

# A Re-Examination of the Validity and Reliability of the School-Age Temperament Inventory

Sandra Graham McClowry ▼ Charles F. Halverson ▼ Ann Sanson

- ▶ **Background:** As a construct, temperament provides a framework for understanding differences among individuals in reaction to their life experiences. The measurement of the construct concerns both researchers and clinicians.
- ▶ **Objective:** The purpose of this study was to examine whether the School-Age Temperament Inventory continued to demonstrate reliability and validity when retested with three existent samples of parent respondents.
- ▶ **Method:** Sample 1 was a sociodemographically and racially heterogeneous group of 200 children from New England in the United States. Data for Sample 2 was provided by 589 mothers and fathers from the state of Georgia in the United States. In Sample 3, data was provided by parents (principally, mothers) of 1,391 adolescents from Australia. Orthogonal Procrustes rotations were conducted to examine the underlying structure of the inventory when it was contrasted with the results obtained in the original standardization of the tool.
- ▶ **Results:** The total coefficients of congruence for the samples were .88 to .97, while those of the four factors ranged from .84 to .98. Across the samples, Cronbach alphas for the dimensions ranged from .80 to .92. Independent *t*-tests identified that boys were significantly more active and less task persistent than girls. However, regression analyses revealed that sex accounted for only 5% of the variance in task persistence and activity.
- ▶ **Discussion:** The results provide substantial additional support for the reliability and validity of the School-Age Temperament Inventory. Recommendations for future research are offered which include exploring the role of temperament in contributing to developmental outcomes in children and examining cross-cultural samples.
- ▶ **Key Words:** child • research instruments • temperament

As a stylistic individual characteristic, temperament is an integral facet of human development. Temperament is the consistent reaction style that an individual demonstrates across a variety of settings and situations,

particularly those that involve stress or change (McClowry, 2003). An understanding of the construct of temperament can assist nurses and other clinicians to recognize why clients react differently to life experiences, plan clinical strategies that appropriately account for individual differences, and assess the effectiveness of various interventions.

Most theorists in the temperament field and clinicians who use it as a guiding framework for their practice acknowledge that temperament encompasses interactive biological and environmental components (Rothbart & Bates, 1998; Chess & Thomas, 1984). The complexity of these transactions is reflected in the measurement of temperament. Although a variety of methods exist, such as interviews and observational techniques, the most frequently used is Likert-type questionnaires because they are easy to administer and low in cost.

When gathering data on child temperament, parents, because they have multiple opportunities to observe their children in a variety of situations, are often asked to be informants. The reliability and validity of parental reports, however, remains a source of debate for several reasons (Rothbart & Bates, 1998). According to Kagan (1998), parental perceptions of their children's temperament can be biased by their own personality and behavioral expectations. Parents may also have a limited knowledge of normative growth and development on which to compare their child with others. Moreover, the expression of temperament and respondents' perception of it, are also influenced by age, ethnicity, race, sex, and socioeconomic status (Kohnstamm, 1989). Thus, temperament, although stable, is not static. Instead, it is influenced over time by the environment and by biological maturation as various genes are activated.

---

*Sandra Graham McClowry, PhD, RN, FAAN, is Professor, Division of Nursing, Steinhardt School of Education, New York University, New York.*

*Charles F. Halverson, Jr. PhD, is Professor, Department of Child and Family Development, University of Georgia, Athens.*

*Ann Sanson, PhD, is Associate Professor, Deputy Director, Australian Institute of Family Studies, Melbourne, Victoria, Australia.*

Parental reports of temperament are also controversial, because the questionnaires have been developed primarily with homogeneous samples of White, non-Hispanic, middle class families from the United States (US). In addition, many of the instruments have not undergone extensive psychometric evaluation. Instead, many researchers have assumed that questionnaires, once developed, are valid and can be used reliably in their studies. Nunnally and Bernstein (1994), however, assert that instrumentation validation is an unending process.

Large epidemiological cross-cultural samples with representative samples are needed to thoroughly explore the many compounded variables that potentially influence parental reports of child temperament. The principal disadvantages of such studies are the cost and time involved. A more modest approach is to examine whether the psychometric estimates of temperament instruments remain stable when completed by samples that differ from the one engaged in the development of the tool. Such studies can be low in cost when conducted on existent data sets. They also can provide additional evidence of reliability and validity of an instrument and focus future investigations. By examining the results in relation to the differences among the samples, hypotheses may be generated regarding the influence of compounding factors that can be best answered with large epidemiologically-based cross-cultural studies.

McClowry (1995) developed the School-Age Temperament Inventory (SATI) with a predominantly White, non-Hispanic middle class sample. The purpose of the present study was to re-examine the reliability and validity of the SATI when completed by three additional samples of parent respondents.

### The School-Age Temperament Inventory

The SATI was developed (McClowry, 1995) as a parental report of the temperament of children 8–11 years of age. The conceptualization of the instrument was based on a review of item-based factor analytic studies of child temperament questionnaires. Four temperament dimensions were hypothesized: (a) negative reactivity, (b) task persistence, (c) approach/withdrawal, and (d) energy.

The initial development of the tool included 435 mothers and 228 of their spouses as parent informants. The families were recruited from three New England school districts. Although the families were from all socioeconomic groups as measured on Hollingshead's (1975) Four Factor Index, the majority was middle class. Sixty-two percent had clerical or technical occupations. The mothers reported that a majority of the children (89%) were White, non-Hispanic.

To examine the underlying structure of the SATI, maternal data were subjected to principal components factor analysis with varimax rotation. In general, the four conceptualized dimensions were supported as follows: (a) negative reactivity described the intensity and frequency

**Parents, because they have multiple opportunities to observe their children, are often asked to be informants**

with which the child expresses negative affect; (b) task persistence depicted the degree of self-direction that a child exhibited in fulfilling tasks and other responsibilities; (c) approach/withdrawal portrayed the child's initial response to new people and situations; and (d) activity described the amount of large motor activity of the child.

Once developed, the SATI consisted of 38 items with a Likert-type format ranging from *never* (1) to *always* (5). After reverse coding the appropriate items, scores were obtained by averaging the items in each dimension. Higher scores indicated that the child was high in negative reactivity, task persistent,

had a tendency to withdraw from new situations, and was highly active.

Cronbach alphas for the dimensions (provided by the mothers and fathers, respectively) were .85 to .90 for the data. Correlations between maternal and paternal reports were .51 to .68. Test-retest reliability after 4–6 months was .80 to .89. The impact of age and sex was minimal, never explaining more than 5% of the variance.

### Factor Analytic Studies of Child Temperament

Although the SATI appeared to have adequate reliability and validity when completed by a group of predominately middle class White parents from New England, the appropriateness of its use with other samples was unknown. Reports on other temperament questionnaires also included limited evidence of testing for performance across diverse groups, as is evident from an examination of item-based factor analytic studies of child temperament questionnaires: Boer and Westenberg (1994), McClowry, Hegvik, and Teglassi (1993), Presley and Martin (1994), Sanson, and Smart, Prior, Oberklaid, and Pedlow (1994).

None of these studies includes a heterogeneous cross-national sample with representative subsamples of diverse race, ethnicity, and socioeconomic status. The reported race/ethnicity of the families in the studies were predominantly White, non-Hispanic (Keogh et al., 1982; Presley & Martin, 1994; Rowe & Plomin, 1977; Windle & Lerner, 1986) or were not described at all (Baker & Velicer, 1982; Czeschlik, 1992; Finegan et al., 1989; Sanson et al., 1994). The studies that did not discuss race were conducted in countries whose populations are predominantly White and many were conducted before the recent emphasis on ethnic diversity. The study by Czeschlik (1992) was conducted in Germany. Boer and Westenberg (1994) studied children from the Netherlands. The sample in Sanson et al. (1994) was Australian and Finegan (1989) et al.'s study was conducted in Ontario. The rest of the studies were conducted in various parts of the US. Four were in the Northeast (Baker & Velicer, 1982; McClowry et al., 1993; McClowry, 1995; Windle & Lerner, 1986) and three were in the West (Buss & Plomin, 1975; Rowe & Plomin, 1977). Presley and Martin (1994) combined samples from the Northeast, West, and Southeast.

The socioeconomic status of the families was predominantly middle class (Boer & Westenberg, 1994; McClowry et al., 1993; McClowry, 1995; Presley & Martin, 1994; Windle & Lerner, 1986), or middle class and upper class (Buss & Plomin, 1975; Rowe & Plomin, 1977), or upper class alone (Finegan et al., 1989). In two cases, the entire range of classes were included (Keogh et al., 1982; Sanson et al., 1994). No mention of socioeconomic status was made in two other studies (Baker & Velicer, 1982; Czeschlik, 1992).

The failure to check the psychometrics of an instrument across a variety of samples is evident in the temperament field as it is in many areas of child development. Without such psychometric data, it is easy to assume that any statistically significant differences obtained by comparing a new sample with the one used in the initial development of the tool are theoretically and clinically meaningful. Kohnstamm (1989) reported that the temperament field has used two methods for cross-sample comparisons: (a) factor loadings (usually obtained from principal components factor analysis) visually examined for similarities and differences, or (b) means and *SDs* compared between samples. Another way is to examine congruence coefficients calculated after conducting an orthogonal Procrustes rotation (Hurley & Cattell, 1962; McCrae, Zonderman, Costa, Bond & Paunonen, 1996) results in an assessment for factor similarity of personality inventories. Procrustes rotation examined the degree of fit among factor structures derived from new samples (when compared to the first one). When applying the most conservative type (orthogonal Procrustes rotation) factors are rotated to minimize the sums of squares of deviation from the target matrix (under the constraint of maintaining orthogonality) (McCrae et al., 1996).

To further assess the reliability and validity of the SATI across three additional samples of parental respondents, two research questions were asked: (a) Do the same the dimensions of negative reactivity, task persistence, approach/withdrawal, and activity emerge across the samples? and (b) Do the factors emerging across the samples show high replicability?

## Method

### Subjects

Data from three existent samples of parent respondents were compared with the results obtained in the original development of the questionnaire (McClowry, 1995). There were 200 maternal reports from New England in Sample 1. The children were 8–11 years of age with an average age of 9.51 years ( $SD = 1.02$ ). Fifty-one percent were girls. The average age of the mothers was 35.9 years ( $SD = 5.97$ ). The reported ethnicity of the children was White, non-Hispanic (59%), Black (22%), Hispanic (16%), and Native American (2.5%). Fifty-four percent of the children lived with their mothers and fathers, 36.5% lived in homes headed by single mothers, and 9.5% lived

***Parental perceptions of their children's temperament can be biased by their own personality and behavioral expectations***

in blended families. Based on the Hollingshead (1975) Four Factor Index, the socioeconomic status of the families were unskilled (16%), semi-skilled (17%), clerical (30%), technical (26.5%), and professional (10.5%).

Sample 2 included 589 children from Georgia in the US. The data included 451 maternal and 138 paternal reports. Fifty-three percent of the children were girls. The children were 8–13 years of age with an average age of 10.3 years ( $SD = 1.09$ ). The sample consisted of White, non-Hispanic (87%), Black (9%), Asian (2%), and Hispanic, non-Black (1%). The children's mothers (a) had not finished high school (3%), (b) had some college (11%), (c) had graduated from college (43%), or (d) had completed at least some graduate school (43%).

Sample 3 included 1,391 Australian 13–14 year olds ( $M = 13.78$ ,  $SD = 2.74$ ) who were members of the longitudinal Australian Temperament Project (ATP) (Prior, Sanson, Smart, & Oberklaid, 2000). Most of the children's parents were born in Australia (84% of the mothers and 78% of the fathers). An additional 12% came from Europe, primarily from the United Kingdom. The remaining mothers (4%) and fathers (8%) came from a variety of countries (predominately Asian). The sample was approximately 90% White, but it reflected the multicultural nature of the Australian society. Thirty-one percent of the families lived in rural areas. Fifty-two percent were boys. Eighty-three percent of the children lived with their mothers and fathers, 11% were from single-parent homes, and 5% lived in blended families. Parental educational levels varied from 8<sup>th</sup> grade or lower (4%) to postgraduate degrees (20%). The average educational level was grade 11–12. The fathers' occupations included: semiskilled and unskilled (18%), clerical or skilled (41%), and professional (40%). The mothers were classified as unskilled or semiskilled (22%), clerical or skilled (52%), or professional (27%).

### Design and Procedure

Maternal informants for Sample 1 were recruited from two school districts by sending letters of invitation to approximately 1,800 mothers of children in regular 3<sup>rd</sup> to 5<sup>th</sup> grade classrooms. Mothers were instructed to indicate on an enclosed stamped postcard if they were willing to participate. A total of 293 postcards were received. Packets with stamped return envelopes were then mailed to the homes of the 278 mothers who met the criteria for participation. A 73% return rate yielded 204 responses of which all but four were useable.

To obtain Sample 2, alumni from the College of Family and Consumer Sciences of the University of Georgia were contacted by letter. Parents, willing to participate, were asked to return the enclosed postcard. Packets with stamped return envelopes were then mailed to their homes. Sixty-four percent of the alumni who had children returned their questionnaires.

Sample 3 included the participants in the ATP who had already participated in the previous eight waves of the study. Questionnaires were also sent to the families' homes with stamped return envelopes. A majority (84%) of those sent questionnaires returned them. To adapt the SATI for an Australian population, slight alterations were made in the wording of a few items. In addition, because the subjects were older, item 28, "Runs to get where he/she wants to go," was reworded to "Moves fast to get where s/he want to go;" and item 13, "Runs when entering or leaving the house" was changed to "Prefers active outdoor activities to quiet indoor ones." This item was excluded from the analysis since there were no data similar to it in the original sample.

## Results

The first question asked whether the SATI continues to support the dimensions of negative reactivity, task persistence, approach/withdrawal, and activity. The analysis for this question was conducted in several steps. First, separate principal components factor analyses with varimax rotation were conducted on Samples 1, 2, and 3. Based on each of the scree plots, four factors were retained with eigenvalues  $> 1$ . The total amount of explained variance was: Sample 1 (81%), Sample 2 (57%) and Sample 3 (56%).

Next, the varimax matrixes obtained in the first step were rotated by orthogonal Procrustes procedure (Schonemann, 1966) to the targeted one that was reported by McCloskey (1995). Finally, the coefficients of congruence of the rotated factor patterns were estimated for proportional similarity (Harman, 1976). The results included an overall fit coefficient, one for each of the dimensions, and for the individual items.

The factor analytic results demonstrated a great deal of similarity to the original solution. All items across the samples loaded most strongly on the appropriate factor, thus supporting the four-dimensional structure of the instrument (Table 1).

Internal consistency estimates using Cronbach alphas also supported the original SATI solution. The alphas ranged from .80 to .92 for the four dimensions and are shown in comparison with the maternal data from the original sample (Table 2).

Replicability, which was the focus of the second question, was assessed by examining the coefficients of congruence. The total coefficient of congruence was .97 for Sample 1, .92 for Sample 2, and .88 for Sample 3. Across the samples, the coefficients of congruence of the factors range from .84 to .98. The item variable coefficients of congruence ranged from .56 to 1.00, averaging .92 over the three sets (Table 3).

Replicability was further tested by exploring whether the means and standard deviations of the dimensions differed across the samples. An analysis of variance (ANOVA) was conducted on each of the dimensions with the sample as the between-subject variable. Bonferroni post hoc comparisons were conducted for significant effects. Because the total sample was large, and thus might provide significant, but clinically meaningless results, linear regression with effects coding was used to determine

the magnitude of the identified effects on temperament. Only effects accounting for more than 3% of variance were considered.

Significant differences were found across the samples on the dimensions of negative reactivity, task persistence, and approach/withdrawal. However, the regression analysis showed that the differences in the samples only accounted for 2% of the variance in negative reactivity and 1% in task persistence and approach/withdrawal. Such differences were not regarded as clinically significant and were not further interpreted.

Finally, replicability was examined by exploring whether the means and standard deviations of the dimensions differed by sex. In McCloskey (1995) boys were significantly higher in activity and lower in task persistence. First independent tests were conducted for each of the dimensions with the three samples combined. Significant sex differences were found in task persistence and activity. Boys ( $M = 3.42$ ,  $SD = 0.80$ ) were significantly less persistent than girls ( $M = 3.78$ ,  $SD = 0.73$ ),  $t(2562) = 12.00$ ,  $p < .001$ . Girls ( $M = 2.82$ ,  $SD = 0.69$ ), however, were significantly less active than boys ( $M = 2.82$ ,  $SD = 0.75$ ),  $t(2597) = -11.07$ ,  $p < .0001$ . Regression analysis revealed that sex (dummy coded), accounted for 5% of the variance in task persistence and in activity. Table 4 summarizes the means and standard deviations of the dimensions by sex. The data from the original sample is also included.

## Discussion

The results of this study provide additional evidence of the validity and reliability of the SATI. The presented data demonstrates a remarkable and robust convergence of factor structures when completed by samples that are primarily White middle class parents varying in geographic locations. The results also yield additional support for the construct validity of four dimensions of school-age temperament: (a) negative reactivity, (b) task persistence, (c) approach/withdrawal, and (d) activity. These four dimensions, with slightly different names, have been consistently found in factor analytic studies of childhood temperament measures (Martin, Wisenbaker, & Huttunen, 1994; Rothbart & Bates, 1998).

Interpreting the results of the Procrustes rotations is more complicated because the level of the congruence coefficient (required for determining replicability) remains controversial. Factor equivalence is generally accepted as .90 or above (McCrae et al., 1996; Mulaik, 1971). However, Barrett (1986, p. 337) regarded .80 as a "conceptual lower bound."

Based on these guidelines, the replicability of the factors is supported. An excellent fit occurred with Sample 1; Sample 2 was nearly as good. An adequate (but less satisfactory fit) emerged with Sample 3 perhaps due to the older age of the subjects in this sample or because they were from Australia rather than the US. All of the overall congruence coefficients and those for the dimensions exceeded the lower bound level of .80. Some concern, however, is warranted regarding task persistence on an item level in Samples 2 and 3 since they included four items on which the variable congruence was less than .80.

TABLE 1. Weighted Factor Loadings From Samples 1, 2, and 3, and Commonalities

Item Number and Stem Commonalities	Factor				Commonalities
	NR	TP	A/W	Act	
<i>Negative Reactivity</i>					
23 When angry, yells or snaps at others	.79	.05	.04	.12	.64
29 Responds intensely to disapproval	.78	.11	.06	.12	.65
26 Moody when corrected for misbehavior	.75	.10	.13	.02	.60
32 Makes loud noises when angry	.74	.10	.02	.02	.59
14 Reacts strongly to a disappointment	.73	.12	.05	.15	.57
10 Gets mad when mildly criticized	.73	.09	.12	.01	.55
33 Gets upset when there is a change in plans	.66	.14	.17	.12	.50
20 Gets very frustrated	.64	.14	.15	.18	.48
5 When he/she disagrees, speaks quietly <sup>a</sup>	.63	.12	.02	.23	.47
17 Gets angry when teased	.63	.07	.12	.02	.42
2 Gets upset when he/she can't find something	.62	.09	.07	.12	.42
37 Has off days when he/she is moody	.54	.15	.13	.00	.33
<i>Task Persistence</i>					
22 Stays with homework until finished	.04	.84	-.00	.06	.71
8 Does not complete homework <sup>a</sup>	.04	.82	-.01	.07	.70
30 Has difficulty completing assignments <sup>a</sup>	.08	.78	.01	.04	.62
16 Remembers to do homework	.02	.78	-.00	.07	.61
6 Returns to responsibilities	.13	.77	.03	.04	.61
25 Goes back to the task at hand	.15	.77	.06	.06	.63
18 Quits routine household chores <sup>a</sup>	.23	.74	.02	.02	.60
11 Leaves own projects unfinished <sup>a</sup>	.11	.68	.02	.05	.48
15 Gets frustrated with projects and quits	.25	.63	.07	.08	.47
36 When an activity is difficult, gives up <sup>a</sup>	.25	.63	.10	-.00	.47
4 Switches from one activity to another <sup>a</sup>	.18	.60	.02	.12	.41
<i>Approach/Withdrawal</i>					
21 Bashful when meeting new children	.09	.01	.80	-.01	.65
9 Is shy with adults he/she doesn't know	.02	-.00	.78	-.00	.62
38 Seems uncomfortable at someone's house	.11	.06	.75	.06	.59
34 Avoids new guests	.09	.06	.74	.06	.57
3 Approaches children his/her age <sup>a</sup>	.02	-.10	.69	-.00	.48
12 Seems nervous or uncomfortable in a new place	.18	.04	.69	.01	.51
27 Moves right into a new place <sup>a</sup>	.13	.06	.69	-.10	.50
7 Smiles or laughs with new adults	.04	.09	.66	-.00	.45
31 Prefers to play with someone he/she knows	.11	-.10	.59	-.00	.37
<i>Activity</i>					
13 Runs when entering or leaving	.16	.03	-.00	.79	.65
28 Runs to where he/she wants to go	.09	.07	-.10	.77	.61
24 Runs or jumps when going down	.18	.11	.05	.74	.59
35 Seems to be in a hurry	.19	.08	.01	.70	.53
19 Bursts loudly into the room	.35	.17	-.00	.64	.56
1 Walks quietly in the house <sup>a</sup>	.26	.15	-.00	.61	.46

Note. The commonalities listed in Table 1 were derived from the averaged loadings for each variable.

<sup>a</sup>Reverse coded before factor analyzed.

**TABLE 2. Cronbach Alpha of the SATI Dimensions**

Dimensions	Original	Sample 1	Sample 2	Sample 3
Negative reactivity	.90	.89	.90	.92
Task persistence	.90	.89	.91	.92
Approach/withdrawal	.88	.84	.86	.92
Activity	.85	.80	.86	.92

Further evidence of reliability was also found. The Cronbach alphas found that these three samples, like those from McClosry (1995), were at least .80 for each of the dimensions. The finding that girls are less active and more task persistent is also consistent with McClosry (1995). In that study, sex contributed 4% of the variance versus the 5% in this study. The variance in task persistence was 5% in both studies.

Examined in their totality, the results lend considerable support to the validity and reliability of the SATI and its four dimensions. The tool also demonstrates good psychometric properties for young adolescents in addition to the school-age children for whom it was originally intended. It is also likely that the derived factors will prove useful in studying developmental outcomes of interest to nurses and other clinicians. ▀

Accepted for publication February 1, 2003.

Portions of this study were funded by the National Institute for Nursing Research (NR03059), the Fan Fox and Leslie R. Samuels Foundation, the National Institute of Mental Health, and the National Health and Medical Research Council of Australia.

The authors thank Paul T. Costa, Jr., PhD, Chief, Laboratory of Personality & Cognition, National Institute on Aging, for his assistance in the development of this article.

Corresponding author: Sandee Graham McClosry, PhD, RN, FAAN, Division of Nursing, Steinhardt School of Education, New York University, 246 Greene Street, 8<sup>th</sup> Floor, New York, NY 10003 (e-mail:sm6@nyu.edu).

**TABLE 3. Summary of the Coefficients of Congruence Rotated to Original Sample**

Dimensions	Sample 1	Sample 2	Sample 3
Negative reactivity	.98	.91	.84
Task persistence	.96	.88	.84
Approach/withdrawal	.98	.97	.97
Activity	.93	.92	.88
Total coefficient of congruence	.97	.92	.88

Note. SATI = ???

**TABLE 4. Means and Standard Deviations of the SATI Dimensions by Sex**

Dimensions	Boys	Girls
Negative reactivity	2.96 (0.75)	2.91 (0.73)
Task persistence	3.42 (0.80)	3.78 (0.73)
Approach/withdrawal	2.58 (0.72)	2.57 (0.72)
Activity	2.82 (0.75)	2.51 (0.69)

**References**

Baker, E. H., & Velicer, W. F. (1982). The structure of reliability of the Teacher Temperament Questionnaire. *Journal of Abnormal Child Psychology*, 10, 531-545.

Barrett, P. (1986). Factor comparison: An examination of three methods. *Personality and Individual Differences*, 7, 327-340.

Boer, F., & Westenberg, P. M. (1994). The factor structure of the Buss and Plomin EAS Temperament Survey (Parental Ratings) in a Dutch sample of elementary school children. *Journal of Personality Assessment*, 62, 537-551.

Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development*. New York: Wiley.

Chess, S., & Thomas, A. (1984). *Origins and evolution of behavior disorders: From infancy to early adult life*. New York: Brunner/Mazel.

Czeschlik, T. (1992). The Middle Childhood Temperament Questionnaire: Factor structure in a German sample. *Personality and Individual Differences*, 13, 205-210.

Finegan, J. A., Niccols, A., Zacher, J., & Hood, J. (1989). Factor structure of the Preschool Characteristics Questionnaire. *Infant Behavior and Development*, 12, 221-227.

Harman, H. H. (1976). *Modern factor analysis* (3<sup>rd</sup> ed.). Chicago: University of Chicago Press.

Hollingshead, A. B. (1975). *Four Factor Index of Social Status*. New Haven, CT: Department of Sociology, Yale University.

Hurley, J. R. & Cattell, R. B. (1962). The Procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral Science*, 7, 258-262.

Kagan, J. (1998). Biology and the child. In W. Damon (Ed.), *Handbook of child psychology*, Vol. 3 (5<sup>th</sup> ed., pp. 177-235). New York: Wiley.

Keogh, B. K., Pullis, M. E., & Cadwell, J. (1982). A short form of the Teacher Temperament Questionnaire. *Journal of Education Measurement*, 19, 323-329.

Kohnstamm, G. A. (1989). Temperament in childhood: Cross-cultural and sex differences. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 483-508). New York: Wiley.

Martin, R. P., Wisenbaker, J., & Huttunen, M. (1994). Review of factor analytic studies of temperament measures based on the Thomas-Chess structural model: Implications for the Big Five. In C. F. Halverson, Jr., G. A. Kohnstamm, and R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 157-172). Hillsdale, NJ: Erlbaum.

McClosry, S. G. (1995). The development of the School-Age Temperament Inventory. *Merrill-Palmer Quarterly*, 41, 271-285.

McClosry, S. G. (2003). *Your child's unique temperament: Insights and strategies for responsive parenting*. Champaign, IL: Research Press.

- McClowry, S. G., Hegvik, R. L., & Teglasi, H. (1993). An examination of the construct validity of the Middle Childhood Temperament Questionnaire. *Merrill-Palmer Quarterly*, 39, 279-293.
- McCrae, R. R., Zonderman, A. B., Costa, Jr., P. T., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, 70, 552-566.
- Mulaik, S. A. (1971). *The foundations of factor analysis*. New York: McGraw-Hill.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Presley, R., & Martin, R. P. (1994). Toward a structure of preschool temperament: Factor structure of the Temperament Assessment Battery for Children. *Journal of Personality*, 62, 415-448.
- Prior, M., Sanson, A., Smart, D., & Oberklaid, F. (2000). *Pathways from infancy to adolescence: Australian Temperament Project 1983-2000*. Melbourne, Australia: Australian Institute of Family Studies.
- Rothbart, M. K., & Bates, J. E. (1998). Temperament. In W. Damon (Ed.), *Handbook of child psychology* (5th ed., pp. 105-174). New York: Wiley.
- Rowe, D. C., & Plomin, R. (1977). Temperament in early childhood. *Journal of Personality Assessment*, 41, 150-156.
- Sanson, A.V., Smart, D.F., Prior, M., Oberklaid, F., & Pedlow, R. (1994). The structure of temperament from age 3 to 7 years: Age, sex and sociodemographic influences. *Merrill-Palmer Quarterly*, 40, 233-252.
- Schoenemann, P. H. (1966). A generalized solution of the Orthogonal Procrustes problem. *Psychometrika*, 31, 1-10.
- Windle, M., & Lerner, R. M. (1986). Reassessing the dimensions of temperamental individuality across the life span: The Revised Dimensions of Temperament Survey (DOTS-R). *Journal of Adolescent Research*, 1, 213-229.