and verbal achievement served as indicators of School Achievement. Children's tendencies to respond in normatively desired ways were measured by an adapted short version of the Social Desirability subscale of the German Junior Eysenck Personality Questionnaire developed by Rost & Hartmann (1993) which is based on the Junior Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975). Table D9 in Appendix D shows the English and German items.

3.4 Treatment of the Data

3.4.1 Identification of Sociometric Groups

Based on the method suggested by Coie et al. (1982), the total number of positive and negative nominations received were calculated for each child and z-standardized within classrooms. Then, the sum of a child's received positive plus negative nominations was computed to yield the child's social impact-score which indicates the child's social visibility. Additionally, the number of each child's negative nominations minus the number of positive nominations yielded the child's social-preference score indicating the child's general likability. The social impact and social preference scores were z-standardized and were then used to identify children for the five distinct social status groups as described by Coie et al. (1982).

The popular group consisted of all those children who (a) received a social preference score of greater than 1.0, (b) a standardized positive score of greater than 0, and (c) a

standardized negative score of less than 0. Thus, the popular children are exceptionally well liked by their peers.

The rejected group consisted of those children who received (a) a social preference score of less than -1.0, (b) a standardized negative scores of greater than 0, and (c) a standardized positive score of less than 0. Thus, the rejected children are exceptionally disliked by their peers.

The neglected group consisted of those children who received (a) a social impact score of less than -1.0 and (b) an absolute positive score of 0. The neglected children thus had no one identifying them as among the three classmates they liked most. They differed from the rejected children in that the rejected children reiceived many negative votes, whereas the neglected children did not. Thus, the neglected children are not explicitly disliked but rahter overlooked by their peers.

The controversial group consisted of those children who received (a) a social impact score of greater than 1.0 and (b) positive and negative standardized scores that were each greater than 0. Thus, members of this group were all above the mean for both positive and negative nominations. These children stand out, because they have many classmates who like them and many others who dislike them.

The average group consisted of those children who received a standardized social preference score that was greater than -.5 and less .5. Thus, these children are not especially liked or disliked, but they differ from the neglected children in that they are not overlooked by their classmates.

By following these criteria, 476 (74.1%) children were identified of the original sample of 642 as fitting into one of the five social status types. Specifically, there were 99 (15.4%) children in the popular group, 84 (13.1%) in the rejected group, 15 (2.1%) in the neglected group, 22 (3.4%) in the controversial group, and 256 (39.3%) in the average group. The remaining 166 (25.9%) children have not met Coie et al.'s stringent classification criteria and could not be positively classified into one of the sociometric groups. This situation is typical for the Coie et al. classification method which is aimed at identifying 'pure' average and extreme sociometric groups. The relative sizes of the various sociometric groups and the remaining unclassified children in the present sample are in accordance with the relative sizes

usually reported when employing this method (e.g., Brendgen et al., 2000; Coie et al., 1982; Dodge, 1983; Coie & Dodge, 1988; Patterson et al., 1990; Kupersmidt & Coie, 1990).

3.4.1.1 Does the Participation Rate of the Children Affect the Reliability of the Sociometric Classifications?

Crick and Ladd (1989) found that nominator attrition and reliability of this sociometric classification are inversely related. The authors found that beginning with a participation rate of lower than 80% classification errors were most apparent among the controversial and the unclassified groups. Of the remaining four groups, the average and neglected groups were least resistant, and the rejected group most resistant to classification errors due to nominator attrition. Hence in the present study, nominatior attrition can be assumed to have the largest impact on the average group and the lowest impact on both the popular and rejected groups. Nominator attritition reached in 13 (33.3%) of the 39 assessed classrooms the critical benchmark of 80%. Specifically, in these classrooms only 42.8% to 78.4% (Md = 70.4 %) of the children participated. Removing the 204 children of these 13 classrooms from the analyses would have reduced the sample sizes of the investigated groups to a considerable extent. Importantly, conducting the analyses with the reduced sample sizes, generally, did not affect the results. Thus, I report the results based on the unselected sample. In addition, this procedure appears justified because researchers often combine the average group with the unclassified group (see, e.g., Coie & Dodge, 1983; French & Waas; Patterson et al., 1990) which indicates that unreliability of identifying average children may be of negligible concern.

3.4.2 Identification of Groups of Friended and Friendless Children across the Sociometric Status Groups (i.e., Popular, Average, and Rejected Children)

In a first step, those children had to be identified who had at least one of their friendship nominations in class reciprocated as well as those children whose friendship nominations were not reciprocated. Of the total 642 children 537 (83.6%) children had, at least, one mutual friendship. Specifically, 120 (18.7%) children had three mutual friendships, 202 (31.5%) children had two mutual friendships, 215 (33.5%) children had one mutual friendship, and 105 (16.4%) children had no mutual friendship. A conducted mean-level comparison showed that the number of mutual friendships was independent from children's grade level (see

Appendix G for analyses investigating the developmental differences of the relationships of the target constructs with the number of mutual relationships). Across each grade level the number of mutual friendships ranged between 0 and 3. Table E1 in Appendix E provides the descriptive statistics of the number of mutual friendships separately for each grade level.

In the next step, friended and friendless children were identified across various sociometric status groups. Sociometric status assessed according to the procedures proposed by Coie et al. (1982) (see Section 2.4.1). Specifically, popular, average, and rejected children were identified whose friendship nominations were either reciprocated or not reciprocated by the nominated friends. As shown in Table 4, 97 popular children (8.1 - 13.2 years, M age = 10.7 years, SD = 1.3), 219 average children (8.1 – 12.9 years, M age = 10.2 years, SD = 1.2), and 46 rejected children (8.0 – 12.3 years, M age = 10.3 years, SD = 1.3), at least, had one friend who reciprocated the friendship nomination. Only two (2.0%) children[50] of the popular group nominated friends who did not reciprocate the friendship nomination. 37 (14.5%) children of the average group (8.3 - 13.3 years, M age = 10.2 years, SD = 1.5) and 38 (45.2%) children of the rejected group (8.3 – 12.6 years, M age = 10.5 years, SD = 1.3) were not nominated by the nominated children to be one of the three best friends at school. Conducted mean-level comparisons revealed that, on average, the number of mutual friendships differed across the three sociometric groups (see Section 4.2.2 for analyses investigating the differences of the relationships of the target constructs with the number of mutual relationships across the sociometric groups).

Generally, the finding that only about half of the rejected children had, at least, one mutual friendships is in line with findings from previous studies (e.g., Brendgen et al., 2000; George & Hartmann, 1996; Ladd, 1983; Parker & Asher, 1993; Rose & Asher, 1999; Uhlendorff & Krappmann, 1999). This finding is in line with the assumption that peer status and friendships are related goals (see Section 2.4.3)

[50] A nine year old popular girl (grade 3) nominated a girl and two boys who participated in the study. An eleven year old popular boy (grade 5) nominated three boys of whom one did not participate in the study.

Table 4
Age, Gender, Number of Nominated Friends, Number of Nominated Friends who Participated of Friended and Friendless Children across the Groups of Popular, Average, and Rejected Children

	Target Groups				
	Friended Children			Friendless Children	
Characteristic	popular	average	rejected	average	rejected
Group size (total N = 437), N (percent)	97 (22)	219 (50)	46 (11)	37 (9)	36 (9)
Age (years)					
Mean (SD)	10.7 (1.3)	10.2 (1.2)	10.3 (1.3)	10.2 (1.5)	10.5 (1.3)
Range	8.1 - 13.2	8.1 - 12.9	8.0 - 12.03	8.3 - 13.3	8.3 - 12.6
Gender, N (percent)					
male	39 (40)	117 (53)	27 (59)	17 (46)	23 (61)
female	58 (60)	102 (47)	19 (41)	20 (54)	15 (39)
Nominated Friends, Mean (SD)	3.0 (0.1)	2.9 (0.3)	2.7 (0.6)	2.8 (0.5)	2.8 (0.6)
Nominated Friends who participated, Mean (SD)	2.7 (0.5)	2.6 (0.6)	2.3 (0.7)	2.4 (0.7)	2.4 (0.8)
N (percent) of children of whom only a single friend participated	3 (3)	15 (7)	5 (11)	4 (11)	6 (16)
Mutual Friendships					
total, Mean (SD)	2.3 (0.7)	1.7 (0.7)	1.2 (0.4)		
Range	1 - 3	1 - 3	1 - 2		

Note. In the popular group only two (2%) children's friendship nominations were not reciprocated by the friends. These two children were excluded from the analyses.

In line with findings by Rizzo (1988) and Brendgen et al. (2000), rejected children established friendships not only with other rejected children, but with children from each sociometric group. In contrast to the findings reported by Brendgen et al. who report that the mean proportions of average, popular, and rejected children's friendships with children of the specific sociometric groups did not differ, in the present study a conducted one-factorial multivariate analysis of variance (MANOVA) with sociometric groups (popular, average, and rejected children) as independent variable and mean proportions of friendships with children of sociometric groups (i.e., popular, average, controversial, neglected, and rejected) as

multiple dependent variables showed that there was a multivariate relationship; multivariate $F_{(10, 710)} = 6.40$, $p < .01$.

Specifically, conducted one-way ANOVAs revealed that these mean proportions differed across sociometric groups for friendships with rejected children ($F_{(2, 359)} = 16.92$, $\eta^2 = .09$, $p < .01$), average children ($F_{(2, 359)} = 5.17$, $\eta^2 = .03$, $p < .01$), and popular children ($F_{(2, 359)} = 13.99$, $\eta^2 = .07$, $p < .01$) but not for friendships with neglected, ($F_{(2, 359)} = 0.90$, $\eta^2 = .01$, $p = .41$; $M = 0.60\%$, $SD = 5.25$, $M = 0.00\%$, $SD = 0.00$, $M = 1.08\%$, $SD = 7.34$, for average, popular, and rejected children, respectively), controversial, ($F_{(2, 359)} = 0.15$, $\eta^2 = .00$, $p = .86$; $M = 4.11\%$, $SD = 15.17$, $M = 5.15\%$, $SD = 16.55$, $M = 4.34\%$, $SD = 14.24$, for average, popular, and rejected children, respectively) and unclassified children, ($F_{(2, 359)} = 0.84$, $\eta^2 = .00$, $p = .43$; $M = 25.87\%$, $SD = 37.27$, $M = 28.01\%$, $SD = 34.32$, $M = 33.69\%$, $SD = 44.78$, for average, popular, and rejected children, respectively). Conducted aposteriori Scheffe tests ($p < .05$) indicated that rejected children, on average, had a lower percentage of popular friends ($M = 4.34\%$, $SD = 17.72$) than both average children ($M = 18.64\%$, $SD = 33.50$) and popular children ($M = 34.36\%$, $SD = 38.55$). The latter two groups also differed significantly from each other. However, rejected children, on average, had a similar percentage of average friends ($M = 35.87\%$, $SD = 46.74$) as both average children ($M = 46.42\%$, $SD = 41.61$) and popular children ($M = 31.27\%$, $SD = 33.61$). In contrast, average children's percentage of friendships with other average children was significantly higher than popular children's percentage of friendships with average children. Rejected children, on average, had a higher percentage of friendships with other rejected children ($M = 20.65\%$, $SD = 40.42$) than average children ($M = 4.33\%$, $SD = 16.3$) and popular children ($M = 1.20\%$, $SD = 6.90$). The latter two groups did not differ significantly from each other.

The majority of the 1182 nominated friendships 1113 (94.2%) were with children of the same sex and only 69 (5.8%) of the nominated friendships were with children of the opposite sex. With regard to the friendship nominations that were reciprocated by the nominated friends none of them involved a relationship with a friend of the opposite sex.

3.4.2.1 Does Participation Rate of the Children's Friends Affect the Reliability of the Classification of Friended and Friendless Children?

The present section assesses the reliability of the employed classification of friendship reciprocity. Specifically, the reliability of the employed classification of children as being friendless would be affected if the friends whom they have nominated haven't participated in the study. If none of the friends has participated, then, no friend can reciprocate the friendship nomination. As a consequence, children were removed from the analyses if there were no valid responses of a single nominated friend available (see Point 3.1). Similarly, if the friends of children who haven't received a single friendship nomination, generally, participated less frequently in the study compared to the friends of children who have reciprocated friendships, then, this would indicate that the employed classification may not be reliable. Table 4 depicts the mean levels of number of nominated friends and participation rate of the nominated friends across the groups.

A conducted 3-x-3-Chi-Square test revealed that the friends of children who haven't received a single friendship nomination did not participate less frequently in the study as the friends of average children who have reciprocated friendships; $\mathbf{x^2_{(4)}} = 7.72$, $p = .10$, two-tailed test. Thus, the classification of children as being friendless may not be due to a lower probability of receiving a nomination because the nominated friends have not participated in the study. This result provides support for the assumption that the classification of friendless and friended children may be reliable.

3.4.2.2 Do Sociometric Groups of Friended and Friendless Children Differ in Age and Gender?

In order to test whether friended and friendless children differ in age and whether there is an interaction with sociometric status, I performed a performed one-way ANOVA on the groups of popular, friended and friendless average, friended and friendless rejected children. The results showed that, on average, the children's age did not differ across the five target groups (i.e., friended-popular, friended-average, friended-rejected children, as well as friendless-average and friendless-rejected children); $F_{(4,\ 432)} = 0.38$, $\eta^2 = .00$, $p = .82$. A conducted 5-x-2-Chi-Square test revealed that the distribution of gender was not significantly different across the five target groups, $\mathbf{x^2_{(4)}} = 7.88$, $p = .10$, two-tailed test. This result

supports the assumption that neither friendship participation nor sociometric status or their interaction were related to gender.

In sum, there was no evidence that friended and friendless children differed in mean age and gender across sociometric groups.

3.4.2.3 Do Sociometric Groups of Friended and Friendless Children Differ in Aggression?

As described in the theory part (see Section 2.4.3.2), I expected that the groups would differ in the peer-nominated aggression scores. Specifically, I expected that friended-rejected children evince the highest levels of aggression compared to the remaining groups. With regard to friendless-rejected children I expected that the finding would replicate that they were less aggressive than both aggressive-rejected and popular children (French, 1988).

In order to test whether friended and friendless children differ in their levels of aggressiveness across sociometric groups, I performed a performed one-way ANOVA on the groups of popular, friended and friendless average, friended and friendless rejected children. The results showed that the five groups differed significantly[51] in peer nominations of aggression ($F_{(4, 432)} = 7.92$, $\eta^2 = .07$, $p < .01$). As follow-up tests, I conducted a series of t-tests for independent samples. The results showed that friended-rejected children had significantly higher levels of aggression than friended-average children ($M = -0.61$, $SD = 1.19$ and $M = -0.07$, $SD = 0.85$, for friended-rejected children and friended-average, respectively); $t_{(551)}$ [52] $= -3.69$, $p < .01$, one-tailed test. Thus, the expectation that the groups of friended-rejected children has a large overlap with the group of aggressive-rejected children found some support. Contrary to the expectations, friendless-rejected children were found to be significantly more aggressive than popular children ($M = 0.36$, $SD = 1.22$ and $M = -0.11$, $SD = 0.84$, for friendless-rejected and popular children, respectively); $t_{(51.5)}$[53] $= -2.18$, $p < .05$. Moreover, friendless-rejected children did not significantly differ from their aggressive counterparts ($M = 0.36$, $SD = 1.22$ and $M = -0.61$, $SD = 1.19$, for friendless-rejected and

[51] Despite the nonnormal distribution of aggression in the subsamples of popular and average children (see Table O1 in Appendix O) the conducted parametric tests (i.e., one-way ANOVA and aposteriori Scheffe tests) can be assumed to yield reliable test statistics because each of the groups is comprised of more than 30 subjects. Moreover, a conducted Kruskal-Wallis test and aposteriori conducted Wilcoxon U-tests yielded similar results as the reported parametric tests.

[52] The variances of aggression differed significantly across the groups of friended-rejected children and friended-average children; $F\ (45,22) = 1.95$, $p < .01$. Thus, the t-value refers to the respective test for unequal variances.

[53] The variances of aggression differed significantly across the groups of friendless-rejected children and popular children; $F\ (37,96) = 2.09$, $p < .01$. Thus, the t-value refers to the respective test for unequal variances.

friended-rejected children, respectively); $t_{(82)} = 0.95$, $p = .34$. Thus, the expectation that the groups of friendless-rejected children has a large overlap with the group of nonaggressive-rejected children was not supported.

Finally, conducted aposteriori Scheffe tests ($p < .05$) did not indicate that the friendless-average group differed from any of the other groups in their levels of aggressiveness. This finding suggests that these tests lacked the statistical power to detect differences. For example, it was to be expected that this group either significantly differed from the popular or the friended-rejected group .

3.4.3 General Remarks to the Analyses Employing Sociometric Status and Friendship Status (Friendless vs. Friended Children) as Potential Moderator Variables

As outlined in the theoretical part of this dissertation (see Section 2.5), in the present study sociometric status is hypothesized to moderate the relationships among perceived control, action strategies, and friend-rated friendship quality as well as the number of mutual friendships. In addition, the present dissertation investigates whether the relationships among perceived control, action strategies, and self-rated friendship quality are invariant across sociometric groups of friended and friendless children; that is, a lack of moderating effects of both sociometric status and friendship status were hypothesized.

A moderator is a qualitative or quantitative variable that affects the direction and/or strength between a predictor (i.e., independent) variable and a criterion (i.e., dependent) variable (cf., Baron & Kenny, 1986). In analysis of variance terms, a basic moderator effect can be represented as an interaction between an independent variable and a factor specifying the conditions for its operation. Thus, moderators specify when certain effects will occur.

Children's sociometric status was used as a measure of objective control conditions which are assumed to moderate children's action control in friendship relationships. Such an approach is quasi-experimental (Cook & Campbell, 1979) because sociometric status can't be randomly assigned to participants. In addition, the present study can be characterized as field research (Bickman & Henchey, 1972) because participants were assessed in real-life settings (Stroebe, Hewstone, Codol, & Stephenson, 1992). Thus, the present study can be characterized as a quasi-experimental field-research (cf., Bortz, 1984) which has the advantage to be high on external validity. However, such an approach suffers from threats

(e.g., selectivity) to internal validity. Consequently, controlling for variables which may affect the results of the study was important. However, it is impossible to control for all potential variables that may cause selection effects. Hence, in the present study constructs were controlled that in previous research have been found to influence peer relationships (see Section 2.4.3.2).

Specifically, gender, age, Raven intelligence, school achievement, peer nominations of aggressive behaviors, and behaving in normatively and socially desirable ways have been demonstrated to be related to peer relationships. Some of these variables (e.g., intelligence, aggressive behaviors) can be considered as mediating the relationships of agency and means-ends beliefs, Goal Difficulty, Goal Importance, and Action Strategies with the friends' views of friendship quality. However, some of these variables can be also thought as third variables influencing agency and means-ends beliefs, Goal Difficulty, Goal Importance, and Action Strategies as well as the friends' views of friendship quality (e.g., school achievement). Because age and gender were assumed to be most influential their effects were controlled in each of the conducted analyses.

3.4.4 Preparation of the Raw Data for Analyses

3.4.4.1 Parceling

For purposes of factor analyses, items indicating the same underlying construct were collapsed (i.e., calculating the unit-weight composite of a set of items) into parcels or aggregate indicators (i.e., subscales). In factor analyses, analyzing parcel scores instead of items is advantageous in several respects: (a) aggregate indicators are more reliable and have greater generality (i.e., validity), (b) response biases and other characteristics that are idiosyncratic to individual items are likely to have less influence, (c) the ratios of manifest variables to inferred factors and to estimated parameters are increased, and (d) the distributional characteristics of parcel scores are less likely to cause problems than distributions of items (e.g., Kishton & Widaman, 1994; Marsh, Antill, & Cunningham, 1989). Several procedures of collapsing items into subscales have been suggested in the literature (e.g., Kishton & Widaman, 1994). The procedure employed for all constructs was an assignment of items into three aggregate indicators per construct according to the guide lines provided by the respective coding guides of the instruments (see, Little & Wanner, 1997 for

the Multi-CAM instrument, and Little et al., 1997 for the Friendship Inventory). This parceling procedure follows the rationale of internal consistency and unidimensionality (Kishton & Widaman, 1994).

In order to just identify a measurement model, the constructs should not be represented by more than three indicators, since only then the number of equations in the measurement model equals the number of the parameters to be estimated (i.e., just identified model; Bollen, 1989), which minimizes the probability of random dual factor loadings (Sullivan & Feldman, 1979).

A summary of the combination of items for each construct indicator is provided in Table D5 in Appendix D.

Screening the data to eliminate missing data, outliers, and skewed distributions was done because each can seriously bias parameter estimates in structural equation modeling (Gallini & Casteel, 1987; Boomsma, 1987).

3.4.4.2 Treatment of the Missing Data

Overall, on item level only 3.44% of the 72 items assessing agency beliefs (36 items), means-ends beliefs (18 items), action strategies (12 items), goal difficulty (3 items), and goal importance (3 items) were missing. When only one or two items of a set of six items assessing each dimension of agency beliefs (i.e., effort, ability, personal attributes, luck, parents, and teachers) or when only one item of a set of three items measuring each dimension of means-ends beliefs (i.e., effort, ability, personal attributes, luck, parents, and teachers), action strategies (i.e., direct action, seeking help, doing nothing, and avoidance) as well as goal difficulty, and goal importance had missing values, missing values were replaced by the mean of the remaining items. This procedure was applied for 1.70% of the item responses.

The same procedures were applied for the items assessing friendship quality of each of the three nominated school friendships. When only one or two responses of the six items assessing intimacy of a specific friendship (i.e., a specific child's first, second, or third friendship) was missing the values of the missing responses were replaced by the mean of the non-missing responses of the specific friendship. This procedure was applied for 0.30% of the responses of the first friendship regarding friendship intimacy, 0.33% of the responses of the second friendship, and 0.36% of the responses of the third friendship. When only one

response of the three items assessing conflict of a specific friendship was missing, the values of the missing responses were replaced by the mean of the non-missing responses of this friendship. This procedure was applied for 0.29% of the responses of the first friendship, 0.33% of the responses of the second friendship, and 0.34% of the responses of the third friendship. Then, these items were collapsed across the the three nominated friendships (i.e., calculating the mean of the items assessing the three nominated friendships) into aggregate items representing the children's typical perceptions of the quality of the nominated friendships.

When only two responses of the six aggregate items assessing intimacy of the nominated friendships was missing the values of the missing responses were replaced by the mean of the non-missing responses of this friendship characteristic. This procedure was applied for 0.05% of these aggregate scores. When only one response of the three aggregate items assessing conflict was missing the values of the missing responses were replaced by the mean of the non-missing responses of this friendship feature. This procedure was applied for 0.22% of these aggregate scores.

After replacing the missing values on item level, items measuring agency beliefs were parceled into three indicators or parcels[54]. Then, the missing values of the items and parcels were estimated using saturated regression equations to estimate any missing value from non-missing values (Lösel & Wüstendorfer, 1974; see also Little, Das, Carlson, & Yachimowicz, 1993). This procedue was applied for 1.25% of the data points on parcel level (i.e., 642 subjects * 63 parcels).

[54] The aggregate items of friendship intimacy were parceled after completing both the missing and the outlier treatments. Then, the three sets of items assessing self-rated friendship intimacy of the three nominated friendships, self-rated friendship intimacy of the reciprocated friendships, and the friend-rated intimacy were aggregated into parcels comprised of the same friendship characteristics (see Table D6 in Appendix D).

3.4.4.3 Outlier Treatment

Assessments of the distributional characteristics (i.e., skewness, kurtosis, and outliers) were conducted on the parcel level. Outliers were identified for each individual by using regression procedures. Specifically, each parcel was predicted by saturated regression equations. Any score falling outside the 1% isodensity contour of the regression distribution was deemed an outlier. Those cases identified as outlier were pulled back toward the mean (i.e., reweighted) and were assigned a score that reflected the 5th percentile of the isodensity contour of the regression distribution. Overall, 1.28% of the data points were identified and adjusted as outliers by means these multivariate procedures.

Descriptive statistics (i.e., means, standard deviations, minimum, maximum, skewness, and kurtosis) for the indicators of the constructs used in the present study, are presented in Appendix E Table E1 by grade level, Table E2 for friended children by sociometric status, and E3 showing them for friendless children by sociometric status; see also Table E4, Table E5, Table E6, and Table E7 showing the internal consistencies of agency and means-ends beliefs, measures of friendship quality, Goal Difficulty and Importance, and action strategies, respectively. As the tables show, in general, the distributional characteristics of the indicators (i.e, means, standard deviations, skewness, and kurtosis) satisfy the assumption of normality after data treatment.

3.4.4.4 Aggregating Items Assessing Friendship Quality into Indicators of the Children's Typical Views of Friendship and their Own Views of Reciprocated Friendships and their Friends' Views of the Friendship

I used the items assessing children's views of their specific friendships to aggregate the constructs representing their own and their friends' views of mutual friendships. If these items had still missing values after the first step of missing replacement (i.e., replacing a missing value by the mean of the items assessing the same constructs for a specific friendship), then, missing values were replaced with the mean of the cleaned aggregate items assessing the specific dimension of friendship quality (i.e., intimacy and conflict). This procedure was applied for 0.38% of the responses of the first friendship regarding intimacy, 0.53% of the responses of the second friendship, and 0.45% of the responses of the third friendship. With regard to conflict, this procedure was applied for 0.15% of the responses of the first

friendship, 0.29% of the responses of the second friendship, and 0.07% of the responses of the third friendship. Then, the items assessing friendships which were reciprocated by the friends in a first step were aggregated into items that represented the own views or mutual friendships. In a second step, these aggregate items were parceled according to the same guidelines as the items assessing the children's typical views of their friendships. Similarly, the item responses assessing the friends' views of friendship with a specific child were identified and aggregated into aggregate items. Then, these aggregate items representing the friends' views were parceled following the same guidelines that were implemented when parceling (a) the items assessing the children's typical views of their friendships and (b) the items assessing the children views of their mutual friendships. Specifically, for each item, the three friends' ratings of the relationship with specific child were averaged. These friends were the same ones that the specific child's self-report ratings were averaged for. Thus, for each child in the sample, the most salient friends' typical (i.e., average) evaluation of friendship intimacy and conflict was obtained. Using this procedure, any existing bias in the average friendship perception (i.e., a systematically more or less positive view of friendship quality as compared to the friends' views) can be assessed for each specific child. As a consequence, at the end there were three parallel sets of parcels assessing (a) the children's typical views of friendship quality, (b) the children views of the mutual friendships, and (c) the friends' views of the mutual friendships. Table D7 in Appendix D shows the combinations of items comprising the specific parcels.

3.4.4.5 Reduction of Statistical Dependencies when Analyzing the Friends' Mutual Views of their Relationships across Sociometric Groups

Dependence between observations is a concern for data analyses because the standard inferential statistics assume that errors in observations are independently sampled (e.g., Kenny & Judd, 1996). The child's and the friends' estimates of friendship quality were averaged (see below Section Treatment of the Data). An advantage of this procedure is that redundancies are avoided, thereby maintaining sufficient sample size for the analytic procedures. Using this procedure, any existing bias in the average friendship perception (i.e., a systematically more or less positive view of friendship quality as compared to the friends' views) can be assessed for each child. Importantly, Brendgen et al. (2000) found that this procedure of aggregating

across more than one friendship does not result in attenuated correlations across the partners' perspectives. Among those children with only one mutual friend, however, there were 18 (4.97%) dyads where the two partners had only received reciprocated friendship nominations from each other, but not from any other child they had nominated. In these cases each child had a self-report and friend-report score referring to the same single friendship. In order to avoid duplicate, dependent information in these cases, one partner of each of these dyads was randomly excluded from the sample. However, the random selection of the partners was restricted to the condition that the friend was dropped if a child belonged to a smaller target group . For example, if the child belonged to the rejected group while the friend belonged to the larger average group, the friend was dropped from the analyses. Specifically, 14 (3.87%) children of the average group and 4 (1.10%) children of the rejected group were excluded from the analyses. Notably, even the random exclusion of one partner from exclusive dyads does not provide complete independence of the data, which could not be achieved without severe loss of subjects. The aggregation procedure across several friends, however, avoids duplicate information. Moreover, Brendgen et al. (2000) found that cases where children with two friends or children with three friends mutually nominate only each other are extremely rare. Moreover, friends who reciprocated a specific child's nominations, often, were part of other sociometric groups than the three target groups which may increase independence of the data. Consequently, the results of the study are probably not overly attenuated by a lack of independence of the data.

As a result the final sample sizes used in the analyses investigating the relationships among agency, and means-ends beliefs, goal difficulty, and action strategies, and the friends' evaluations of friendship quality are reduced. Specifically, these analyses are based on 205 average children (8.1 - 12.9 years, M age = 10.2 years, SD = 1.2), 97 popular children (8.1 - 13.2 years, M age = 10.3 years, SD = 1.3), and 42 rejected children (8.1 - 12.3 years, M age = 10.3 years, SD = 1.3).

3.5 General Statistical Procedures

3.5.1 Structural Equation Modelling as a Statistical Approach

The hypotheses of this dissertation were tested employing either standard structural equation model (SEM) techniques (using LISREL 8; Jöreskog & Sörbom, 1993) or its extension multiple-group mean and covariance structures analyses (MACS, Little, 1997). In MACS analyses mean-level information is analyzed in addition to variance-covariance information (Cole, Maxwell, Arvey, & Salas, 1993; Horn & MacArdle, 1992; Little, 1997; McArdle & McDonald, 1984). A MACS analysis has three main advantages: (a) it tests whether the hypothesized factor structure of the item responses is tenable, thereby representing a powerful hypothesis-testing framework; (b) it corrects for measurement error (i.e., disattenuation) to allow more veridical estimates of, for example, the true latent covariances of the constructs and the true construct means; and (c) it explicitly tests for equivalence of comparison (i.e., metric invariance), which is an important precondition for any comparison across groups. When metric invariance is obtained, the basic precondition for between-group comparisons that all comparisons are made on the same underlying constructs is supported (Baltes, Reese, & Nesselroade, 1988; Little, 1997). Furthermore, representing multiple constructs in a model assists the validity assessment because the constructs' relations with other constructs are explicitly represented (e.g., controls for suppressor effects, see, e.g., Saris & Stronkhorst, 1984). The maximum likelihood procedure was applied as a method of parameter estimation.

3.5.1.1 Controlling for Gender and Grade Effects

Because gender and grade level have been found to be influential factors in children's friendship relations (e.g., Buhrmester, 1990; Buhrmester & Furman, 1987; Furman & Buhrmester, 1985, 1992; Parker & Asher, 1993; Berndt & Perry, 1986) such effects could possibly bias any conclusions drawn from the analyses. Therefore, each of these covariates were represented by a single indicator in all specified SEM models and, hence, they were represented as error-free variables[55]. Gender was represented as a dummy-coded variable (females had the higher value). Consistent with those who stress the influence of the social

[55] The representation of gender and grade as error-free variables is reasonable because these constructs can be assumed to be reliably measured.

life or, more specifically, the school environment on children's sociocognitive development (e.g., Higgins & Parsons, 1983), children's age was represented in terms of grade in school (i.e., years of schooling). In addition to the linear effects of grade, the quadratic effects of grade were controlled. Quadratic trends were computed outside of LISREL by regressing age squared on age and saving the residuals (i.e., the quadratic component that is orthogonal to the linear component of age). In the models employed for the age-group comparisons only gender was represented in the models.

3.5.1.2 Evaluation of the Overall Model Fit

Model evaluation is not necessarily a simple procedure because no well-established guidelines for testing goodness-of-fit exist. The approach that is often recommended is to examine multiple goodness-of-fit indices rather than to rely on a single piece of information (e.g., Bollen, 1990; Marsh, Balla, & McDonald, 1988; Mulaik, James, Alstine, Bennett, Lind, & Stilwell, 1989; Raykov, Tomer, & Nesselroade, 1991; Tanaka, 1993). LISREL provides several fit indices for the model as a whole (for a review and assessment of these indices, see Marsh et al., 1988). In the present study, the following overall goodnes-of-fit measures were used to assess the fit between the hypothesized models and the sample data:

3.5.1.2.1 $\mathbf{x^2}$ - Value. $\mathbf{x^2}$ - Value with its associated degrees of freedom and probability level. When the sample size is sufficiently large, $\mathbf{x^2}$ is a likelihood ratio test statistic that can be used to test the fit between a restricted hypothesized model and the unrestricted sample data. A non-significant $\mathbf{x^2}$ value indicates that a certain model is tenable (i.e., it possesses acceptable fit). The assumption of the right model finds further support when the model specification is based on a strong theory and no additional estimates or relaxing of constraints have been necessary. The sensitivity of the $\mathbf{x^2}$ likelihood ratio test to sample size and model size, as well as to the violation of various model assumptions (linearity, multinormality, additivity) is widely known (see, e.g., Bollen, 1989). For example, with the enormous statistical power that comes when investigating large samples, almost every reasonable model will be rejected if only the $\mathbf{x^2}$ value and the associated probability are considered. For example, the assumption of no differences between the estimated and the observed matrix is regularly not tenable because of disturbances. According to Saris and Stronkhorst (1984), disturbances resulting in nonperfect fit are caused by (a) unknown variables, (b) known but

omitted variables (e.g., because the assessment of an exceedingly number of variables would be necessary for an exact description of social processes), (c) the randomness of human behavior, and (d) measurement error. Given the difficulties that are involved when interpreting the significance of the $\mathbf{x^2}$ value as a goodness-of-fit-index, the following goodness-of-fit indices were evaluated as well:

3.5.1.2.2 $\mathbf{x^2}$ / df Ratio. A variety of acceptable values for the $\mathbf{x^2}$ / df have been proposed, ranging from a low value of $\mathbf{x^2}$ / df < 1.50 through values of $\mathbf{x^2}$ / df < 3.0 to $\mathbf{x^2}$ / df <5.0. Following the recommendation of Byrne (1989), in the present study, a ratio < 2.0 is conceived to represent an adequate model fit.

3.5.1.2.3 Root-Mean-Square-Error-of-Approximation (RMSEA). The RMSEA indicates the average discrepancy between the elements in the sample and hypothesized covariance matrices. RMSEA values range from zero to one, thereby small values indicate a good model fit. Based on extensive experience it has been suggested that values of the RMSEA under .05 can be taken as indicative of reasonably good model fit (e.g., Browne & Cudeck, 1993). Thus, in the present study a value of RMSEA < .05 is used as indicating acceptable model fit.

3.5.1.2.4 Relative Goodness-of-Fit Indices. Relative or practical goodness-of-fit indices are especially important in cases of large samples when the power of the statistical test underlying the structural equation model approach is very high (see above). Relative goodness-of fit indices are derived from the comparison of some specified (i.e., restricted) model with a null model (i.e., usually one that posits complete independence of all measured variables). Thus, relative fit indices provide measures of complete covariation in the data. These indices range from zero to one. Generally, model fit is deemed to be acceptable when the resulting practical fit indices are of .90 or larger (Bentler & Bonett, 1980). The following three relative goodness-of-fit indices were used to assess overall model fit: The non-normed fit index[56] (NNFI, Bentler & Bonett, 1980; Tucker & Lewis, 1973), the incremental fit index (IFI; Bollen, 1989), and the comparative fit index (CFI, Bentler, 1990).

[56] The non-normed fit index is also known as the Tucker-Lewis index (RHO or TLI).

3.5.1.2.5 Local Model Fit. In addition to the overall model fit, local model fit, which refers to the model components was scrutinzed. Local model fit, is indicated when there are (a) sensible parameter estimates (e.g., no negative variances), (b) no salient standardized residuals that follow a normal ogive curve in the Q plot, and (c) no salient modification indices (i.e., no extremely high modification indices relative to the modification indices of other estimates). LISREL provides a modification index for each parameter in the model that is fixed; the modification index (Sörbom, 1989)[57] represents the improvement in the overall $\mathbf{x^2}$ test of model fit that would be achieved if that specific parameter were set free.

3.5.1.3 Testing for Factorial Invariance in Multi-Sample Analyses

LISREL methodology can be used to analyze data from a single sample at a certain time point. LISREL can also be used to analyse data from several samples[58] simultaneously with some or all parameters constrained to be equal across groups. In general, any degree of invariance can be tested, from the one extreme where all parameters are assumed to be invariant across groups, over time, or both to the other extreme when no constraints exist.

When employing MACS models, six types of parameter estimates can be evaluated: (a) the unstandardized regressions of the indicators on the factors (i.e., the common factor loadings of the indicators), (b) the intercepts or means of the indicators, (c) the residual variances of the indicators (i.e., the unique factors of the indicators or the measurement errors), (d) the latent factor means, (e) the latent factor variances, and (f) the latent factor covariances or correlations (McArdle & Cattell, 1994; Meredith, 1993). The first three types of parameters belong to the measurement model and the latter three represent components of the latent mean and covariance structure or latent variable model (Little, 1997). The latent variable model specifies the structural equations that encompass the relationships between latent variables or constructs (e.g., factor variances, covariances and means). The measurement model summarizes the structural equations that represent the link between the latent and observed variables (e.g., factor loadings or unstandardized factor regressions and the intercepts or means of the indicators). Thereby, the observed or manifest variables (indicators) of a latent variable contain measurement errors and unique variance due to related

[57] The modification index is equivalent to the Lagrange multiplier test (Bentler, 1986).
[58] The LISREL methodology can also be extended to several time points.

constructs which are not included in the model (i.e., specific variance not common with the other measured variables defining the same latent construct). These disturbances are represented as the residual variances of the indicators. Notably, in all conducted models the residual variances were specified to be uncorrelated. Moreover, no dual factor loadings were allowed in any of the conducted models. Appendix F provides a more detailed description of the specification of the models.

3.5.1.3.1 Factorial invariance. Factorial invariance refers to the equivalence of both the measurement level and the latent-variable level of a given model across various samples. Factorial invariance is established by the assessment of both measurement invariance and structural invariance. Measurement invariance ist established for the measurement model and structural invariance is established for the latent variable model. Testing factorial invariance follows a two step process. First, the invariance of the measurement model is tested. Measurement invariance addresses whether a construct is measured the same way when comparing two or more groups of individuals (e.g., Horn, 1991; Little, 1997; MacCallum & Tucker, 1991). Consequently, measurement invariance is a precondition for testing the invariance of the structural model (e.g., P. B. Baltes, Reese, & Nesselroade, 1978; Horn & McArdle, 1992; Little, 1997). If one does not know that measurement invariance obtains, any differences in the structural components (e.g., differences in factor means and covariances) could be interpreted as indicating qualitatively different constructs, or quantitative differences between constructs, or some combination of both. To test that the constructs of the present study do possess the same meaning across groups, both configural invariance assumptions and metric invariance assumptions (see e.g., McArdle & Cattell, 1994) were tested. Configural invariance means that loading patterns fitted to the data of different groups contain the same salient and nonsalient loadings, but the former are allowed to differ in magnitude among the groups (i.e., data sets). In addition to the salient factor loadings also the variances, covariances, intercepts, and uniqueness are free to vary across the groups. When applying the stricter criterion of metric invariance, the factor loadings and the intercepts of the indicators are constrained to be equal[59]. Appendix F provides a more detailed description of the

[59] The invariance of the residual variances of the indicators was not tested, because residual variances represent both errors and unique factors. Because errors are assumed to be random, it makes little sense to expect that the

procedures implemented to test measurement invariance. Assuming measurement invariance was obtained, differences (or similarities) between the selected groups regarding the latent constructs' moments (i.e., factor variances, factor covariances, and factor means) can be evaluated.

3.5.1.3.2 Two Rationales for Testing Factorial Invariance. The invariance of the factor structure across groups of participants was analyzed using hierarchical model testing. An equivalence test is conducted as a nested-model comparison between a model in which specific parameters are constrained to equality across groups and one in which these parameters (and all others) are freely estimated in all groups. Two different rationales can be used for testing cross-group restrictions (i.e., factorial invariance): A statistical rationale and a modeling rationale (Little, 1997). Adopting a statistical rationale, the difference in $\mathbf{x}^2$ between the two models is a test of the equality restrictions (with degrees of freedom equal to the difference in their degrees in freedom). If the $\mathbf{x}^2$-difference test is nonsignificant then the statistical evidence indicates no cross-group differences between the equated parameters. If the test is significant, then evidence of cross-group inequality exists. With a modeling rationale, on the other hand, model constraints are evaluated by comparing the practical fit indices of the restricted and the less restricted models. Small differences (e.g., NNFI <= .05) in practical levels of fit between the restricted and the less restricted model are considered to be negligible if (a) overall model fit is acceptable, (b) indices of local mis-fit are uniformly and unsystematically distributed for the constrained parameters (residuals were normally distributed and modification indices were uniformly negligible), and (c) the accepted model is substantively more meaningful and parsimonious (i.e., no "garbage" parameters are estimated in any group and the equivalence model contains fewer parameter estimates than the non-invariance model) than the alternative model (see Little, 1997).

A modeling rationale is appropriate if large models with numereous constrained parameters are evaluated because the $\mathbf{x}^2$ statistic is very sensitive to sampling errors, especially when estimated on large sample sizes (e.g., Marsh, et al., 1988). Specifically,

---

residuals would be invariant across samples (MacCallum & Tucker, 1991). If the degree of random error is not exactly equal across groups, the nonequal portions of the random error are forced into other parameters of a given model, thereby indroducing potential sources of bias (see, Little, 1997).

testing measurement invariance involves evaluating the general tenability of an imposed indicator-to-construct structure via overall model fit indices. Thus, in such cases Little (1997) recommends to employ a modeling rationale for assessing measurement invariance. However the author refers to studies where samples stemming from various cultural contexts are compared. Studies involving various different contexts are more likely to involve sources of bias and error such as cultural bias, translation errors, varying conditions of administration than a study that was conducted in a single sociocultural context under standardized assessment procedures. The present study took place under the latter circumscribed conditions and, in addition, the subsamples (i.e., sociometric-status groups and age groups) stem from the same overall sample. As described in Section 3.2.1, the overall sample did not evince systematic differences across the three school contexts. In addition and importantly, the models assessing factorial invariance of the various sets of constructs (i.e., agency beliefs, means-ends beliefs, action-control strategies, self-rated friendship quality, goal difficulty and importance) separately don't involve more than three constructs and, thus, can be characterized as being small models. In contrast, models assessing the interrelationships among the various sets of constructs are larger models. In the present study a modeling rationale was employed when testing large models (i.e., number of constructs >= 3) while a statistical rationale was employed when testing measurement invariance of small models (i.e., number of constructs < 3).

Generally, a statistical rationale will be employed when testing substantive hypotheses concerning the latent parameters because the latent level reflects error-free effects among constructs (see, Little, 1997). The substantive hypotheses were tested by constraining the respective parameters to be equal in a model which was nested in the measurement-invariant model. Then, the $\mathbf{x^2}$ statistic of this model and was compared with the $\mathbf{x^2}$ statistic of the measurement-invariant model. If the difference of the $\mathbf{x^2}$ statistic was associated with a significant p-value the constrained parameters that represented sources of misfit were identified by scrutinizing the indicators of local misfit (see Section 3.5.1.2.5). In a next step, one of these parameters was freed and the $\mathbf{x^2}$ statistic of this model was compared to the $\mathbf{x^2}$ statistic of the previous model. If the resulting $\mathbf{x^2}$ difference indicated a significant increment in fit the freed parameter was accepted to be significantly different. In the following step, the

χ^2 statistic of the model with the freed parameter was compared to the χ^2 statistic of the measurement-invariant model. Generally, if the comparison of the χ^2 statistic of the model containing a freed parameter and the χ^2 statistic of the measurement-invariant model results in a nonsignificant χ^2 difference all potential sources of misfit have been identified. If the resulting χ^2 difference still indicated a significant decline in fit another constrained parameter that represented a source of misfit was identified by scrutinizing the indicators of local misfit. Then, the same univariate testing procedures were applied as for the previous parameter in order to test whether this parameter was significantly different from the remaining constrained parameters.

3.5.1.3.3 Significance levels. Generally, testing measurement invariance aims at rejecting the alternative hypothesis of existing cross-group differences and accepting the null hypothesis of no cross-group differences. Thus, a Type-II error protection may be of concern (e.g., Bortz, 1989; Bortz, Lienert, & Boehnke, 1990; Cohen, 1962; Cohen & Cohen, 1983). On the other side, testing measurement invariance involves many parameters and may be affected by measurement errors. Hence, these omnibus tests are very powerful and minor sampling errors or fluctuations may lead to reject the null hypothesis resulting in a Type-I error. The significance level of testing measurement invariance was set at $\alpha < .01$. However, if measurement errors are more likely to affect testing of measurement equivalence the employment of a modeling rationale is indicated (e.g., Little, 1997). Thus, I employed for models which were comprised of more than 4 constructs a modeling rationale. When measurement invariance was enforced and the overall model fit of larger models was still satisfactory, I accepted the measurement invariant models.

When testing substantive hypotheses, alternative hypotheses will be accepted when the probability of Type I error is smaller than five percent. When testing the null hypothesis of no cross-group differences on latent parameters, a Type-II error protection will be taken into account by specifying the Type I error level at $\alpha < .20$ (cf., Bortz, 1989; Bortz, Lienert, & Boehnke, 1990; Cohen, 1962; Cohen & Cohen, 1983). Specifically, when testing equivalence of covariate effects (i.e., gender as well as grade, and quadratic effects of grade) across groups a significane level of $\alpha <= .20$ was employed. However, when testing the hypothesis that both friendship status and sociometric status do not moderate the relationships among the

constructs (see Hypothesis B1) a significane level of α <= .10 was employed. This significance level was justified because on the one side, generally, the statistical power of the conducted tests was negatively affected by the rather small sample sizes of three of the investigated groups (i.e., friended rejected children, friendless rejected children, and average children). On the other side, the large number of tested parameters increased the probability of significant results due to chance.

Finally, in order to counteract chance-related significant results multivariate tests were conducted. Follow-up univariate tests were only conducted if a previously conducted multivariate test was significant.