Psychological Assessment
1998, Vol. 10, No. 3, 241-249

1040-3590/98/$3.00

Structure and Validity of Parent and Teacher Perceptions of Children's Competence: A Multitrait–Multimethod–Multigroup Investigation

David A. Cole, Dawn M. Gondoli, and Lachlan G. Peeke
University of Notre Dame

Parents and teachers completed their respective versions of S. Harter's (1985b) Rating Scale of Children's Actual Behavior to assess 5 domains of competence in 321 3rd- and 403 6th-grade children. Peers completed a peer nomination index of multiple competencies. Confirmatory factor analyses revealed 5 factors in both the teacher and parent scales: Academic Competence, Social Acceptance, Athletic Competence, Physical Appearance, and Behavioral Conduct. Two higher order factors also emerged: a Well-Behaved/Good Student factor and an Athletic/Attractive/Popular factor. All 5 subscales of the teacher and parent scales manifested a high degree of discriminant validity. Significant levels of convergent validity emerged for most of these subscales.

Kendall and Morris (1991) argued persuasively that the psychological assessment of children must be a multimethod, if not multiinformant, process. Children often exhibit different strengths and liabilities in different settings, in which they are observed by different people. Parents rarely see their children in the classroom. Teachers rarely see their students in the children's homes. Consequently, setting-specific behaviors can result in low levels of agreement between different informants (Achenbach, McConaughy, & Howell, 1987). Because such cross-informant disagreements can be taken (or mistaken) as evidence of poor convergent validity, researchers and clinicians must be extremely careful in their choice of measurement instruments.

In both research and practice, parents' and teachers' appraisals of children's competencies are important dimensions of a complete assessment. For example, investigators often administer parent and teacher assessments in therapy outcome studies to judge the generalizability of interventions to home and school environments (e.g., Kazdin, 1993; Kendall, Reber, McLeer, & Epps, 1990; Peck, Carlson, & Helmstetter, 1992). Others may attempt to predict children's responses to therapy with use of data obtained from multiple informants (e.g., Kazdin, 1995). Still others may use parent and teacher appraisals to assess the consistency of particular behavior patterns across settings (e.g., Feldman, Salzinger, Rosario, & Alvarado, 1995; French & Waas, 1985; Nassau & Drotar 1995). In diagnostic work, one must often wrestle with discrepant information from multiple informants (e.g., Biederman, Faraone, Milberger, & Doyle, 1993).

In a complete assessment, scientists and practitioners must consider not only deficits and signs of psychopathology in the child, but strengths and signs of competence as well. At different stages of development, some domains of competence may be more important than others. Harter (1990) identified five domains of competence that are of particular importance in middle childhood: academic competence, social acceptance, athletic competence, physical appearance, and behavioral conduct. Harter (1985b) also developed the Self-Perception Profile for Children (SPPC), a widely used and well-validated self-report inventory of children's perceptions of self-competence in these domains. In the interest of multimethodism, Cole and White (1993) introduced and factor analyzed the Peer Nomination of Multiple Competencies to assess peers' perceptions of children's competence in these same five domains. Harter (1985b) also published the Teacher's Rating Scale of Child's Actual Behavior (TRS) and suggested that the teacher form can be reworded for use with parents (Parent's Rating Scale; herein called the PRS) to obtain additional perspectives of children's competence. Although the TRS and the PRS have been used in a variety of studies (e.g., Granleese, Turner, & Trew, 1989; Nottelmann, 1987; Thomas, Forehand, Neighbors, 1995; Wierson, Nousiainen, Forehand, & Thomas, 1992), the factor structures of these instruments have not been established. One goal of the present study was to examine the relation of TRS and PRS items to the factors they were designed to represent.

A second goal was to examine the relation of the factors to one another. Multiple domains of perceived competence are clearly interrelated; however, the nature of these interrelations is less clear. On the one hand, a general competence factor (analogous to a general intelligence factor in the IQ literature) may account for the interrelatedness between competence domains. On the other hand, two or more higher order factors (analogous to verbal and performance in the IQ literature) may better explain the correlations between competence domains. Cole and White's (1993) study of peers' perceptions of competence suggested the existence of two higher order factors. In one factor, two domains of competence (Academic Competence and Behavioral Conduct) converged into a Well-Behaved/Good Student dimension. In the other factor, three domains (Athletic

David A. Cole, Dawn M. Gondoli, and Lachlan G. Peeke, Department of Psychology, University of Notre Dame.

Correspondence concerning this article should be addressed to David A. Cole, Department of Psychology, University of Notre Dame, Notre Dame, Indiana 46556. Electronic mail may be sent to david.a.cole.1@nd.edu.

Competence, Physical Appearance, and Social Acceptance) converged into an Athletic/Attractive/Popular dimension. The emergence of such higher order factors implies that peers tend to perceive certain competencies as co-occurring. Academically competent students tend to be perceived as well behaved. Similarly, athletic and attractive students tend to be perceived as socially competent. Of particular concern to clinicians is whether the perceived co-occurrences of these competencies are veridical or whether they represent spurious stereotypes and illusory correlations to be guarded against in the assessment process (Chapman, 1967). One step toward answering this question is to determine the extent to which other informants perceive the same associations between children's competencies.

Our third goal was to assess the relative validity of various sources of information about children's competence. The ideal context for assessing convergent and discriminant validity is the multitrait–multimethod design (MTMM; Campbell & Fiske, 1959). The optimal statistical approach for analyzing MTMM data is confirmatory factor analysis (CFA; Cole, 1987; Kenny & Kashy, 1992; Widaman, 1985). Discovering that the TRS and PRS subscales load significantly onto the expected underlying trait factors constitutes evidence of convergent validity. Discovering that subscales do not load significantly onto trait factors that they were not designed to assess represents evidence of discriminant validity. Such trait factors are most easily estimated when one has at least three measures (not merely two). Consequently for the present study, we obtained not just parent and teacher reports, but peer nomination estimates of the five competency domains as well.

In the present article, we approached the first goal (i.e., to examine the internal structure of these measures) by applying CFA methods twice, once to the TRS and then to the PRS. (The factor structure of our peer nomination measure was examined in a previous study with an independent sample: Cole & White, 1993.) We use CFA instead of exploratory methods because the items were explicitly constructed to represent one and only one domain of competence.[1] In this manner, CFA allowed us to test how well the anticipated factor structure fit the observed data. To address the second question (about the higher order factor structure), we applied a second level of complexity to the initial CFA models. Instead of allowing the five factors simply to correlate, we tested two higher order models, one with a single higher order factor and one with two such higher order factors. To address the third question (about convergent and discriminant validity of the measures themselves), we applied CFA to the subscales (not the items) of the parent, teacher, and peer assessments, with use of the techniques described by Kenny and Kashy (1992). We conducted all analyses on a relatively large, nonclinic, heterogeneous sample of third- and sixth-grade boys and girls and examined the generalizability of our findings across grade level and gender.

Method

Participants

We collected data from teachers, parents, and peers as part of a larger study. Participants were 724 elementary school students, 49 teachers, and 485 parents. Students attended either third or sixth grade in one of nine public schools in a mid-size, midwestern school district. Each school contained children from urban and suburban neighborhoods. This sample was retained from a larger pool of students, after excluding children in self-contained, special education classrooms, children with such poor reading or attentional skills that they could not complete the questionnaires even with assistance, children who refused to participate, children whose teacher refused to participate, and parents who did not grant informed consent. Classrooms ranged in size from 16 to 28 students (M = 21.1, SD = 6.2). Classrooms did not differ significantly with regard to students' gender or ethnicity ($ps > .40$). The final sample consisted of 321 third graders and 403 sixth graders (48% were girls; 53% were boys). Approximately 67% of parents of participating children completed their questionnaires. The sample was racially heterogeneous, including White (66%), African American (30%), Hispanic (2%), Multiethnic (2%), and Other (1%) children. The mean age was 8.9 years (SD = 0.5) for third graders and 11.9 years (SD = 0.5) for the sixth graders. Family size (i.e., the number of people living in the home) ranged from 3 to 9 (M = 5.5, SD = 1.0). Approximately 37% of the children had at least 1 parent with a previous divorce. Parents reported education levels for mothers ranging from 10 to 20 years (M = 14.0, SD = 2.3), education levels for fathers ranging from 11 to 20 years (M = 14.0, SD = 2.3), and annual family incomes ranging from less than $10,000 to more than $90,000 ($Mdn$ = $35,000). No significant differences emerged between schools on any of the student demographic characteristics ($ps > .50$). Demographic differences between grade levels were also nonsignificant ($ps > .30$).

Measures

Teacher's Rating Scale. Harter's (1985b) TRS is a 15-item report of teachers' appraisals of children's competence. The inventory is of similar form and content to the Self-Perception Profile for Children with the same five competence scales: Academic Competence, Social Acceptance, Athletic Competence, Physical Appearance, and Behavioral Conduct. On the TRS, teachers report how they perceive children's competence, not how they believe children perceive their own competence. Responding to each item is a two-step process. First, teachers select one of two statements about the child. For example, a pair of statements pertaining to academic competence is "This child is really good at his/her school work" and "This child can't do the school work assigned." Next, the teacher indicates whether their choice is "Sort of true" or "Really true" about the child. Items are converted to 4-point rating scales such that high scores reflect greater perceived competence. In previous research (Cole, Martin, Powers, & Truglio, 1996), TRS subscales manifested high levels of internal consistency (Cronbach's alphas ranged from .93 to .97) and good test–retest reliability (correlations ranged from .67 to .73 over a 4-month interval).

Parent Rating Scale. As recommended by Harter (1985b), we reworded the TRS for use by parents about their own children. An example item is "Which is more true about *your child*? *My child* is really good at his/her school work," or "*My child* can't do the school work assigned" (emphasis added). Cole et al. (1996) reported that subscales of the PRS manifested moderately high levels of internal consistency. Cronbach's alphas ranged from .82 to .89 except for Physical Appearance, which was .59. The lower internal consistency of Physical Appearance may have been the result of restriction in range, as few parents rated their children as unattractive. Cole et al. (1996) also reported that the PRS subscales demonstrated relatively strong test–retest reliability estimates, ranging from .60 to .80 over a 4-month interval.

Peer Nominations of Multiple Competencies. To assess the same

[1] Confirmatory factor analyses were conducted with LISREL 8 (Jöreskog & Sörbom, 1993).

five domains of competence from peers' perspectives, we used the Peer Nomination of Multiple Competencies (PNMC; Cole, 1990, 1991a; Cole & White, 1993). The assessment technique is similar to that used in many studies of children's social status (e.g., Coie, Dodge, & Coppotelli, 1982). For each domain of competence, participants are asked two questions, one positive and one negative. For example, the social competence items are "Who has lots of friends in your class?" and "Who does not have many friends at all?" The questions are printed across the top of an optical scan sheet, and classmates' names are printed along the left side. To select classmates for a particular characteristic, students shaded in circles across from the classmate's name and below the appropriate question. Information obtained from each student contributed to the scores of other, not oneself. Students obtained two scores for each domain of competence (i.e., 10 scores in all). Each score represented the proportion of their classmates who nominated them for a particular characteristic (e.g., doesn't have many friends). For each domain, we subtracted the score on the negative item from the score on the positive item, such that higher scores on the composite indicated that the child received more positive than negative nominations. The PNMC subscale has a high degree of stability and is negatively associated with a wide variety of maladaptive behaviors and outcomes (Cole, 1991b; Cole & Carpentieri, 1990). In CFA work, the subscales loaded significantly onto their respective trait factors and only modestly onto a method factor, reflecting strong evidence of convergent and discriminant validity (Cole, 1990).

Procedures

All assessments occurred approximately 8 to 12 weeks into the fall semester of the school year. Research assistants delivered the TRS to teachers at approximately the same time that parents received the PRS by mail. Participants returned completed questionnaires directly to the University of Notre Dame in self-addressed, stamped envelopes. Research assistants made phone calls and sent postcards as reminders whenever necessary. Substitute teachers did not participate in the study; regular teachers completed the surveys after their absence. Teachers received $30 each for their participation. Only 1 parent participated from each household. Nonresponding parents were prompted by postcard and phone. Parents were given the option either to receive $10 for their participation or to have $10 donated to their children's school for the purchase of educational materials.

Results

Preliminary Results

Before addressing the primary goals of this study, we conducted several preliminary analyses. With only 67% of parents responding to the survey, we were concerned about potential self-selection bias. In a series of tests, we compared responders with nonresponders on variables on which we did have complete data. Chi-square tests revealed no significant differences between responders and nonresponders with regard to child's grade level, gender, or race ($ps > .20$). Separate multivariate analyses of variance (MANOVAs) for the TRS and the PNMC revealed no significant differences between children of responding versus nonresponding parents.

To describe the sample more completely, we present the means and standard deviations for all subscales of the TRS and PRS, broken down by grade level and gender (see Table 1). To test the significance of grade and gender effects, we conducted two separate 2 × 2 (Grade × Gender) MANOVAs, one for the five TRS subscales and one for the five PRS subscales. For the parent subscales, MANOVA revealed only a significant gender main effect, $F(5, 477) = 12.63$, $p < .001$. Univariate tests (using Bonferroni's correction for Type I error: $\alpha/5 = .01$) revealed that parents rated boys as more athletically competent than girls and girls as better behaved than boys. For the teacher subscales, we found significant multivariate effects for grade, $F(5, 716) = 6.14$, $p < .001$; for gender, $F(5, 716) = 12.98$, $p < .001$; and for the interaction, $F(5, 716) = 2.82$, $p < .02$. Univariate tests revealed that teachers gave third graders higher ratings than sixth graders with regard to Social Competence, Athletic Competence, and Physical Appearance. Teachers also rated boys higher than girls on Athletic Competence but rated girls higher than boys on Behavioral Conduct. The gender differ-

Table 1
Means and Standard Deviations of Teacher's Rating Scale of Child's Actual Behavior of (TRS) and Parent's Rating Scale (PRS) Subscales Broken Down by Grade and Gender

	Third-grade girls		Third-grade boys		Sixth-grade girls		Sixth-grade boys		
Subscale	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	Univariate effect ($p < .01$)
					PRS				
Academic	6.82	2.79	6.43	2.91	7.10	2.24	6.44	2.60	
Social	7.17	2.41	7.20	2.27	6.96	2.67	7.02	2.53	
Athletic	4.86	2.45	6.17	2.18	4.79	2.55	6.01	2.44	Gender
Appearance	8.60	0.90	8.43	1.06	8.46	1.09	8.25	1.23	
Behavior	7.44	2.13	6.47	2.58	7.50	2.13	6.85	2.40	Gender
					TRS				
Academic	5.90	3.03	5.39	3.20	5.97	3.07	5.33	3.22	
Social	6.39	2.54	6.20	2.57	5.72	2.70	5.47	2.79	Grade
Athletic	5.46	2.09	6.21	2.24	5.17	2.48	5.54	2.53	Gender, grade
Appearance	7.67	1.89	7.67	1.73	6.95	2.30	6.98	2.10	Grade
Behavior	7.70	2.67	6.21	3.08	7.69	2.10	5.54	3.34	Gender, Gender × Grade

ence on Behavioral Conduct was stronger in third than in sixth grade.

Structure of the Items

To examine the structure of the TRS and PRS, we conducted a series of multigroup CFAs. We examined the TRS items and the PRS items separately. We included all 724 child participants in our analyses of the TRS and 485 of the participants whose parents responded in our analyses of the PRS. (Analysis of the TRS with only the 485 participants with complete data yielded essentially the same results.) In each CFA, we allowed each item to load onto only one of five oblique factors, according to the domain of competence it was designed to represent. All other factor loadings were constrained to be zero. None of the disturbances was allowed to correlate with the others. In each analysis, the input data consisted of four variance–covariance matrices, one for each of the gender–grade groups. We compared three hierarchically nested models, not only to test the integrity of this model but also to test the generalizability of this model across groups.

We used several criteria to test each model. We first examined the chi-square. A nonsignificant chi-square indicates that discrepancies between the model and the data are negligible. Unfortunately, with large samples such as those in the present study, very small discrepancies between the model and the data can lead to a significant chi-square. Other criteria must be used to assess whether the magnitude of the discrepancies is large enough to be of substantive concern. Among such indices are the goodness-of-fit index (GFI), the comparative fit index (CFI), and the root mean squared error of approximation (RMSEA). A large GFI and CFI (e.g., greater than .90) indicate that the model has accounted for a large proportion of the covariance between the measures. A small RMSEA (e.g., less than .05) indicates that only a very small amount of information was not explained by the model. In the present article, our large sample sizes consistently produced statistically significant chi-squares. Therefore, we relied heavily on the alternative fit indices in the interpretation of our results.

In Model 1, we placed no cross-group equality constraints on any parameters. This model provided an excellent fit to the data for both the TRS and the PRS. As shown in Table 2, all fit indices (except the chi-square) confirmed the good fit. In Model 2, we constrained the factor loadings to be equal across groups. This model also provided an excellent fit to the data for both the TRS and PRS. Finally in Model 3, we constrained the factor intercorrelations (as well as the factor loadings) to be equal across groups. As shown in Table 2, this model also provided a good fit to the data on both measures. In other words, the structure of both the TRS and the PRS appeared to be equivalent across gender and across grade level. (See Model 4 below for a discussion of particular parameter estimates.)

Higher Order Factor Structure

To test our hypothesis about the higher order factor structure of the TRS and PRS, we constructed a fourth CFA model. This model was identical to Model 3 (above) except that instead of allowing all five factors to correlate freely, we constrained them to correlate by virtue of two higher order factors. We allowed the Behavioral Competence factor and the Academic Competence factor to load onto a higher order Well-Behaved/Good Student factor, and we allowed the Social Acceptance, Athletic Competence, and Physical Appearance factors to load onto a higher

Table 2
Goodness-of-Fit Indexes for Nested Confirmatory Factor Analysis (CFA) Models of the Teacher Rating Scale (TRS) and Parent Rating Scale (PRS)

Model	df	χ^2	GFI	CFI	RMSEA	p
TRS (n = 724)						
1. Five first-order factors, no cross-group constraints	320	518.56	.93	.98	.029	>.99
2. Five first-order factors, factor loadings equal across groups	350	591.73	.92	.98	.031	>.99
3. Five first-order factors, factor loadings and correlations equal across groups	395	684.58	.91	.97	.033	>.99
4. Five first-order factors and two higher order factors, equal across groups	399	689.78	.91	.97	.034	>.99
PRS (n = 485)						
1. Five first-order factors, no cross-group constraints	320	487.17	.92	.95	.033	>.99
2. Five first-order factors, factor loadings equal across groups	350	536.30	.91	.94	.033	>.99
3. Five first-order factors, factor loadings and correlations equal across groups	395	613.41	.90	.93	.034	>.99
4. Five first-order factors and two higher order factors, equal across groups	399	617.56	.90	.93	.034	>.99

Note. GFI = goodness-of-fit index; CFI = comparative fit index; RMSEA = root mean squared error of approximation.

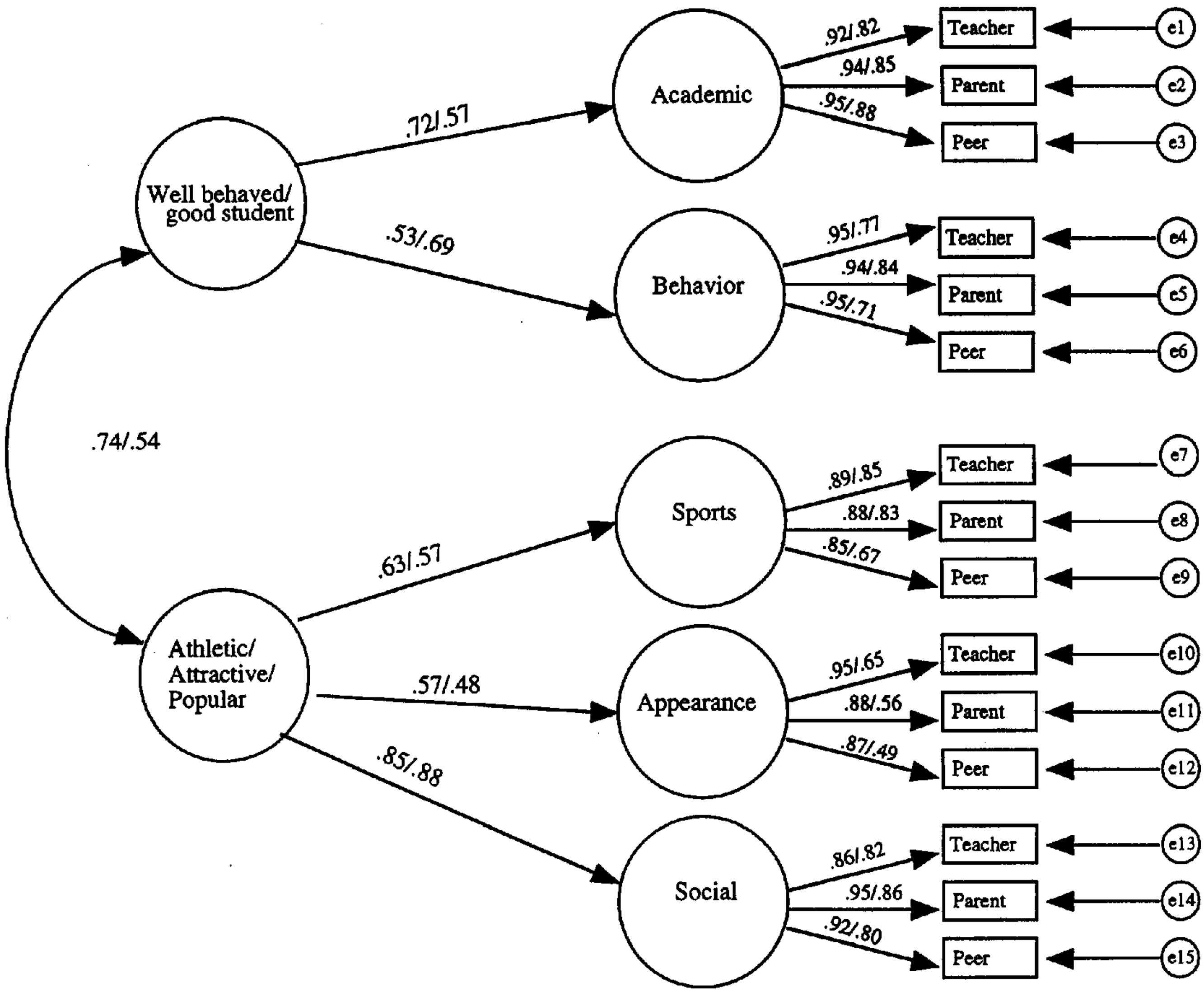

Figure 1. Path diagram of two-factor confirmatory factor analytic models with parameter estimates for the Teacher Rating Scale (left) and the Parent Rating Scale (right). All parameters were constrained to be equal across groups. Error terms associated with manifest variables are represented by e1–e15.

order Athletic/Attractive/Popular factor. We also constrained all higher and lower order factors (and the correlation between the higher order factors) to be equivalent across all four groups. As above, we applied this model to the TRS and the PRS data separately.

Model 4 provided a good fit to both the TRS and the PRS data by all criteria (except the chi-square). This model and its parameter estimates are presented in Figure 1. For the TRS, first-order factor loadings ranged from .85 to .95. All were significantly greater than zero ($ps < .001$). Second-order factor loadings for the TRS ranged from .53 to .85 ($ps < .001$). The correlation between the two higher order factors was .74. This correlation was significantly less than unity, indicating that combining them into a single higher order factor would not be appropriate. (For this reason, we did not test a more restrictive model with only one [general] higher order factor. The present results demonstrate the inadequacy of such a model.) For the PRS, the first-order factor loadings were slightly smaller than for the TRS, ranging from .49 to .88. All were significantly greater than zero. The smallest loadings appeared on the Physical Appearance factor, which was probably a consequence of the range restriction of parents' responses to these items. Few parents rated their children as unattractive. Higher order factor loadings on the PRS ranged from .48 to .88 ($ps < .001$). The correlation between the two higher order factors was .54; again this value was significantly less than unity, reflecting the distinction between these factors. In summary, Model 4 fit both the TRS and the PRS data well, indeed equally well in all four groups.

Convergent and Discriminant Validity

To assess the convergent and discriminant validity of these measures, we constructed a multitrait–multimethod covariance matrix for each of the four groups. Each 15 × 15 matrix contained the variances and covariances for three methods (teacher ratings, parent ratings, and peer nominations) of measuring the five domains of competency. The teacher ratings were the five TRS subscales; the parent ratings were the five PRS subscales; and the peer nominations were the five subscales of the PNMC. We estimated the convergent and discriminant validity of the teacher and parent scales by means of a multigroup–multitrait–multimethod CFA. We allowed each measure to load onto only one of five oblique trait factors. All other factor loadings were constrained to equal zero. These five trait factors were oblique. We allowed for the anticipated shared method variance by use of Kenny and Kashy's (1992) correlated disturbance approach. That is, we allowed correlations between the disturbance terms of measures that used the same method. In particular, the disturbance terms for the five TRS scales were allowed to correlate; the disturbances for the five PRS scales were allowed to correlate; and the disturbances for the five peer nomination scales were allowed to correlate. A well-fitting model without cross-loadings would constitute evidence of discriminant validity. The emergence of significant trait factor loadings would constitute evidence of convergent validity. Anticipating that teachers might be more objective than parents in certain domains, we compared the trait factor loading for each TRS scale with its PRS counterpart. Assuming that the validity of the TRS and PRS scales might differ from gender to gender or from third to sixth grade, we also compared trait factor loadings across groups.

First, we tested a base model in which we placed no within- or between-groups constraints on any of the factor loadings. (We did, however, constrain factor variances and covariances to be equal across groups.) This model provided a good fit to the data by all criteria except the chi-square: $\chi^2(230, N = 495) = 382.52$, GFI = 91, CFI = 95, RMSEA = .054. The fact that this multigroup model fit the data model without allowing scales to cross-load constitutes evidence of discriminant validity in all four groups. The factor loadings for the PRS and TRS varied from subscale to subscale (see Table 3).

Then we conducted two series of follow-up tests. In one series, we compared the TRS factor loadings with their counterparts on the PRS within each group. These tests enabled us to identify subscales that had higher (or lower) factor loadings depending on whether they were administered to teachers or parents. In the second series, we compared the TRS and PRS factor loadings across groups. These tests enabled us to identify subscales with higher (or lower) factor loadings in specific subgroups.

Academic Competence. For Academic Competence, factor loadings ranged from .68 to .87 for the TRS and from .65 to .86 for the PRS, reflecting a high degree of convergent validity in all four groups. We compared the TRS factor loadings with their counterparts on the PRS within each group. None of the loadings was significantly different from the others. We also compared each of the TRS and PRS factor loadings across groups. These comparisons were also nonsignificant.

Social Acceptance. For Social Acceptance, factor loadings ranged from .56 to .85 for the TRS and from .35 to .68 for the PRS. All factor loadings were significantly greater than zero, and all differences between groups were nonsignificant, suggesting good convergent validity in all groups. For sixth-grade boys, however, the TRS loading was significantly greater than the PRS

Table 3
Trait Factor Loadings From Multigroup–Multitrait–Multimethod Confirmatory Factor Analysis

Measure	Third grade: Girls	Third grade: Boys	Sixth grade: Girls	Sixth grade: Boys	Between-subjects pairwise comparison
Academic Competence					
Teacher	.83	.87	.67	.83	
Parent	.80	.85	.65	.72	
Peer	.61	.56	.69	.93	
Social Acceptance					
Teacher	.56	.77	.73	$.85_a$	
Parent	.36	.56	.69	$.49_a$	
Peer	.48	.38	.93	.73	
Athletic Competence					
Teacher	.47	.68	.69	.81	
Parent	.49	.42	.65	.53	
Peer	.45	.44	.88	.75	
Physical Appearance					
Teacher	$.59_b$	$.31_c$	$.53_d$	$.45_e$	
Parent	$.18_b$	$.09_c$	$.29_d$	$.22_e$	
Peer	.70	.39	.98	.39	
Behavioral Conduct					
Teacher	$.58_f$	$.97_g$	.38	$.98_h$	b3 > g3; b6 > g6
Parent	$.36_f$	$.59_g$	.47	$.74_h$	b3 > g3; b6 > g6
Peer	.65	.69	.39	.79	

Note. In a given column, loadings with the same subscripts represent pairwise comparisons (teacher vs. parent) that were significant ($\alpha = .05$).

loading, suggesting that teachers may be better informants about social acceptance than are parents in this group.

Athletic Competence. For Athletic Competence, the TRS factor loadings ranged from .47 to .81, and the PRS loadings ranged from .42 to .65. Differences across groups or between measures were all nonsignificant. All loadings were significantly greater than zero, reflecting good convergent validity.

Physical Appearance. For Physical Appearance, factor loadings for the TRS ranged from .30 to .59. Although all loadings were significantly greater than zero, their overall magnitudes were not large, reflecting only moderate levels of convergent validity. For the PRS, loadings ranged from .09 to .29. For third-grade boys and girls, the loadings were not significantly greater than zero, reflecting serious problems with convergent validity. In all four groups, the PRS loadings were significantly smaller than their TRS counterparts, suggesting that parents were less valid reporters of physical appearance than were teachers. As mentioned above, these problems appear to be due to a restriction of range in parents' ratings of their children's attractiveness.

Behavioral Conduct. For Behavioral Conduct, TRS loadings ranged from .38 to .98. All loadings were significantly greater than zero; however, the loadings for boys were significantly larger than those for girls at both grade levels. For boys, the TRS behavior scale showed strong evidence of convergent validity. For girls, the TRS behavior scale had only moderate levels of convergent validity. For the PRS, factor loadings ranged from .39 to .73. Again the loadings for boys were significantly larger than those for girls, reflecting good convergent validity for boys and moderate validity for girls. In three of the four groups, the loadings for the PRS were significantly lower than those for the TRS, suggesting that teachers may be better informants about behavioral conduct than are parents.

Discussion

Four major findings emerged from the present study. First, CFAs of the items in the parent and teacher forms of Harter's (1985b) Rating Scales of Children's Actual Behavior revealed highly interpretable and generalizable factor structures that corresponded closely to the theoretical structure that governed the construction of the instruments. Second, we confirmed that the five subscales from both instruments reflected two higher order dimensions: a Well-Behaved/Good Student factor and an Athletic/Attractive/Popular factor. Third, all TRS and PRS subscales showed evidence of discriminant validity in all four groups. Fourth, all of the teacher subscales manifested at least moderately high levels of convergent validity, as did all but one of the parent subscales. Finally, teachers' and parents' appraisals of boys' behavioral competence were more valid than were their appraisals of girls' behavioral competence. We discuss each of these findings below.

Interpretable and consistent factor structures emerged for both the parent and teacher measures of perceived competence. For third- as well as sixth-grade children and for boys as well as girls, all items loaded significantly onto the factor that clearly represented the dimension they were designed to assess. Indeed, the structures of the teacher and parent scales were essentially identical to those previously reported for Harter's (1985b) original measure of children's self-reported competence. At a practical level, these findings justify the interpretation of the TRS and PRS subscales, in a manner comparable to that recommended for the SPPC. At a more theoretical level, Harter's decomposition of children's self-perceived competence into five component parts lent credence to Wylie's (1974) and Harter's (1985a) positions that self-esteem is not a single, unitary, global construct. In a similar fashion, the present study suggests that other's perceptions of children's competencies can also be construed as consisting of at least five distinct components that appear parallel to children's self-perceptions.

Of additional importance is that this five-factor structure generalized across grade level (third and sixth) and across gender. Formal comparisons revealed that even the magnitudes of the factor loadings were equivalent across groups. These results have important implications for the use of the TRS and PRS to study developmental and gender-related differences. An assumption underlying such research is that the measures are structurally equivalent across groups (or across time). These measures appeared to tap into the same constructs in all four of the groups we examined. These results suggest that the TRS and PRS hold considerable promise for gender studies and developmental research into children's competence.

We also found that the five subscales within each instrument were correlated in a manner that reflected two higher order dimensions: a Well-Behaved/Good Student factor and an Athletic/Attractive/Popular factor. This result does not imply that the theoretical five-factor model is wrong or should be abandoned in favor of this simpler structure. It does mean, however, that the five factors correlate with one another for two more general reasons that deserve further scrutiny. The emergence of a Well-Behaved/Good Student factor suggests that parents and teachers who perceive a child as academically competent will also tend to view that child as well behaved. Similarly, the emergence of an Athletic/Attractive/Popular factor suggests that children who are perceived as competent in sports or as physically attractive will tend to be perceived as popular as well. Perhaps more important from a clinical perspective is that children who are perceived as incompetent in one domain may be regarded as incompetent in the associated domains as well. These associations may reflect perceptual stereotypes in the eyes of beholders, or they may represent true associations between naturally occurring personal characteristics. That similar findings emerged from peer nomination data (Cole & White, 1993) suggests that these perceptions are at least highly generalizable. In future studies, appraisals by teachers, parents, and peers could be compared with more objective measures of competence to disentangle perceptual stereotypes from actual patterns of competence.

The magnitude of the correlation between the Well-Behaved/Good Student factor and the Athletic/Attractive/Popular factor differed somewhat from the TRS (r = .74) to the PRS (r = .54). The larger correlation between these two factors on the TRS reflects greater conflation of these two dimensions in the eyes of teachers. Such conflation may be a reflection of global influences such as halo effects. Teachers may have more of a tendency than parents to perceive good students as good in multiple domains (and poor students as relatively incompetent in multiple domains). In part, such perceptions may be due to the relatively limited number of settings in which teachers actually observe children. Parents, however, observe their children in a wider variety of contexts (e.g., sports) and receive feedback about their children in still other domains (e.g., school). Such diversity of information may reduce overgeneralizations about children's competencies. In clinical evaluations, a major task is often to identify children's strengths as well as weaknesses. In this assessment process, clinicians may need to corroborate globally positive (or negative) appraisals from teachers by seeking confirmation from an independent source.

Confirming the internal structure of an instrument, however, implies nothing about the validity of the instrument vis à vis other measures or external criteria. Through the power of a multitrait–multimethod design, we accumulated evidence of discriminant and convergent validity of the TRS and PRS. Both measures showed a high degree of discriminant validity in all four of our groups. In other words, no TRS or PRS subscale loaded onto any factor other than the one it was designed to represent. We also found evidence of convergent validity for all five of the TRS subscales and for four out of five of the PRS subscales. The only weak subscale was parents' report of their children's physical appearance, which failed to load significantly onto the Physical Appearance factor in all four of our groups. Examination of the means and standard deviations of this scale reveals evidence of ceiling effects. Not surprisingly, most parents regarded their children as rather attractive. Teachers and peers were more discriminating.

Teachers' and parents' appraisals of boys' behavioral conduct had higher levels of convergent validity than did teachers' and parents' appraisals of girls' behavioral conduct. This finding is especially interesting given the often cited mean difference between boys' and girls' levels of misbehavior, conduct, and overt aggression (Loeber & Keenan, 1994). Gender differences in behavioral conduct also emerged in the present study. In general, boys appear to misbehave more than girls. Rock, Werts, and Flaugher (1978) demonstrated that manifest mean differences can emerge between groups that have no mean difference on the underlying true scores when the measures are not equally valid in all groups under investigation. Cole, Maxwell, Arvey, and Salas (1993) underscored this point in their discussion of alternatives to traditional MANOVA techniques.

These possibilities are especially interesting given recent evidence that girls and boys may misbehave at similar rates but in different ways. Crick and Grotpeter (1995) suggested that boys exhibit higher rates of over and physical aggression, but that girls exhibit higher rates of relational aggression (i.e., the tendency to aggress toward others through social relationships). When overt and relational aggression are considered jointly, the gender difference in aggressive behavior appears to diminish or disappear (Crick, Bigbee, & Howers, 1996; Crick & Grotpeter, 1995). These findings in conjunction with our own lead us to speculate that traditional measures of behavioral conduct may underrepresent the kinds of misbehavior often exhibited by girls, hence attenuating the validity of such measures in young female populations. Further refinement and testing of new behavioral conduct measures may be needed.

Several caveats regarding the present study suggest avenues for future research. One limitation of the study may reflect a limitation of the instruments themselves. The TRS and PRS were designed for use in middle childhood. Consequently, we included only third and sixth graders in our investigation. Although other forms of Harter's measures exist for use at earlier and later grade levels (e.g., Harter, 1988), the present versions may well be valid for use beyond the third- to sixth-grade range. Indeed for longitudinal studies of the transition from middle childhood into adolescence, having instruments that remain valid and appropriate for use in middle and high school populations becomes of paramount importance. Further validation studies of these measures in older samples would be a valuable next step. A second limitation is the absence of objective or behavioral measures of competence. Although the appraisals of parents, teachers, and peers represent three important and diverse sources of information, the comparison of these measures to a set of behavioral criteria would address additional questions of interest. In the present study, we observed that parents and teachers tended to associate academic competence with good behavior and popularity with physical attractiveness and athletic competence. Although previous research suggests that peers do the same (Cole & White, 1993), we do not know if these are truly correlated characteristics or the correlations exist in the eyes of the perceiver. In future studies, the addition of objective measures of competence would help to resolve this question.

References

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232.

Biederman, J., Faraone, S. V., Milberger, S., & Doyle, A. (1993). Diagnoses of attention-deficit hyperactivity disorder from parent reports predict diagnoses based on teacher reports. *Journal of the American Academy of Child and Adolescent Psychiatry, 32,* 315–317.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior, 6,* 151–155.

Coie, J. D., Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status in the school: A cross-age comparison. *Child Development, 59,* 815–829.

Cole, D. A. (1987). The utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology, 55,* 584–594.

Cole, D. A. (1990). Relation of social and academic competence to depressive symptoms in childhood. *Journal of Abnormal Psychology, 99,* 422–429.

Cole, D. A. (1991a). Change in self-perceived competence as a function of peer teacher evaluation. *Developmental Psychology, 27,* 682–688.

Cole, D. A. (1991b). Preliminary support for a competency-based model of depression in children. *Journal of Abnormal Psychology, 100,* 181–190.

Cole, D. A., & Carpentieri, S. (1990). Social status and the comorbidity of child depression and conduct disorder. *Journal of Consulting and Clinical Psychology, 58,* 748–757.

Cole, D. A., Martin, J. M., Powers, B., & Truglio, R. (1996). Modeling causal relations between academic and social competence and depression: A multitrait–multimethod longitudinal study of children. *Journal of Abnormal Psychology, 105,* 258–270.

Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin, 114,* 174–184.

Cole, D. A., & White, K. (1993). Structure of peer impressions of children's competence: Validation of the peer nomination of multiple competencies. *Psychological Assessment, 5,* 449–458.

Crick, N. R., Bigbee, M. A., & Howers, C. (1996). Gender differences in children's normative beliefs about aggression: How do I hurt thee? Let me count the ways. *Child Development, 67,* 1003–1014.

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development, 66,* 710–722.

Feldman, R. S., Salzinger, S., Rosario, M., & Alvarado, L. (1995). Parent, teacher, and peer ratings of physically abused and nonmaltreated children's behavior. *Journal of Abnormal Child Psychology, 23,* 317–334.

French, D. C., & Waas, G. A. (1985). Behavior problems of peer-neglected and peer-rejected elementary-age children: Parent and teacher perspectives. *Child Development, 56,* 246–252.

Granleese, J., Turner, I., & Trew, K. (1989). Teachers' and boys' and girls' perceptions of competence in the primary school: The importance of physical competence. *British Journal of Educational Psychology, 59,* 31–37.

Harter, S. (1985a). Competence as a dimension of self-evaluation: Toward a comprehensive model of self-worth. In R. Leahy (Ed.), *The development of the self* (pp. 55–121). New York: Academic Press.

Harter, S. (1985b). *Manual for the Self-Perception Profile for Children.* Denver, CO: University of Denver.

Harter, S. (1988). *Manual for the Self-Perception Profile for Adolescents.* Denver, CO: University of Denver.

Harter, S. (1990). Causes, correlates, and the functional role of global self-worth: A life-span perspective. In R. J. Sternberg & J. Kolligan, Jr. (Eds.), *Competence considered* (pp. 67–97). New Haven, CT: Yale University Press.

Jöreskog, K., & Sörbom, D. (1993). *LISREL: User's reference guide.* Chicago: Scientific Software International.

Kazdin, A. E. (1993). Changes in behavioral problems and prosocial functioning in child treatment. *Journal of Child & Family Studies, 2,* 5–22.

Kazdin, A. E. (1995). Child, parent, and family dysfunction as predictors of outcome in cognitive–behavioral treatment of antisocial children. *Behaviour Research & Therapy, 33,* 271–281.

Kendall, P. C., & Morris, R. J. (1991). Child therapy: Issues and recommendations. *Journal of Consulting and Clinical Psychology, 59,* 777–784.

Kendall, P. C., Reber, M., McLeer, S., & Epps, J. (1990). Cognitive–behavioral treatment of conduct-disordered children. *Cognitive Therapy & Research, 14,* 279–297.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait–multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112,* 165–172.

Loeber, R., & Keenan, K. (1994). Interaction between conduct disorder and its comorbid conditions: Effects of age and gender. *Clinical Psychology Review, 14,* 497–523.

Nassau, J. H., & Drotar, D. (1995). Social competence in children with IDDM and asthma: Child, teacher, and parent reports of children's social adjustment, social performance, and social skills. *Journal of Pediatric Psychology, 20,* 187–204.

Nottelmann, E. D. (1987). Competence and self-esteem during transition from childhood to adolescence. *Developmental Psychology, 23,* 441–450.

Peck, C. A., Carlson, P., & Helmstetter, E. (1992). Parent and teacher perceptions of outcomes for typically developing children enrolled in integrated early childhood programs: A statewide study. *Journal of Early Intervention, 16,* 53–63.

Rock, D. A., Werts, C., & Flaugher, R. L. (1978). The use of analysis of covariance structures for comparing the psychometric properties of the multiple variables across populations. *Multivariate Behavioral Research, 13,* 403–418.

Thomas, A. M., Forehand, R., & Neighbors, B. (1995). Change in maternal depressive mood: Unique contributions to adolescent functioning over time. *Adolescence, 30,* 43–52.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for the multitrait–multimethod data. *Applied Psychological Measurement, 9,* 11–26.

Wierson, M., Nousiainen, S. D., Forehand, R., & Thomas, A. M. (1992). Parent–adolescent congruence on decision-making and its relation to adolescent adjustment. *Journal of Family Violence, 7,* 9–18.

Wylie, R. C. (1974). *The self-concept* (Vol. 1, Rev. ed.). Lincoln: University of Nebraska Press.

Received July 29, 1997
Revision received February 24, 1998
Accepted April 6, 1998 ■