Introduction

The HOPE Teacher Rating Scale (referred to as HOPE Scale in this manual) is an instrument created at Purdue University designed to aid in the identification of gifted and talented students. Classroom teachers complete the HOPE Scale on each of their students (ideally as a universal assessment for all students) by responding to 11 items using a 6-point frequency-response scale. This instrument was developed as part of Project HOPE (Having Opportunities Promotes Excellence), a 3-year project designed to identify and serve high-potential students from low-income families in out-of-school enrichment programs funded by the Jack Kent Cooke Foundation. Its development continued under Project HOPE+—another project funded by the Jack Kent Cooke Foundation—this time serving high-potential Native American youth in grades 5–12 from five communities on four different reservations. One of the goals of Project HOPE was to develop procedures for recognizing potential among children from low-income backgrounds. Children from low-income and culturally diverse families represent the smallest proportion of students performing at the highest levels of achievement (Plucker, Burroughs, & Song, 2010; Plucker, Hardesty, & Burroughs, 2013). Youth from low-income families who do achieve at high levels are less likely to persist as high achievers; twice as likely to drop out of high school; less likely to attend selective colleges; more likely to attend the least selective colleges; and if they do attend college, are less likely to graduate than their higher income academic peers (Wyner, Bridgeland, & DiIulio, 2009). Additionally, students from low-SES households face unique challenges as their families may lack social capital or access to resources necessary to adequately prepare and support student achievement in school (Demi, Coleman-Jensen, & Snyder, 2010; Schunk, Pintrich, & Meece, 2008). The HOPE Scale was developed because the recognition of talent is a precursor to those talents being nurtured. The HOPE Scale can be used to identify students from all income and ethnic groups for gifted programming, but it was carefully designed to more equitably identify students from low-income families. It includes two subscales—Academic and Social—that take into account the multifaceted dimensions of giftedness. Items on each of these subscales were developed to be as income and culturally neutral as possible. The HOPE Scale then underwent extensive study with diverse samples of students. Its development and further refinement continues today.

Literature Review

Using Teacher Nominations and Rating Scales

Teacher rating scales are one type of instrument commonly used in the identification of students for gifted program services. Teacher referrals are very common and are often the first step in the identification process preceding a student's formal identification as gifted (National Association for Gifted Children [NAGC], 2013). Other criteria/methods used across the United States include multiple measures, IQ scores, achievement data, and state-approved assessments (NAGC, 2013). Many teacher-rating scales exist (e.g., Pfeiffer & Jarosewich, 2003; Renzulli et al., 2013; Ryser & McConnell, 2004) and are frequently used by school personnel as part of their identification process. However, many educators in many schools and school districts make the common mistake of creating their own instrument, and in doing so, compromise the validity and reliability of any scores they generate from data collected with their "homemade instrument."

Research findings have been mixed regarding the efficacy of teacher nominations for programs for the gifted. Pegnato & Birch (1959) argued against the use of teachers as identifiers of giftedness, concluding that group intelligence tests had the best combination of effectiveness and efficiency. Alternatively, Gagné (1994) concluded that teacher nominations were just as effective as other methods of identification; while others found that teachers were effective identifiers of giftedness when training on characteristics of gifted students was provided (Hunsaker, Finley, & Frank, 1997).

In his 2006 study, McBee found that teachers did a better job referring Asian, Anglo American, and Native American students for gifted programs than they did referring African American and Hispanic students. Teachers nominated almost 10% of Asian students for gifted programs and only 1.81% of Hispanic students. McBee discussed the implications of his results and raised some interesting questions regarding the quality of teacher referrals for gifted programs. He questioned whether the referral process as a whole was biased against low-SES, Black, and Hispanic students, but did not reach any final conclusion regarding that question. He suggested that the differences in quality and quantity of teacher referrals of students from traditionally underrepresented groups for gifted programs could indicate "racism, classicism, or cultural ignorance" (p. 109). He also questioned whether this was a sign that ability is not evenly distributed across different races and SES groups, or whether the problem was in the nomination process or in the screening. McBee reiterated that, regardless of the causes of these issues, his analyses provided statistical evidence that the quality of teachers' referrals for gifted programs was higher than what previous studies had reported (e.g., Pegnato & Birch, 1959; Gagné, 1994).

Researchers have suggested that teachers are in a prime position to recognize talent in their classrooms (High & Udall, 1983; Peterson, 1999) and that teacher rating instruments have potential as screening tools or additional pathways to identifying students from low-income families (Stambaugh, 2007; VanTassel-Baska, 2008). In their review of the literature, Hodge and Cudmore (1986) concluded that with explicit definitions of giftedness and a well-developed instrument, "the use of teacher judgments in the identification of gifted children should be continued, and, in fact, expanded" (p. 192).

However, in many studies, teacher referrals did not follow any defined criteria. Rather, teachers were merely asked to refer students for gifted programs and were not provided with an instrument or structure to help them identify characteristics that are commonly found among gifted students. Jarosewich, Pfeiffer, and Morris (2002) discussed some of the benefits of teacher rating instruments as a tool for nominating students for gifted programs, suggesting that teacher rating scales were an "inexpensive, efficient, and structured means of collecting teacher data, which can complement scores on academic and intellectual tests" (p. 332). Thus, it seems prudent to work toward developing quality tools that can help teachers recognize student potential and provide valuable, quality information for use in the identification process that is different from commonly and frequently exclusively (NAGC, 2013) used achievement and aptitude scores.

HOPE Scale Subscales

Some teacher-rating scales used for nomination and identification of gifted students contain as many as 14 separate subscales that are often highly correlated, making it difficult for raters to be able to distinguish among them and also making them extremely time-consuming to complete (Peters & Gentry, 2010). The two dimensions of giftedness measured by the HOPE Scale, Academic and Social, were selected in accordance with the federal definition acknowledging that gifted, creative, and talented youth exist within all cultural and economic groups in a variety of areas of human endeavor. This same definition acknowledged outstanding talent or potential for outstanding talent as recognizable when young people are compared with others "similar in age, experience, or environment" (United States Department of Education, 1993, p. 3). Because teachers have firsthand knowledge of their students' academic and social characteristics, they are uniquely qualified to provide an accurate rating of their students' behaviors based upon their day-to-day experiences in the classroom

Academic subscale. Academics have been the primary focus for most gifted identification processes and programming (NAGC, 2013). Renzulli (2005) identified two types of giftedness: schoolhouse and creative-productive. Schoolhouse giftedness refers to those students who display several "teacher-pleaser" character-

istics, including maturity, ease of learning, and test-taking ability. Students with creative-productive giftedness can be risk-takers and abstract thinkers who "will actually have an impact on others and cause change" (Renzulli, 2005, p. 253). Renzulli does not view these two types of giftedness as mutually exclusive, but suggested that schoolhouse giftedness is more frequently identified (and valued) in schools, although it may actually be less important in the long run. Consequently, the Academic subscale requires teachers to think about both of these types of giftedness and includes items related to high levels of academic performance, eagerness to explore new concepts, and using alternative processes. Academic items were written to be culturally and economically unbiased and include the following:

Item 1. Performs or *shows potential* for performing at remarkably high levels.
Item 6. Is eager to explore new concepts.
Item 7. Exhibits intellectual intensity.
Item 9. Uses alternative processes.
Item 10. Thinks "outside the box."
Item 11. Has intense interests.

Social subscale. By also emphasizing social aspects of giftedness, the Social subscale allows teachers to evaluate behaviors outside of academics, giving them the opportunity to identify students with potential talent who might otherwise have been overlooked in an academic-focused nomination process. Often, economically disadvantaged students and those from culturally and linguistically diverse backgrounds find themselves faced with unique social, cultural, and psychological pressures (Moore, Ford, & Milner, 2005). As a result they are frequently not identified or underidentified for gifted and talented programs, and those who are identified often drop out of these programs due to a lack of cultural peers and culturally sensitive teachers and inappropriate preparation for academic rigor (Bernal, 2007; Olszewski-Kubilius, Lee, Ngoi, & Ngoi, 2004; Worrell, 2007). The inclusion of a Social subscale was especially important because one goal was to equitably identify students from these underserved populations who are often not identified when traditional aptitude and achievement measures are the only pathway for identification. The Social subscale includes items related to leadership, interactions with adults and older students, and self-awareness. Statements from the Social subscale include:

Item 2. Is sensitive to larger or deeper issues of human concern.
Item 3. Is self-aware.
Item 4. Shows compassion for others.
Item 5. Is a leader within his/her group of peers.
Item 8. Effectively interacts with adults or older students.

Development of the HOPE Scale

Project HOPE

Project HOPE (Having Opportunities Promotes Excellence) was funded by the Jack Kent Cooke Foundation from 2007–2010. The goals of the project were to (a) develop procedures for recognizing ability and talent among low-income children (i.e., the HOPE Scale); (b) provide services for identified students from low-income families in Purdue's Gifted Education Resource Institute (GERI) programs for high-ability students; (c) develop follow-up services for the project participants; and (d) evaluate the effects of the project on the students and on the identification process in the participating schools (Gentry, 2007). One of the outcomes of Project HOPE was the HOPE Scale, an instrument that may be useful to help identify culturally and linguistically diverse students, as well as students from low-income families for gifted education programs. Teachers used the HOPE Scale to rate academic and social characteristics of their students.

HOPE Scale Development Summary

After reviewing the literature related to involving teachers in the identification process and definitions of giftedness and characteristics of gifted students, especially those from low-income backgrounds, and due to the scarcity of teacher rating scales developed for identifying those students with strong validity and reliability information, the Project HOPE research team developed, piloted, and studied the HOPE Scale as a rating scale that can help teachers recognize potential among children from low-income backgrounds. The scale was originally designed to be used with students in kindergarten through fifth grade, and more recently has also been used with students in grades 6 through 12.

The Project HOPE team, which included four faculty members and two graduate students, reviewed the literature on gifted and talented student behaviors and created a list of items describing such behaviors. The goal was to create items describing observable characteristics associated with academic and social components of giftedness and talent that are normally observed by teachers in their daily interactions with students. Another goal for developing such items was to allow teachers to compare students to others of similar background and experience. Thus, the following direction was included in the HOPE Scale: "When completing this form, please respond by thinking about the student compared to other children similar in age, experience, and/or environment" with the purpose of helping teachers nominate the top students in the various age, ethnic, and experience groups within their classrooms or schools (Lohman, 2006). The Project HOPE team created many items that could potentially be included in the HOPE

Scale, including items with similar wording. The team reviewed all of the items and revised any that might have been unclear before they agreed on the final list of 13 items to be included in the first version of the HOPE Scale.

This version used a Likert-type frequency response format ranging from never to always (1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = almost always, 6 = always). Comrey (1988) recommended that rating scales contain at least five points, and a 6-point scale was used to eliminate neutral scores. The first version, as well as all subsequent versions of the HOPE Scale, also contained one item providing teachers with an opportunity to indicate all content areas in which a student shows talent, and demographic items (race, ethnicity, free or reduced lunch status) to allow for comparison of scores across different groups and development of local norms. Teachers are encouraged to provide additional information on each child's potential. Because the scale is short, teachers in the sample were willing to complete it on all of their students, resulting in a high response rate and providing evidence about teachers' willingness to use the scale.

Content-related evidence of validity. The process of gathering evidence of the validity of a test's content involves making sure a test includes items representative of the proposed universe of content (Cronbach, 1971). The Project HOPE research team reviewed the literature on gifted and talented student behaviors and definitions of giftedness and wrote items related to two components often considered when identifying gifted students: Academic and Social. Initial content validity was established by sending the items to 19 experts on the behavior of gifted and talented students, including program coordinators, graduate students, university faculty, and veteran instructors of gifted and talented students. The experts were asked to assign items to a category. Any item that was correctly assigned by 17 of the 19 experts and had a confidence mean of 2.5 or greater on a 3-point scale was included in the first version of the scale. The research team made word-level changes using the feedback from the content experts, and the first version of the instrument included three items related to social and 10 items related to academic components of giftedness and talent.

Exploratory and confirmatory factor analysis procedures. Initial construct validation was completed using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) on a sample of 5,955 HOPE Scales completed by 349 teachers. Exploratory factor analysis procedures were used to determine the initial factor structure of the scale. These results indicated that a two-factor solution accounted for 99% of the total variation in the data collected using the first version of the scale (Peters & Gentry, 2010). Results of confirmatory factor analysis (CFA) procedures indicated a relatively good model fit for the first version of the scale (Peters & Gentry, 2010). Next, the Project HOPE research team used the results of the EFA and CFA to make changes to the original HOPE Scale to create an instrument that addressed the two latent factors (i.e., the Academic and Social

subscales). The team removed and deleted items and created a revised instrument with 17 items. This second version of the HOPE Scale was completed by 71 teachers on 1,700 K–5 students. Indicators of model fit and parsimony were used to remove any items with poor fit or high residual values or modification indices (Peters & Gentry, 2012). The third version of the HOPE Scale (see Appendix A) contains 12 items together with demographic information and is the instrument contained in this manual for use in schools with teachers and students. Teachers respond to 11 of those items using a Likert-type frequency response format ranging from never to always (1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = almost always, 6 = always), and to item 12 indicating content areas in which a student shows talent. They also complete demographic items and can make additional comments on a student's potential using space provided on the scale.

Model fit statistics for the third version of the HOPE Scale indicate moderate to good model fit (RMSEA = .101; CFI = .96; GFI = .91). Internal consistency estimates were strong, with .96 for the Academic scale and .92 for the Social scale. The interfactor correlation between the two scales is also high at .89, but because there are only two factors, a higher order factor would not help explain the data, and a single factor in place of the two factors does not fit the theory. When tested as a single factor, the result was poor fit in the base model. For these reasons, we elected to keep two correlated scales on the instrument despite the high intercorrelation (Peters, 2009).

HOPE Scale Development Details

Several studies as summarized above were undertaken in order to pilot, revise, and provide validity evidence for the HOPE Scale. These samples and methods are detailed in the following pages to provide the reader with a complete understanding of the development samples, and the steps taken to create a quality, parsimonious, and user-friendly instrument that should yield valid and reliable data to provide information from teachers in the process of identifying students for gifted program services.

Elementary development sample. The teachers and students who comprised the initial development sample were all from Indiana. A group of 357 teachers completed HOPE Scales on 5,995 students in kindergarten through fifth grade. The first version of the HOPE Scale, which contained 13 items, was used with this sample. The racial/ethnic and income demographics of this sample can be seen in Table 1.

The sample used for additional confirmatory factor analysis procedures on the second version of the HOPE Scale included 1,700 students from three school districts. Table 2 includes demographic information on these schools.

Table 1

Elementary Sample (First Version) Demographic Characteristics by School Corporation (n=5,995)

State	Indiana	Indiana	Indiana	Indiana	Indiana
Designation	Rural	Rural	Rural	Metro	Metro
HOPE Scales Returned	405	557	692	1,528	2,813
FARM Eligible	36%	38%	34%	62%	58%
Caucasian	96%	90%	91%	59%	60%
African American	0%	<1%	<1%	<1%	10%
Hispanic	2%	5%	8%	37%	21%
Asian	<1%	<1%	0%	<1%	<1%
Multiracial	<1%	4%	1%	3%	8%
Native American	0%	<1%	0%	<1%	<1%

Note. FARM eligible are those students who are eligible for the federal free and reduced meal program. Adapted from "Multi-group construct validity evidence of the HOPE Scale: Instrumentation to identify low-income elementary students for gifted programs," by S. J. Peters and M. Gentry, 2010, *Gifted Child Quarterly, 54,* p. 302. Copyright 2010 by National Association for Gifted Children. Adapted with permission.

Table 2

Elementary Sample (Second and Third Versions) Demographic Characteristics by School Corporation (n=1,700)

State	Indiana	Indiana	Illinois
Designation	Metro	Rural	Metro
FARM Eligible	59%	14%	43%
Caucasian	22%	96%	33%
African American	33%	1%	17%
Hispanic	16%	1%	36%
Asian	20%	<1%	12%
Multiracial	5%	1%	0%
Native American	1%	<1%	2%
Unknown Ethnicity	3%	0%	0%
Gender	51% Male	51% Male	48% Male

Note. Adapted from "Additional validity evidence and across-group equivalency of the *HOPE Teacher Rating Scale*," by S. J. Peters and M. Gentry, 2012, *Gifted Child Quarterly, 57,* p. 90. Copyright 2012 by National Association for Gifted Children. Adapted with permission.

Secondary development sample. Since 2011, the HOPE Scale has been used with secondary education students. The Jack Kent Cooke Foundation funded a grant, Project HOPE+, in 2011 to bring 50–70 Native American students in grades 6–12 from four reservation communities to Purdue University each summer to participate in a 2-week residential program for gifted students. Four reservation communities that had established relationships with individuals associated with the project were selected to partner in this initiative. A fifth reservation community was added in 2013 after school officials contacted the faculty member in charge of the project and asked to participate in the HOPE Scale secondary validation study.

Native American participants. The focus of the initial studies on using the HOPE Scale with secondary students was on identifying Native American students, so the secondary sample primarily includes students from the communities that partnered with Purdue University to send students to the university's summer programs. A list of the tribes involved in the project, the number of participating teachers, completed scales at each site, and demographic information of the students are shown in Table 3.

Non-Native American participants. A second sample of 359[1] non-Native American sixth- through 12th-grade students was collected from 13 teachers in five schools located in Arizona, Indiana, Missouri, and Ohio. The racial/ethnic and income demographics of the students in this sample, and completed scales at each site are shown in Table 4. This sample included 177 (49.3%) males. Individuals were recruited for the study by a call for participants posted on the website of the university sponsoring the study and by word of mouth.

Construct Validity Support

Factor structure. With 11 items, the HOPE Scale, after initial revision, has two subscales: Academic and Social. The Academic subscale items include: 1, 6, 7, 9, 10, and 11, and the Social subscale items include: 2, 3, 4, 5, and 8. Table 5 shows the factor loadings and items for each scale for the elementary sample.

Factor loadings for the elementary sample ranged from .828 to .937 for the Academic subscale and from .794 to .879 for the Social subscale. All factor loadings are sufficiently high, with a minimum t-value of 77.03, which indicates that the items fit the hypothesized model. Model fit statistics were in the acceptable range with Goodness of Fit Index (GFI) estimated at 0.910, which is greater than the optimal 0.90 or greater, Comparative Fit Index (CFI) of 0.956, and the less than ideal RMSEA of 0.111, which is greater than cutoff for good fit of 0.08.

Factor loadings for the secondary Native American sample ranged from .798 to .927 for the Academic subscale and from .725 to .935 for the Social subscale. All

1 Data collection is ongoing.

Table 3

Secondary Native American Sample
Demographic Characteristics (n=1,095)

State	Arizona	Arizona	Minnesota	Minnesota	South Dakota
Tribe	Diné	Diné	Ojibwe	Ojibwe	Lakota
Designation	Rural	Rural	Rural	Rural	Rural
Teachers	30	8	8	23	6
HOPE Scales Returned	640	113	267	56	19
FARM Eligible	100%	100%	100%	100%	100%
Gender	48% Male	52% Male	45% Male	41% Male	52% Male

Table 4

Secondary Non-Native American Sample
Demographic Characteristics (n=359)

State	Ohio	Arizona	Indiana	Indiana	Missouri
Designation	Urban	Suburban	Suburban	Rural	Rural
HOPE Scales Returned	67	101	131	33	27
FARM Eligible	84%	72%	69%	21%	55%
Caucasian	69%	24%	67%	92%	94%
African American	23%	11%	15%	1%	2%
Hispanic	3%	61%	5%	3%	3%
Asian	<1%	2%	<1%	<1%	<1%
Multiracial	4%	<1%	11%	2%	<1%
Native American	0	1%	<1%	<1%	<1%
Gender	49% Male	50% Male	45% Male	60% Male	52% Male

factor loadings are sufficiently high, with a minimum t-value of 64.62, which indicates that the items fit the hypothesized model. Model fit statistics for the Native American sample were in the acceptable range. The Goodness of Fit Index (GFI) was 0.905, which is greater than the optimal 0.90. However, the Comparative Fit Index (CFI) was 0.983, and the RMSEA was 0.114, which is greater than the 0.08 suggested by previous research (Hu & Bentler, 1999). However, these findings are consistent with previous validation studies of the HOPE Scale by Peters and Gentry (2010, 2012).

Table 5

CFA Factor Loadings for Elementary Sample (n=1,542)

Item	Academic	Social
1. Performs or *shows potential* for performing at remarkably high levels.	0.884	
6. Is eager to explore new concepts.	0.901	
7. Exhibits intellectual intensity.	0.937	
9. Uses alternative processes.	0.891	
10. Thinks "outside the box."	0.924	
11. Has intense interests.	0.828	
2. Is sensitive to larger or deeper issues of human concern.		0.854
3. Is self-aware.		0.879
4. Shows compassion for others.		0.794
5. Is a leader within his/her group of peers.		0.829
8. Effectively interacts with adults or older students.		0.836

Table 6

CFA Factor Loadings for Native American Sample (n=1095)

Item	Academic	Social
1. Performs or *shows potential* for performing at remarkably high levels.	0.878	
6. Is eager to explore new concepts.	0.858	
7. Exhibits intellectual intensity.	0.841	
9. Uses alternative processes.	0.833	
10. Thinks "outside the box."	0.798	
11. Has intense interests.	0.927	
2. Is sensitive to larger or deeper issues of human concern.		0.920
3. Is self-aware.		0.725
4. Shows compassion for others.		0.947
5. Is a leader within his/her group of peers.		0.935
8. Effectively interacts with adults or older students.		0.880

Table 7

CFA Factor Loadings for Secondary Non-Native American Sample (n=359)

Item	Academic	Social
1. Performs or *shows potential* for performing at remarkably high levels.	0.753	
6. Is eager to explore new concepts.	0.862	
7. Exhibits intellectual intensity.	0.865	
9. Uses alternative processes.	0.930	
10. Thinks "outside the box."	0.934	
11. Has intense interests.	0.738	
2. Is sensitive to larger or deeper issues of human concern.		0.909
3. Is self-aware.		0.866
4. Shows compassion for others.		0.822
5. Is a leader within his/her group of peers.		0.700
8. Effectively interacts with adults or older students.		0.804

Factor loadings for the secondary non-Native American sample ranged from .738 to .934 for the Academic subscale and from .700 to .909 for the Social subscale. All factor loadings are sufficiently high, with a minimum t-value of 24.28, which indicates that the items fit the hypothesized model. The Goodness of Fit Index (GFI) was lower than expected at 0.738. However, the Comparative Fit Index (CFI) was 0.927. The RMSEA was also high at 0.198. This finding is consistent with previous validation studies of the HOPE Scale by Peters and Gentry (2010, 2012) and the results of the CFA of the Native American sample. A more robust sample of non-Native American HOPE Scales may result in an improved fit for the model.

Alpha reliability estimates. In general, reliability estimates indicate consistency. One measure of reliability is the alpha internal consistency of the items. Alpha has a possible range from 0 to 1, with higher alpha levels representing higher reliability, and indicating that the items in the measured scale tend to be answered similarly. This is preferred within each subscale because the items are meant together to define and measure the construct (McCoach, Gable, & Madura, 2013). Using the elementary development sample, the academic subscale yielded an alpha reliability estimate of 0.96. The same sample yielded an alpha reliability estimate for the social subscale of 0.92. The secondary Native American sample yielded alpha reliability estimates of 0.97 for the Academic subscale and

0.93 for the Social subscale. These levels of internal consistency are high, and considered reasonable for individual and group decisions. Tables 8 and 9 include the descriptive statistics for the elementary and secondary samples, as well as response percentages, correlations with total, and alpha reliability estimates if removed for each HOPE Scale item.

Criterion-related evidence of validity. Criterion-related validity involves comparing scores on the instrument under investigation with scores on instruments with similar purposes. In the case of a teacher rating scale used to identify gifted and talented students, measures of criterion-related evidence could include the correlation between the HOPE Scale and achievement tests. Two forms of criterion-related validity include concurrent and predictive validity. Estimates of concurrent validity for the HOPE Scale were computed by correlating the HOPE Scale with math, reading, and science achievement scores from the same students and estimates of predictive validity were obtained by correlating HOPE Scale scores with math and reading achievement scores from one year after HOPE Scale scores were collected. Table 10 shows concurrent and predictive validity estimates for the HOPE Scale.

Concurrent validity estimates for the HOPE Scale (see correlations between HOPE Scale and the Midwest State Achievement Test [MSAT; a pseudonym] in Table 10) ranged from .50 to .56 (Peters & Gentry, 2012). These results are similar with concurrent validity estimates of similar teacher rating scales (Ryser & McConnell, 2004). Predictive validity estimates for the HOPE Scale (see correlations between HOPE Scale scores and MSAT 09 in Table 10) ranged from .46 to .55. These estimates of predictive validity are similar to the estimates of concurrent validity, which indicates that the HOPE Scale yields similar results when administered at the same time as an achievement test, or if the achievement test is administered a year later. Peters and Gentry (2012) also investigated the predictive validity of the HOPE Scale using a Tobit regression model and found that the HOPE Scale explained 51.9% (Social scale) to 57.9% of the variation in math and reading achievement scores.

Interrater reliability. Limited information exists on the HOPE Scale interrater reliability because most of the research on the instrument has involved asking classroom teachers to rate all of their students using the HOPE Scale. Thus, most of the studies included only one rating on each student. However, Pereira (2011) investigated the extent to which classroom and English as a second language (ESL) teachers' HOPE Scale scores on the same students differed. Correlations between ESL and classroom teacher scores were .58 (Academic subscale) and .57 (Social subscale). Even though these correlation estimates are within the moderate range, they need to be interpreted with caution because the teachers completing HOPE Scales on the same students had different roles within their schools. Renzulli et al. (2013) reported a correlation of .50 between the ratings of teachers in two different

Table 8

HOPE Scale Descriptive Statistics for Elementary Sample (n=1,542)

Factor	Item	Response Percentage						Mean	*SD*	*r* with total[a]	Alpha if removed[b]	Alpha
		1	2	3	4	5	6					
Academic	1	11	24	24	16	15	9	3.27	1.47	.86	.95	.96
	6	8	20	27	16	15	13	3.51	1.49	.87	.95	
	7	11	26	27	14	13	9	3.21	1.46	.91	.95	
	9	10	27	30	16	12	6	3.11	1.35	.87	.95	
	10	12	28	25	15	12	8	3.10	1.44	.91	.95	
	11	10	26	31	14	11	8	3.12	1.38	.80	.96	
Social	2	10	23	32	18	11	5	3.43	1.35	.81	.90	.92
	3	7	18	34	19	13	10	3.54	1.30	.82	.90	
	4	6	12	33	22	17	11	3.76	1.33	.79	.91	
	5	14	28	27	15	10	7	3.15	1.48	.76	.91	
	8	8	14	31	22	15	10	3.77	1.41	.81	.90	

Note. [a]standardized correlations. [b]standardized coefficients.

Table 9

HOPE Scale Descriptive Statistics for Secondary Native American Sample (n=1,095)

Factor	Item	Response Percentage						Mean	*SD*	*r* with total[a]	Alpha if removed[b]	Alpha
		1	2	3	4	5	6					
Academic	1	6	19	24	25	14	11	3.56	1.39	.86	.96	.97
	6	6	15	21	31	16	12	3.71	1.37	.90	.96	
	7	5	15	20	29	18	12	3.77	1.37	.91	.96	
	9	3	11	22	32	18	13	3.91	1.28	.92	.96	
	10	4	11	22	31	18	14	3.90	1.32	.91	.96	
	11	6	14	21	32	16	12	3.76	1.36	.86	.96	
Social	2	10	23	32	18	11	5	3.69	1.30	.81	.90	.93
	3	7	18	34	19	13	9	3.30	1.23	.83	.90	
	4	6	11	33	22	17	11	3.45	1.28	.85	.90	
	5	14	28	27	15	10	7	3.78	1.34	.75	.92	
	8	8	14	31	22	15	10	3.24	1.24	.77	.91	

Note. [a]standardized correlations. [b]standardized coefficients.

Table 10

Concurrent and Predictive Correlations Among HOPE Scale and Achievement Scores

Subscale	HOPE Academic	HOPE Social	MSAT Reading[1]	MSAT Math[1]
HOPE Academic	-			
HOPE Social	.87 (519)	-		
MSAT 08 Reading	.54 (176)	.51 (177)	-	
MSAT 09 Reading	.55 (265)	.49 (268)	-	
MSAT 08 Math	.56 (179)	.54 (180)	.82 (276)	-
MSAT 09 Math	.55 (266)	.51 (269)	.83 (456)	-
MSAT 08 Science	.57 (82)	.50 (82)	.84 (84)	.85 (85)
MSAT 09 Science	.54 (90)	.46 (91)	.86 (183)	.84 (184)

Note. Sample sizes are in parentheses next to each correlation coefficient; [1]Year of MSAT scores is the same as that in the horizontal row of the cell; Adapted from "Additional validity evidence and across-group equivalency of the HOPE Teacher Rating Scale," by S. J. Peters and M. Gentry, 2012, *Gifted Child Quarterly, 57,* p. 96. Copyright 2012 by National Association for Gifted Children. Adapted with permission.

content areas for the Scales for Rating the Behavioral Characteristics of Superior Students, and Ryser and McConnell (2004) reported correlations between the ratings of teachers and parents ranging from .43 to .60, so the estimates of inter-rater reliability for the HOPE Scale are comparable to those of the Renzulli et al. scales. It is also important to note that Pereira (2011) found that ESL teachers rated students significantly higher than classroom teachers. The implication of this finding for those seeking to identify underserved populations, like ELLs, is that they should include HOPE Scale nominations from specialized teachers who may know the student better and understand the student in different contexts than a classroom teacher.

Measurement invariance testing. Analyses of measurement invariance have found that the HOPE Scale, when used with the development sample, was invariant between students from low-income families and others from non-low-income families, and across Caucasian, Black, and Hispanic students (Peters & Gentry, 2012); whereas noninvariance existed between English language learners and English proficient students (Pereira, 2011) and gender (Peters & Gentry, 2012). However, students who were eligible for the free or reduced lunch program received lower mean scores on both subscales than their peers who do not qualify for this program, with Cohen's *d* averaging .30, a small but not trivial effect. The difference between the mean scores for males and females had effect

sizes of *d*=.13 (Academic) and *d*=.36 (Social). The Caucasian group was used as the referent group for racial/ethnic group comparisons of means, and effect sizes measured by Cohen's *d* were trivial, ranging from .005 to .11. Because of these slightly lower scores, we recommend any test users make within-group comparisons based on local norms, which can be generated when the test is scored. This process is described in Appendix B. Tables 11, 12, and 13 include means and standard deviations for the various gender, income, and racial/ethnic groups from the normative development samples. For additional information on Developmental Research and Technical Data, please see Appendix C.

To date, the HOPE Scale has been subjected to rigorous development and validation studies using diverse samples of students, which facilitated investigation into how it can be used with subgroups including students from low-income families, students who are Black, Hispanic, and Native American, as well as those who are learning English. In its development, underserved students were oversampled, allowing for testing and generalizations to these populations. Invariance existed, meaning that across subgroups as reported, the HOPE Scale measures the same latent construct, although the mean scores are not equal across these subgroups. Appendix D includes summaries of research studies focusing on the HOPE Scale to date. Using local norms and comparing students within subgroups can help achieve equity in identification. Equally important is developing equitable programming and a continuum of services that serve to develop potential once it has been identified. It is important to identify, recruit, serve, and retain students in gifted education programming and to provide a continuum of services (Gentry, 2009).

Table 11

Descriptive Statistics, Mean Differences, and Effect Sizes for the Male and Female Groups on Academic and Social Subscales

HOPE Scale Subscale	Male (*n*=553)		Female (*n* =872)		*Mdiff* [a]	*d* [b]
	M	*SD*	*M*	*SD*		
Academic	18.81	7.90	19.86	7.79	1.05	.13
Social	16.60	5.80	18.72	6.04	2.12	.36

Note. [a]Difference between means of male and female students. [b]Cohen's *d*.

Table 12

Descriptive Statistics, Mean Differences, and Effect Sizes for the FARM and Non-FARM Groups on Academic and Social Subscales

HOPE Scale Subscale	Non-FARM (*n*=1027)		FARM (*n*=515)		*Mdiff* [a]	*d* [b]
	M	*SD*	*M*	*SD*		
Academic	20.25	8.11	17.53	7.00	2.12	.36
Social	18.22	6.07	16.43	5.72	1.79	.30

Note. [a] Difference between means of FARM and Non-FARM students. [b]Cohen's *d*.

Table 13

Descriptive Statistics for Racial/Ethnic Groups on Academic and Social Subscales

HOPE Scale Subscale	Asian (*n* =164)		African American (*n*=217)		Hispanic (*n* =228)		Caucasian (*n* =892)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Academic	19.27	7.70	18.64	7.34	19.07	7.05	19.52	8.22
Mdiff [a]	0.25		0.88		0.45			
d [b]	0.03		0.11		0.06			
Social	17.63	6.27	17.24	5.98	17.78	5.34	17.66	6.11
Mdiff [c]	0.03		0.42		-0.12			
d [d]	0.005		0.07		0.02			

Note. [a] Difference between means for each racial/ethnic group in the Academic subscale and the mean for Caucasian students in the Academic subscale. [b]Cohen's *d* for Academic subscale. [c]Difference between means for each racial/ethnic group in the Social subscale and the mean for Caucasian students in the Social subscale. [d]Cohen's *d* for Social subscale.