Gifted Child Quarterly
Volume 52 Number 2
Spring 2008 160-169

10.1177/0016986208315802
http://gcq.sagepub.com
hosted at
http://online.sagepub.com

Validation of the Gifted Rating Scales–School Form in China

Huijun Li
Harvard University
Steven I. Pfeiffer
Yaacov Petscher
Florida State University
Alper T. Kumtepe
Anadolu University, Turkey
Guofang Mo
Yunnan Teachers University

Abstract: The Gifted Rating Scales–School Form (GRS-S), a teacher-completed rating scale, is designed to identify five types of giftedness and motivation. This study examines the reliability and validity of a Chinese-translated version of the GRS-S with a sample of Chinese elementary and middle school students ($N = 499$). The Chinese GRS-S was found to have high internal consistency. Results of the confirmatory factor analysis corroborated the six-factor solution of the original GRS-S. Comparison of the GRS-S scores and measures of academic performance provide preliminary support for the criterion validity of the Chinese-translated GRS-S. Significant age and gender differences on the Chinese GRS-S were found. Results provide preliminary support for the Chinese version of the GRS-S as a reliable and valid measure of giftedness for Chinese students.

Putting the Research to Use: This study indicates that the Chinese-translated GRS-S holds promise to assist in the screening of Chinese students for gifted programs. Participating Chinese teachers were each willing to complete 10 GRS-S forms on students in their class, suggesting that the GRS-S could readily be adopted for wide-scale screening of students in China. This is an important finding because China has with about 200 million school-age children (versus 40 million in the United States) and is interested in screening and identifying its most gifted and talented students. Most authorities estimate that 5 to 10% of children are gifted. This means that a gifted screening tool such as the GRS-S could help identify 10 to 20 million gifted children in China—equal to half the school population in the United States. Individualized assessment of Chinese students for giftedness is presently impractical for financial, logistical, manpower shortage, and other practical reasons. An instrument that is brief, inexpensive, and easy to administer and score affords China an exciting opportunity to reach millions of its gifted students. The GRS-S or a similar translated teacher rating scale provides China with an educational tool to further its commitment to gifted education.

Keywords: *gifted rating scales; Chinese children; reliability; validity*

Hundreds of thousands of school-age children in the United States and worldwide are considered annually for programs for the gifted and talented (Pfeiffer & Jarosewich, 2003). Exceptional general intelligence was once thought to be the hallmark of giftedness but now is regarded as only one manifestation of giftedness. Recent theories emphasize the multidimensionality of outstanding abilities (e.g., Gardner, 1983; Sternberg, 1997; Tannenbaum, 2000). Although most people think only of intellectual giftedness, children display artistic, musical, athletic, dramatic, interpersonal, aesthetic, leadership, creative, and other gifts. Screening and identifying different types of gifted students have been regarded as challenging and important tasks to ensure that gifted

students receive the special resources and services that they need to reach their full potential (Gallagher, 2003; Pfeiffer, 2002). One problem with the current identification process is that most instruments are designed to measure only intellectual factors. There are few reliable screening tools available to complement intelligence tests in providing assessment of multiple dimensions of giftedness (Jarosewich, Pfeiffer, & Morris, 2002). Several standardized achievement tests, such as Iowa Tests of Basic Skills (Hoover, Hieronymus, Dunbar, & Frisbie, 2001) and California Basic Educational Skills Test (National Evaluation Systems, 2005), have been used in the identification process. One disadvantage in using standardized achievement tests is that they are not sufficiently difficult to measure students' high ability, knowledge, and skill levels (Davis & Rimm, 2004). Another problem regarding the current identification process is that some instruments do not include representative norms, although they may consist of multiple types of giftedness (Richert, Alvino, & McDonnel, 1982).

A number of teacher rating scales have been developed to assess intellectual and academic giftedness. However, many of the rating scales suffer from limitations that compromise their diagnostic value (Jarosewich et al., 2002). For example, some popular scales consist of nonrepresentative standardization normative samples with low interrater reliability and lack of diagnostic precision (Jarosewich et al., 2002; Siegel & Powell, 2004). Some rating scales only measure specific subject areas such as math, language, foreign language, social studies, business, and shop mechanical skills (Feldhusen & Jarwan, 2000). Some researchers have even voiced concerns about the validity and reliability of teacher nomination and ratings of gifted children (Hoge & Coladarci, 1989). Teachers' judgments on student performance are thought to be subjectively contaminated by their expectations and biases (Hoge, 1983; Hoge & Butcher, 1984; Sharpley & Edgar, 1986; Silverstein, Brownlee, Legutki, & MacMillan, 1983).

Most authorities, however, argue that teachers have sufficient amounts of time to reliably observe and interact with students on a daily basis and that they are capable of making valid student assessments (Kenny & Chekaluk, 1993; Meisels, Bickel, Nicholson, Xue, & Atkins-Burnett, 2001; Perry & Meisels, 1996). Thus, teacher ratings and nomination are still considered very important data in the screening and identification of gifted students (Davidson, 1986; Pfeiffer, 2002).

China has a long history of seeking out child prodigies and providing special opportunities for its most gifted students. However, like the United States and many other countries, one pressing issue facing the gifted field in China is a lack of sound instruments to screen and identify gifted children, particularly instruments that would provide information on the multidimensional aspects of giftedness (Chan, 2004; Shi & Zha, 2000; Stevenson, 1998). Typically, students in China first apply for the gifted examination and then take a Chinese version of the Stanford-Binet Intelligence Test or the Wechsler Preschool or Primary Scale of Intelligence–Revised. Students who pass the Intelligence Quotient threshold (≥130) are administered the Test for Identifying the Cognitive Ability of Supernormal Children (Cooperative Research Group of Supernormal Children of China, 1985). Additional information regarding a student's personality traits and physical status are collected for students who obtain a score on the ICASC two or more standard deviations above the mean, or a score above the mean for children 2 years older, or a score greater than the 95th percentile for children of the same age. Students who exceed the stated criteria are placed in a special class for several weeks to examine their learning processes. If the student does well in meeting the intellectual challenges of the rigorous classroom setting, he or she is permitted to remain in the gifted classroom (Shi & Zha, 2000). Thus, although it is a multistep admission process, the major focus is intellectual and academic performance. In Hong Kong, Chan (2003) has developed the Student Multiple Intelligence Profile, a 24-item self-report checklist to measure multiple dimensions of giftedness such as verbal–linguistic, musical, logical–mathematical, visual–spatial, bodily–kinesthetic, intrapersonal, and interpersonal intelligences. This is the first instrument in Chinese to measure multiple dimensions of giftedness. However, it does not include scales for creativity or leadership, considered by many as types of giftedness. In addition, the limited number of items (3 items) for each scale compromises its breadth of coverage.

One new screening instrument that holds promise to help identify gifted students is the Gifted Rating Scales–School Form (GRS-S; Pfeiffer & Jarosewich, 2003). The GRS-S is a teacher-completed rating scale designed to help screen giftedness with students in

Authors' Note: We thank Ms. Song, Chukun of Affiliated Elementary School of Yunnan Teachers University, and Ms. Li, Juan of Number 24 Middle School, Kunming, Yunnan Province, for their support in the data collection. We convey our appreciation to PsychCorp and Harcourt Assessment for permitting us to translate the Gifted Rating Scales–School Form into Chinese. We thank Ms. Samara Blei for her effort conducting the literature search.
Note: This article accepted under the editorship of Paula Olszewski-Kubilius.

grades one through eight, ages 6:0 through 13:11. The GRS-S has four noteworthy features. One, the GRS-S is based on a multidimensional model of giftedness consisting of five ability factors—intellectual ability, academic ability, creativity, artistic talent, and leadership ability. Two, the GRS-S includes a sixth scale that measures motivation. The motivation scale is not considered a type of giftedness but rather an index of how hard the student works as a means of helping to explain the student's academic performance (Pfeiffer & Jarosewich, 2003). Three, the GRS-S can be used as a first-stage screening instrument to screen an entire class, school, or even school system. It can also be used as a second-stage screening instrument with one or a target subgroup of students who have already been flagged using another screening procedure (e.g., noteworthy performance on an achievement test, group nonverbal measure, or teacher nomination). Fourth, the GRS-S has demonstrated strong psychometric properties (Margulies & Floyd, 2004; Ward, 2005).

Although the GRS-S has many advantages as an assessment tool, it has not yet been translated and validated in other countries, such as China, where there is an urgent need for instruments to measure multiple constructs of giftedness. It was of interest to examine whether a translated version of the GRS-S is low in cultural loading and useful with a Chinese population. Gender differences have been found using GRS-S in American and Korean samples where girls were rated somewhat higher on the scales of the GRS-S (Lee & Pfeiffer, 2006; Pfeiffer, Petscher, & Jarosewich, 2007). Age differences were also examined but no significant differences in either the American or Korean samples were found. It is, therefore, of interest to examine if a Chinese version of the GRS-S would find gender or age differences.

This article reports on the translation and validation of the GRS-S into Chinese. The study was guided by five research foci. First, what is the reliability of the Chinese-translated version of the GRS-S? Second, what is the factor structure of the Chinese version of the GRS-S? Third, are there gender or age differences on the Chinese version of the GRS-S? Fourth, what is the criterion-related validity of the Chinese version of the GRS-S? Fifth, what is the influence of how long or how well a teacher has known a student?

Method

Translation of the GRS-S

With authorization of the Psychological Corporation, Harcourt Assessment Company, we translated the rating scales into Chinese following a carefully articulated and prescribed translation procedure (Bracken & Barona, 1991; Butcher, 1996, Geisinger, 1994). First, the first author translated the GRS-S from English into Chinese. Second, a five-member panel consisting of bilingual educators reviewed the Chinese version of the GRS-S independent of the first author. The panel examined each of the 72 items and the student selection instructions on the GRS-S in terms of word choice, grammatical structure, fidelity of translation, and readability; discussed controversial points; and reconciled differences of opinion. Third, as a pilot test, the translated version was administered to eight educators in China similar to the targeted population of Chinese school teachers. Feedback from the pilot sample included comments on issues such as the wording of items and clarity of the instructions. Following minor revisions based on feedback from the pilot study, the new Chinese version was back-translated to English by a second bilingual translator who was not involved in the earlier translation process and was unfamiliar with the purpose of the study. Finally, the back-translated English version was compared to the original GRS-S to check whether each of the items retained its original meaning and intent. This final step included consulting with the first author of the GRS-S regarding the accuracy of the translation in comparison with the original English version of the GRS-S.

Participants and Procedure

The study was conducted in a metropolitan city in southwest China. Participants were 50 teachers, with 36 teachers from one elementary school (grades one through six, 6 teachers from each grade) and 14 teachers from one middle school (grades seven through eight, 7 teachers from each grade). In China, students have more than one teacher. However, there is one primary teacher who oversees the major academic content areas and is most familiar with each student. These primary teachers were invited to select students from their individual class and complete GRS-S ratings on the students. Class size in the two schools ranged from 40 to 55 students, which was a typical class size in Chinese schools. Middle school teachers in China are required to have a 4-year college degree in education and elementary school teachers have formal education from a 2- or 3-year teachers' college.

In these schools, there were approximately nine classes in each grade. The 50 teachers were instructed to select 10 students in their respective classes who met the following classroom academic performance

criteria: 2 students functioning very above average level, 2 students functioning above average level, 2 students functioning at the average level, 2 students functioning below average level, and 2 students functioning very below average level. This procedure ensured a heterogeneous sample of students based on academic abilities. The selection criteria were consistent with the procedure in other cross-cultural investigations (e.g., Lee & Pfeiffer, 2006). Teachers were also instructed to balance the number of boys and girls across the five categories. Teachers were trained in two sessions to carefully adhere to the GRS-S test instructions and provide objective and independent rating of each item on the six GRS-S scales.

The 50 teachers selected and rated 500 students using the Chinese version of the GRS-S. One GRS-S form did not have the required demographic information and was excluded from the study. The study included a total of 499 student rating forms. The mean age of the 499 students was 10.03 (SD = 1.96); there were 228 girls (46%) and 271 boys (54%).

The participating elementary school was affiliated with the teachers' university of the province, and the students were considered more academically advanced than were students attending regular schools in the city. The middle school, on the other hand, was not considered academically advanced and was a neighborhood school serving residents in the local community.

Instrumentation

GRS-S. The GRS-S for ages 6:0 to 13:11 consists of six scales (intellectual ability, academic ability, creativity, artistic talent, leadership ability, and motivation) with 12 items each for a total of 72 items. Each item is rated on a 9-point scale divided into three ranges: 1 to 3 = *below average*; 4 to 6 = *average*; 7 to 9 = *above average*. The standard (not translated, English) version of the GRS-S reports high levels of internal consistency (0.97 to 0.98), small Standard Error of Measurement and confidence intervals, and strong evidence of validity (Pfeiffer & Jarosewich, 2003). For more detailed information on the psychometric properties of the GRS-S, please refer to Margulies and Floyd (2004), Pfeiffer, Petscher, and Jarosewich (2007), and Ward (2005). For this study, raw scores for each of the scales were converted to standard T scores with a mean score of 50 and standard deviation of 10, following the steps described in the GRS manual. All the analyses were conducted using standard T scores.

Final examination scores. Standardized measures of academic status were not used in Chinese elementary and middle schools. Therefore, the 50 teachers were asked to report each student's most recent final examination scores. Students from all the schools in that city took the same final exams designed for their grade level in a variety of subjects including math, language, music, and art. The final examinations were cross-graded (teachers only graded students from other schools) using the same grading criteria. Thus, although the final exams were not nationally normed, they allowed fair and equitable comparisons of academic achievement level across schools and across students. For the purpose of this study, math, language, music, and art scores were used as criterion measures of the translated GRS-S.

Results

Reliability of GRS-S Chinese Translation

The first set of statistical analyses examined the reliability of the Chinese-translated GRS-S. The coefficient alpha values for the Chinese-translated GRS-S scores were high, with values of .97, .98, .95, .99, .97, and .98 for intellectual, academic, creativity, artistic talent, leadership, and motivation scales, respectively. These values are similar to those reported in the GRS manual (Pfeiffer & Jarosewich, 2003).

Pearson product moment correlations were computed for the six scales on the Chinese-translated GRS-S form; results are presented in Table 1. The highest correlation coefficient was between intellectual ability and academic ability (r[499] = .93, p < .01). This, again, is consistent with the analyses conducted with the U.S. standardization sample (Pfeiffer & Jarosewich, 2003; Pfeiffer et al., 2007).

All correlation coefficients were significant at the p < .01 level. The lowest correlation coefficients were between artistic ability and leadership ability (r[499] = .78) and creativity and leadership ability (r[499] = .78), and these two scales generated the lowest correlation coefficients with other scales (rs between .78 and .85).

Factor Structure of GRS-S Chinese Translation

A confirmatory factor analysis was conducted with the 499 student GRS-S ratings to test the hypothesized six-factor structure of the original GRS-S. As recommended by many authors in the field (e.g., Browne & Cudeck, 1993; Hsueh, Phillips, Cheng, & Picot, 2005; Hu & Bentler, 1998), the estimation of how well the hypothesized factor structure fits the observed data was tested using LISREL 8.7 (Jöreskog & Sörbom, 2004).

Table 1
Correlation Coefficients for Relations Between Chinese Gifted Rating Scales–School Form Scale Scores

	Intellectual Ability	Academic	Creativity	Artistic Talent	Leadership	Motivation
Intellectual Ability	1.00	.93	.90	.78	.79	.84
Academic		1.00	.91	.83	.85	.89
Creativity			1.00	.82	.78	.82
Artistic Talent				1.00	.79	.79
Leadership					1.00	.91
Motivation						1.00

Note: All coefficients are significant at $p < .01$.

Multiple fit indexes were used to assess model fit, including χ^2/df ratio, Comparative Fit Index (CFI), Normed Fit Index (NFI), Relative Fit Index (RFI), Root Mean Square Residual (RMR), and Root Mean Square Error of Approximation (RMSEA). The χ^2/df ratio provides a useful heuristic for evaluating model fit due to the sample size sensitivity of the χ^2 test; estimates less than 3.0 typically indicate good model fit. CFI, NFI, and RFI are fit indexes that compare the specified model to an independence model, with values greater than or equal to .90 being acceptable (Hu & Bentler, 1998). RMR estimates close to 0 indicate better model fit and are based on the square root of the mean of squared residuals. Lastly, the RMSEA is a parsimony measure given the model degrees of freedom. Hu and Bentler (1998) recommended estimates less than .06 as a cutoff for good model fit, although fit may be ascertained upward to .10. Results confirmed that the data fit well to a six-factor model ($\chi^2[2469] = 7269.84$, $p < .001$; CFI = .99; NFI = .99; RFI = .99; RMR = .03; RMSEA = .077; 90% CI = .075, .078; $\chi^2/df = 2.94$).

Standardized factor loadings and the proportion of variance in each variable accounted for by the relationships in the model (R^2s) are reported in Table 2. All items yielded statistically significant standardized factor loadings ($z > 1.96$), ranging between .49 and .96, indicating moderate to very strong correlations between the items and principle domains.

Because of comparatively high correlations among the scales of the Chinese version of the GRS-S, a unidimensional, single-factor model was also tested as a comparative analysis to the theoretical six-factor solution. Results indicated that the model did not fit as well as the initial proposed six-factor measurement model did ($\chi^2[2484] = 20272.96$, $p < .001$; CFI = .97; NFI = .97; RFI = .97; RMR = .06; RMSEA = .21; 90% CI = .20, .22; $\chi^2/df = 8.16$). All relative fit indexes of the single-factor model (i.e., NFI, RFI, CFI) were lower, noncentrality-based indexes (i.e., RMSEA) were higher, and absolute indexes (i.e., χ^2/df ratio, RMR) were higher for the unidimensional factor model than the six-factor model. These findings support a better fit of the six-factor model in comparison with the unidimensional model.

Table 2
Standardized Factor Loading Ranges for Chinese Gifted Rating Scales–School Form Scales

Scale	Standardized Factor Loadings Range	R^2 Range
Intellectual Ability	.52 to .96	.27 to .91
Academic	.83 to .96	.69 to .92
Creativity	.49 to .94	.24 to .88
Artistic Talent	.89 to .95	.61 to .89
Leadership	.53 to .94	.28 to .89
Motivation	.81 to .96	.61 to .91

Gender and Age Differences on the Chinese Version of GRS-S

To examine possible gender and age differences on the GRS-S Chinese version, we ran a 2 (gender) × 7 (age: 7, 8, 9, 10, 11, 12, 13) between-group MANOVA to analyze multivariate main effects. Preliminary analyses revealed that assumptions of multivariate normality and homogeneity of the covariance matrices (Box's $M = 23.39$, $p > .05$) were met. The multivariate main effect for gender was statistically significant, Wilks's Lambda = .84, $F(6, 477) = 14.30$, $p < .001$. The mean scale scores for girls and boys were significantly different for intellectual ability ($F[1,482] = 16.22$, $p < .001$, $d = .30$), academic ability ($F[1, 482] = 28.70$, $p < .001$, $d = .43$), creativity ($F[1, 482] = 14.06$, $p < .001$, $d = .30$), artistic talent ($F[1, 482] = 51.18$, $p < .001$, $d = .61$), leadership ($F[1, 482] = 43.13$, $p < .001$, $d = .57$), and motivation ($F[1, 482] = 45.36$, $p < .001$, $d = .57$). The largest difference was observed for the

Table 3
Means and Standard Deviations for the Scale Scores of Chinese Version of Gifted Rating Scales–School Form by Gender and Age Groups

	Scale											
	Intellectual		Academic		Creativity		Artistic		Leadership		Motivation	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Female (n = 228)	52.21	13.14	52.49	12.55	52.81	12.49	56.03	13.61	51.33	11.70	53.51	11.87
Male (n = 271)	48.25	13.28	47.19	12.29	49.05	12.71	47.96	12.84	44.63	11.67	46.81	11.76
7 years (n = 63)	46.84	10.64	47.33	10.08	50.49	10.26	50.53	10.93	47.81	10.65	48.25	9.02
8 years (n = 78)	48.31	11.95	47.91	11.31	49.60	12.59	50.82	12.91	46.68	11.47	47.78	11.07
9 years (n = 77)	51.18	15.83	49.49	15.01	50.91	14.54	52.88	15.80	47.22	14.05	49.56	13.66
10 years (n = 48)	47.69	14.38	48.73	13.96	48.46	13.70	51.17	16.05	46.27	13.41	49.91	12.86
11 years (n = 92)	51.39	12.59	51.07	12.00	52.12	11.38	52.63	13.69	47.90	11.62	50.97	11.88
12 years (n = 84)	55.20	13.46	54.39	12.45	55.49	13.07	64.65	13.35	51.50	11.11	54.58	12.54
13 years (n = 57)	46.75	12.13	45.96	12.01	45.26	11.18	46.58	12.53	44.77	12.22	46.14	12.82
Total (N = 499)	50.06	13.36	49.61	12.67	50.76	12.73	51.63	13.78	47.68	12.14	49.86	12.26

artistic talent scale (girls M = 56.03; boys M = 47.96). All the gender differences were in favor of girls, where girls were rated significantly higher by teachers. Means and standard deviations for gender are provided in Table 3.

The MANOVA comparing GRS-S scales based on age also yielded significant results, Wilks's Lambda = .82, $F(36, 2097) = 2.73$, $p < .001$. Given the number of planned comparisons among age effects, it was important to provide an alpha correction to control for the inflation of the family-wise error rate. The adjusted alpha rate for the post hoc analyses was .002. Post hoc test using Tukey's Honestly Significantly Different method revealed that the group of 12-year-old students differed from the 13-year-old students in their ratings in three of the six scales: academic ($F[6, 482] = 4.31$, $p < .001$, $d = .69$), creativity ($F[6, 482] = 4.52$, $p < .001$, $d = .84$), and motivation ($F[6, 482] = 4.60$, $p < .001$, $d = .38$), with 12-year-old students being rated higher than those in the 13-year-old age group. Despite the significant main effects, the MANOVA did not reveal significant Gender × Age interaction effects indicating that the differences between age groups were not dissimilar by gender for any of the scales. Means and standard deviations for age are also found in Table 3.

Criterion-Related Validity of the Chinese Version of GRS-S

Evidence for the criterion-related validity of the translated Chinese GRS-S was examined by comparing the scores on six scales on the GRS-S with the final exam scores of the four subjects (math, language, music, and art). With the critical alpha at .002, all GRS-S scales had a significant correlation with each subject of the final exam measure (estimates between .35 and .54). Table 4 provides correlation coefficients between Chinese GRS-S scale scores and the final exam scores.

Table 4
Correlation Coefficients for Relations Between Chinese Gifted Rating Scales–School Form Scale Scores and Final Examination Scores

	Language	Math	Music	Art
Intellectual	.50	.51	.36	.36
Academic	.53	.54	.42	.42
Creativity	.51	.52	.35	.38
Artistic	.47	.43	.41	.38
Leadership	.50	.47	.41	.42
Motivation	.52	.53	.41	.42

Note: All coefficients are significant at p = .000

Relationship Between Length of Time That a Teacher Knows the Student and GRS-S Ratings

We examined whether GRS-S ratings were influenced by how long or how well a teacher had known a student. The GRS-S record form includes the following two questions: "How long have you known the child?" and "How well do you think you know the child?" Responses to these two items range between

1 to 3 months and *over 1 year* for the first item and *not well* to *very well* for the second item. The multivariate main effects for the length of time and how well the teacher knew the child were not significant. As expected, neither was the multivariate interaction effect between these two variables significant.

Discussion

We explored the reliability, validity, and potential age and gender effects of a Chinese-translated version of the GRS-S used by Chinese teachers. We also examined the factor structure and criterion-related validity of teacher ratings and the impact of length of time and how well a teacher knew a student on the Chinese version of the GRS-S.

Findings from this study support the reliability of the Chinese-translated version of the GRS-S. The coefficient alphas were .95 to .99, which were well above the minimal standard of .80 (Bracken, Keith, & Walker, 1998). The findings of high reliability were consistent with those reported in the GRS-S test manual, with alpha coefficients ranging from .97 to .99 (Pfeiffer & Jarosewich, 2003; Ward, 2005). Results from the factor analyses also provided preliminary evidence that the factor structure of the Chinese-translated GRS-S reflected the authors' intended six-factor model (Pfeiffer & Jarosewich, 2003). This methodology provided evidence of the original measurement model outperforming a reduced version (unidimensional model) of the original solution. Future research may wish to test additional alternative models.

Furthermore, results provided preliminary support for the criterion validity of the Chinese-translated GRS-S with Chinese students. The correlation coefficients between the GRS-S scales and the final exam scores were significant. Although the final exams were not standardized, the scores met a standard of consistency as indicated in the instrument section. A closer examination of the pattern of relationships revealed some unanticipated findings. The correlations between final exam scores and leadership and creativity were similar to those between final exam scores and intellectual and academic scores. One explanation for this finding could be that Chinese teachers often appoint intellectually bright students and those who earn good grades as classroom leaders, helping teachers organize different activities, supervise peer academic work, and even provide discipline (Shanxia Wan, Ti Li, personal communication, April 20, 2006). In addition, teachers in China may view the construct of creativity differently than do teachers in the United States even though there was a careful translation of the 12 creativity items. Another related and unanticipated finding is that academic and leadership have higher correlations than creativity does with music and art final exam scores. Some researchers indicate that the Chinese concept of creativity is different from the Western concept (Rudowicz & Hui, 1998; Rudowicz & Yue, 2000). Creativity may not be regarded as a highly desirable characteristic for a Chinese student to possess and it may not be considered a type of giftedness in China (Rudowitz & Yue, 2000). Music and art in China are often regarded as a means to enhance academic performance rather than as unique talents in their own right, independent of academics (Li, Lee, Pfeiffer, & Petscher, in press; Shanxia Wan, Ti Li, personal communication, April 26, 2006). Future research is needed to further explore these fascinating results.

There was no difference in teacher ratings either in terms of length of time or familiarity with the child. This, however, does not necessarily mean that Chinese teachers provide accurate gifted ratings regardless of how well or how long they know the children. The finding indicates that teachers may be equally likely to give children high ratings regardless of length of time or extent of knowledge of the child.

The GRS-S mean scores differed significantly by gender for all six of the scales in the Chinese sample. On all of the scales, Chinese girls were rated higher by their teachers than were Chinese boys. Gender differences across the six scales ranged from modest to high, with girls rated approximately 0.4 *SD* higher (intellectual ability) to 0.8 *SD* higher (artistic talent) than boys, considered a medium-to-large effect size (Cohen, 1977). The gender differences in the Chinese sample in favor of females are consistent with the findings reported by Lee and Pfeiffer (2006) and Pfeiffer et al. (2007). Lee and Pfeiffer (2006) reported in their study with a Korean sample that girls were rated significantly higher on all six scales. Pfeiffer et al. (2007) found that U.S. girls' mean scores were significantly higher than boys' mean scores on three of the scales (artistic talent, leadership ability, and motivation) but not significantly higher for the other three scales. The standardization sample for the GRS-S followed a carefully prescribed and rigorous set of norming procedures; it is unlikely that the GRS-S standardization sample is unrepresentative or biased in a way that might explain the small yet consistent gender differences in favor of females. Gender differences are not only found in screening and

assessing giftedness but also in many areas of human development such as fears and anxieties, where girls are consistently rated as having higher levels of fears and anxieties (e.g., Ollendick, Yule, & Ollier, 1991; Shore & Rapport, 1998). It appears that teachers, both in the United States and in Asian countries, perceive female students in their classes as somewhat more talented artistically and creatively, more motivated, stronger leaders, and more academically and intellectually precocious when compared to their male counterparts. Another possible reason for the gender differences is that teachers are influenced by behavioral differences in girls and boys, which often affects how teachers judge or rate a student's ability level (Cole, Maxwell, & Martin, 1997). Teachers believe that boys generally misbehave more frequently than do females. Thus, teachers' judgments on student performance may be contaminated by their expectations and biases (Hoge, 1983). This, of course, is between-group difference and does not reflect significant within-group differences; many boys, both in the United States and China samples were rated above the mean for girls and, in some instances, extraordinarily high by their teachers.

The results reveal significant age differences that were not found in the United States standardization sample (Pfeiffer & Jarosewich, 2003). Specifically, 12-year-old students were rated significantly higher than were 13-year-old students on the academic, creativity, and the motivation scales. One possible explanation for the age difference between 12- and 13-year-old students is that most of the 13-year-old students were middle school students from a nonacademically challenging neighborhood school. In addition, Chinese students' admission to alternative postelementary school options depends on the results of their performance on a district-wide academic exam and their performance on music, science, and art projects. These highly competitive, district-wide (high-stakes) exams are typically administered when the students are 12 years old, corresponding to their last year enrolled in elementary school (Xiaoti Li, personal communication, May 11, 2005). Chinese students are oftentimes highly motivated or even pushed by their parents and teachers to perform at their highest potential at the end of their elementary school to best position themselves for a competitive postelementary school placement. This may explain, at least in part, the spike in GRS-S ratings by teachers for 12-year-old Chinese students. Further research will want to examine in more detail this fascinating and unanticipated finding. Nevertheless, consistent with the analyses of the original GRS-S standardization sample for students in the United States and a study of the GRS-S Korean version (Lee & Pfeiffer, 2006; Pfeiffer et al., 2007), these results did not show age differences within elementary school students (ages 7 to 12). This indicates that the Chinese version of the GRS-S may work equally well at least across elementary school students. Further studies with larger sample sizes are warranted to validate these findings.

Caution should be taken when interpreting results obtained from rating scales because of the possibility of a halo effect (Feeley, 2002). Teachers could have provided overestimated or inflated ratings because of the influence of a student's overall academic ability (i.e., the undue influence of *g* on the ability to make discriminations on the other nonacademic scales). However, the two sessions of rater training likely reduced or eliminated any halo effect.

China is a nation with vast regional, urban, and rural differences in terms of economic and educational development. Participants for this study were recruited from only the southwest region. Future studies are warranted to examine the psychometric characteristics of the Chinese-translated GRS-S across different regions before it is widely used as a gifted screening instrument in China.

References

Bracken, B. A., & Barona, A. (1991). State-of-the-art procedures for translating, validating, and using psychoeducational tests for cross-cultural assessment. *School Psychology International*, *12*, 119-132.

Bracken, B. A., Keith, L. K., & Walker, K. C. (1998). Assessment of preschool behavior and social-emotional functioning: A review of thirteen third-party instruments. *Journal of Psychoeducational Assessment*, *16*, 153-169.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Butcher, J. N. (Ed.). (1996). *International adaptations of the MMPI-2: A handbook of research and applications*. Minneapolis: University of Minnesota.

Chan, D. W. (2003). Adjustment problems and multiple intelligences among gifted students in Hong Kong: The development of the revised Student Adjustment Problems Inventory. *High Ability Studies*, *14*, 41-54.

Chan, D. W. (2004). Multiple intelligences of Chinese gifted students in Hong Kong: Perspectives from students, parents, teachers, and peers. *Roeper Review*, *27*, 18-24.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic.

Cole, D. A., Maxwell, S. E., & Martin, J. M. (1997). Reflected self-appraisals: Strength and structure of the relation of teacher, peer, and parent ratings to children's self perceived competencies. *Journal of Educational Psychology*, *89*, 55-70.

Cooperative Research Group of Supernormal Children of China. (1985). *Test for identifying the cognitive ability of supernormal children digest of education.* Beijing: Author.

Davidson, K. (1986). The case against formal identification. *Gifted Child Today*, *9*, 7-11.

Davis, G. A., & Rimm, S. B. (2004). *Education of the gifted and talented* (5th ed.). Boston: Pearson Education.

Feeley, T. H. (2002). Comment on halo effects in rating and evaluation research. *Human Communication Research*, *28*, 578-586.

Feldhusen, J. F., & Jarwan, F. A. (2000). Identification of gifted and talented youth for educational programs. In K. A. Heller, F. J. Monks, R. J. Sternberg, & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (2nd ed., pp. 271-282). Oxford, UK: Elsevier.

Gallagher, J. J. (2003). Issues and challenges in the education of gifted students. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 11-13). Boston: Allyn & Bacon.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, *6*, 304-312.

Hoge, R. D. (1983). Psychometric properties of teacher-judgment measures of pupil aptitudes, classroom behaviors, and achievement levels. *Journal of Special Education*, *17*, 401-429.

Hoge, R. D., & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. *Journal of Educational Psychology*, *76*, 777-781.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of the literature. *Review of Educational Research*, *59*, 297-313.

Hoover, H. D., Hieronymus, A. N., Dunbar, S. B., & Frisbie, D. A. (2001). *Iowa tests of basic skills*. Itasca, IL: Riverside.

Hsueh, K. S., Phillips, S. R., Cheng, W. Y., & Picot, S. J. F. (2005). Assessing cross-cultural equivalence through confirmatory factor analysis. *Western Journal of Nursing Research*, *27*, 755-771.

Hu., L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424-453.

Jarosewich, T., Pfeiffer, S. I., & Morris, J. (2002). Identifying gifted students using teacher rating scales: A review of existing instruments. *Journal of Psychoeducational Assessment*, *20*, 322-336.

Jöreskog, K. G., & Sörbom, D. (2004). *LISREL 8.7 for Windows* [Computer Software]. Lincolnwood, IL: Scientific Software International.

Kenny, D. T., & Chekaluk, E. (1993). Early reading performance: A comparison of teacher based and test-based assessments. *Journal of Learning Disabilities*, *26*, 227-236.

Lee, D., & Pfeiffer, S. I. (2006). The reliability and validity of a Korean-translated version of the Gifted Ratings Scales. *Journal of Psychoeducational Assessment*, *24*, 210-224.

Li, H., Lee, D., Pfeiffer, S. I., & Petscher, Y. (in press). Parent rating using Gifted Rating Scales–School Form: Reliability and validity in Chinese students. *Educational and Psychological Measurement*.

Margulies, A. S., & Floyd, R. G. (2004). Test review of the Gifted Rating Scales (GRS). *Journal of Psychoeducational Assessment*, *22*, 275-282.

Meisels, S. J., Bickel, D. D. P., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment in kindergarten to grade 3. *American Educational Research Journal*, *38*, 73-95.

National Evaluation Systems. (2005). *California basic educational skills test*. Retrieved November 16, 2005, from http://www.cbest.nesinc.com/

Ollendick, T. H., Yule, W., & Ollier, K. (1991). Fears in British children and their relationship to manifest anxiety and depression. *Journal of Child Psychology and Psychiatry*, *32*, 321-331.

Perry, N. E., & Meisels, S. J. (1996). *Teachers' judgments of students' academic performance* [Working Paper No. 96-08, National Center for Education Statistics]. Washington, DC: U.S. Department of Education.

Pfeiffer, S. I. (2002). Identifying gifted and talented students: Recurring issues and promising solutions. *Journal of Applied School Psychology*, *1*, 30-50.

Pfeiffer, S. I., & Jarosewich, T. (2003). *Gifted rating scales*. San Antonio, TX: Harcourt.

Pfeiffer, S. I., Petscher, Y., & Jarosewich, T. (2007). The Gifted Rating Scale–School Form: Analysis of the standardization sample based on age, gender, race, and diagnostic efficiency. *Gifted Child Quarterly*, *51*, 39-50.

Richert, E. S., Alvino, J. J., & McDonnel, R. C. (1982). *National report on identification: Assessment and recommendations for comprehensive identification of gifted and talented youth.* Washington, DC: U.S. Department of Education, Educational Information Resource Center.

Rudowicz, E., & Hui, A. (1998). Hong Kong Chinese people's view of creativity. *Gifted Education International*, *13*, 159-174.

Rudowicz, E., & Yue, X. (2000). Concepts of creativity: Similarities and differences among Mainland, Hong Kong and Taiwanese Chinese. *Concepts of Creativity*, *34*(3), 175-191.

Sharpley, C. F., & Edgar, E. (1986). Teachers' ratings vs. standardized tests: An empirical investigation of agreement between two indices of achievement. *Psychology in the Schools*, *23*, 106-111.

Shi, J., & Zha, Z. (2000). Psychological research on and education of gifted and talented children in China. In K. A. Heller, F. J. Monks, R. J. Sternberg & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (2nd ed., pp. 757-764). Oxford, UK: Elsevier.

Shore, G. N., & Rapport, M. D. (1998). The Fear Survey Schedule for Children–Revised (FSSC-HI): Ethonocultural variations in children's fearfulness. *Journal of Anxiety Disorder*, *12*, 437-461.

Siegel, D., & Powell, T. (2004). Exploring teacher biases when nominating students for gifted programs. *Gifted Child Quarterly*, *48*, 21-29.

Silverstein, A. B., Brownlee, L., Legutki, G., & MacMillan, D. L. (1983). Convergent and discriminant validation of two methods of assessing three academic traits. *Journal of Special Education*, *17*, 63-68.

Sternberg, R. J. (1997). The concept of intelligence and its role in lifelong learning and success. *American Psychologist*, *52*, 1030-1037.

Stevenson, H. W. (1998). Cultural interpretations of giftedness: The case of East Asia. In R. C. Friedman & K. B. Rogers (Eds.), *Talent in context* (pp. 61-77). Washington, DC: American Psychological Association.

Tannenbaum, A. J. (2000). A history of giftedness in school and society. In K. A. Heller, F. J. Monks, R. J. Sternberg, & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (2nd ed., pp. 23-51). Oxford, UK: Elsevier.

Ward, S. A. (2005). Gifted rating scales. In R. A. Spiers & B. S. Plake (Eds.), *16th mental measurements yearbook* (pp. 404-407). Lincoln: University of Nebraska Press.

Huijun Li, PhD, is the Director of Multicultural in the Commonwealth Research Center (CRC) at the Massachusetts Mental Health Center (MMHC) and the Beth Israel Deaconess Medical Center (BIDMC) Department of Psychiatry, and a faculty of Harvard Medical School. Her research interests include cross-cultural assessment and treatment of fears, anxieties, and depression in children and adolescents. She is also interested in creating culturally valid instruments. This study was conducted when she was an assistant professor at Florida State University.

Steven I. Pfeiffer, PhD, is a professor at Florida State University, where he serves as the director of the doctoral program of the Combined Counseling Psychology and School Psychology. Prior to his tenure at Florida State University, he was executive director of Duke University's Talent Identification Program. He is on the board of Supporting Emotional Needs of the Gifted and president-elect of the American Psychological Association Consortium of Combined-Integrated Doctoral Programs in Professional Psychology. He frequently lectures on the psychological needs of the gifted. He is editor of *Handbook of the Gifted & Talented: A Psychological Approach*, published by Kluwer/Plenum.

Yaacov Petscher, MS, is the Director of Research at the Florida Center for Reading Research and a doctoral candidate in developmental psychology at Florida State University with a focus on applied quantitative methods and statistics. His research interests include academic achievement motivation, and the applications of item response theory and multilevel frameworks.

Alper T. Kumtepe is an assistant professor at Anadolu University, Turkey. His research interests include early childhood education, giftedness in early childhood, measurement and evaluation, instructional design, and ePortfolios for early childhood teacher education.

Guofang Mo is a professor in Yunnan Teachers' University, Kunming, China.