


Consistency of the Performance and Nonperformance Methods in Gifted Identification: A Multilevel Meta-Analytic Review

Gifted Child Quarterly
1–21
© 2016 National Association for
Gifted Children
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0016986216634438
gcq.sagepub.com


Selcuk Acar¹, Sedat Sen², and Nur Cayirdag³

Abstract

Current approaches to gifted identification suggest collecting multiple sources of evidence. Some gifted identification guidelines allow for the interchangeable use of *performance* and *nonperformance* identification methods. This multiple criteria approach lacks a strong overlap between the assessment tools; however, interchangeable use of the instruments (replacing one for another) entails high regularity. This meta-analytic review investigated the consistency of using performance and nonperformance identification methods by examining the influence of three moderators in two different study analyses. Study 1 focused on correlational and comparison studies by using Pearson r as the index of effect size within a three-level multilevel design. Study 2 was conducted with three diagnostic proportional metrics: efficiency, effectiveness/sensitivity, and specificity. Results from Study 1 indicated the overall correlation between the performance and nonperformance gifted identification methods was medium ($r = .30$). Teacher ratings yielded significantly higher consistency with performance measures than teacher or parent nomination and self-ratings. Study 2 showed that nonperformance methods are relatively strong in terms of specificity (70%) and effectiveness/sensitivity (59%) but not very efficient (39%). Analyses of four diagnostic quadrants indicated that performance and nonperformance gifted identification methods, when used alone, tend to identify different students who would not be identified otherwise despite some amount of convergence between the two. Our findings indicated that nonperformance and performance gifted identification methods cannot replace each other. They should be used concurrently rather than be used alone or consecutively.

Keywords

gifted identification, performance methods, nonperformance methods, efficiency, effectiveness, sensitivity, specificity

Identification of gifted and talented students has been one of the most controversial issues in the field (Heller, 2004; Kirschenbaum, 1983; Renzulli & Reis, 2004). Research on gifted identification often starts with basic questions such as the definition and conceptualization of giftedness (e.g., Renzulli, 1990; Renzulli & Reis, 2004; Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996; Worrell & Erwin, 2011). Identification practices are operational responses to the crucial questions about the nature and education of giftedness. More important, identification procedures result in high-stakes decisions such as the inclusion and/or exclusion of students into specialized programs. Therefore, decisions and strategies of gifted identification must rely on empirical evidence.

Identification procedures have evolved from traditional approaches focused solely on intelligence quotient (IQ) or achievement tests to contemporary approaches involving a wide variety of instruments and strategies such as portfolios, teacher, parent, and self-ratings and nominations, and other authentic assessments used together or in combination (Alvino, McDonnell, & Richert, 1981; Brown et al., 2005;

Coleman & Gallagher, 1995; Evans, 1996; Pfeiffer, 2003). The progressive move toward multiple sources of identification evidence mitigates some of the criticism against the traditional methods such as absence of an agreed-on cutoff score (Kirschenbaum, 1983) and concerns on the fairness of the standardized tests (Frasier, García, & Passow, 1995). This extension is also consistent with the definitions of giftedness that go beyond IQ (Marland, 1972; National Association for Gifted Children, 2009, 2013, 2015) and are inclusive of components of giftedness such as leadership (Haroutounian, 1995).

¹State University of New York at Buffalo, Buffalo, NY, USA

²Harran University, Sanliurfa, Turkey

³Istanbul Sabahattin Zaim University, Istanbul, Turkey

Corresponding Author:

Selcuk Acar, International Center for Studies in Creativity, State University of New York at Buffalo, 1300 Elmwood Ave, Buffalo, NY 14222, USA.

Email: acars@buffalostate.edu

Broadly speaking, there are two major forms of gifted identification methods: performance and nonperformance. Early performance methods relies on the use of tests of intelligence (Martinson, 1974; Pagnato & Birch, 1959; Sternberg, 1986). While research has shown high IQ equaling giftedness to be a myth (J. H. Borland, 2009), IQ tests are among the most frequently used instruments for gifted identification (McClain & Pfeiffer, 2012; National Association for Gifted Children, 2015). As a result of more complex conceptualizations of giftedness such as The Three-ring Conception of Giftedness that includes creativity and motivation (Renzulli, 1978, 2011), tests of achievement, aptitude, and creativity have become part of the gifted identification criteria (Archambault et al., 1993; Council of State Directors of Programs for the Gifted, 1999; Georgia Department of Education, 2014; Hunsaker & Callahan, 1995; Torrance, 1984).

A typical characteristic of the performance approach to gifted identification is that respondents obtain a score from the assessments that is indicative of their *performance* on the required tasks without any involvement of teachers, parents, or others. There is little or no room for the judgments of scores besides the actual performance of the respondent. Opinions, observations, or anecdotal evidence by teachers or parents for or against students do not count toward students' eligibility for identification. The only source of data is the students' performance on the required tasks. Therefore, this approach can be called "performance-based identification methods" or "performance methods."

In spite of their capacity to predict academic achievement and job performance (Neisser et al., 1996), the use of performance measures (e.g., IQ tests) as the only indicator of giftedness is problematic because expected behavioral outcomes such as achievement cannot only be attributed to IQ, which research as shown accounts for about 50% of the variability at best (Anderson & Keith, 1997). More important, such measures can limit the access to gifted education services when not accompanied by alternative approaches (VanTassel-Baska, Feng, & Evans, 2007).

The second category (i.e., nonperformance method) represents all other gifted identification approaches not involving any performance-based assessment. Examples of forms of assessment in this category are teacher rating scales, teacher and parent nominations, self-ratings, and peer-ratings. A common characteristic of the nonperformance method is that judgments, instead of test scores, are highly involved in the assessment. If an instrument (e.g., rating scale) is used, the scores are obtained through judgments of self, peers, parents, or teachers. This method falls under the "alternative assessment" (see Pfeiffer, 2015), but this category will be called "nonperformance methods."

The 2014-2015 State of the States Report (National Association for Gifted Children, 2015) provided information about the current practices on identification strategies. Multiple criteria (Frasier, 1997; Friedman-Nimz, 2009;

National Association for Gifted Children, 2015) was the most common identification model (19 states) and IQ tests were the most common instruments (19 states), along with the use of achievement data (19 states), and followed by nomination (12 states), various state-approved assessments (9 states), and portfolios (8 states). Multiple criteria not only lead to assessment beyond IQ testing but also allowed the inclusion of alternative assessments such as teacher and parent nominations and teacher rating scales (Pfeiffer, 2015).

The trends in identification toward nonperformance gifted identification methods led to the development of many rating scales (Jarosewich, Pfeiffer, & Morris, 2002) such as Gifted Rating Scales (GRS; Pfeiffer & Jarosewich, 2003), Renzulli-Hartmann Scale for Rating Behavioral Characteristics of Superior Students (Renzulli & Hartman, 1971), and Gifted and Talented Evaluation Scales (Gilliam & Jerman, 2015). The inclusion of these types of rating scales is built on the idea that teachers and parents can add meaningful information to the gifted identification process. Teacher recommendations have become a common practice for the identification of the gifted (Davis & Rimm, 2003). Gentry and Mann (2008) suggested that teachers could identify high-ability students who do not perform well on standardized tests because they observe their students in various areas over time rather than utilizing the snapshot perspective obtained from standardized tests. Rating scales and checklists provide additional information that may not be captured by the standardized tests (Chan, 2000; Hoge & Cudmore, 1986; Peterson, 1999; Pfeiffer & Blei, 2008). J. H. Borland and Wright (1994) underlined the importance of using alternative evidence such as observation, dynamic assessment, and the best performance rather than standardized performance tests especially for the identification of gifted students from economically disadvantaged and underrepresented groups.

However, alternative methods such as teacher and parent nomination and the use of ratings may be biased against gifted students with specific personality traits (Rohrer, 1995) and those who are culturally diverse (Kaufman & Harrison, 1986). In addition, some of the alternative methods (e.g., teacher or parent nomination) seem to suffer from a lack of scientific rigor (Pfeiffer & Blei, 2008) in that they categorize students as gifted candidates or not rely on teachers' personal conceptualization of giftedness, which may be limited (Lee, 1999; Moon & Brighton, 2008; Neumeister, Adams, Pierce, Cassidy, & Dixon, 2007). Furthermore, teachers could also favor students who conform to their values, expectations, rules, and instructions (George, 1979), have unusual interests (Siegle, 2001), love to read (Siegle, 2001; Siegle & Powell, 2004), or exhibit socially desirable behaviors (e.g., altruism) that are not necessarily a giftedness characteristic (Siegle, Moore, Mann, & Wilson, 2010). Teachers may also hold biases against girls (Bianco, Harris, Garrison-Wade, & Leech, 2011) and certain ethnic groups (Elhoweris, Mutua,

Alsheikh, & Holloway, 2005) as the accuracy of the nominations tends to be lower for Hispanic and African American students and those students from low socioeconomic status (Alvidrez & Weinstein, 1999; Elhoweris, 2008; Masten & Plata, 2000; Masten, Plata, Wenglar, & Thedford, 1999; McBee, 2006, 2010). Overall, if teachers are not educated on gifted characteristics they may hold unrealistic expectations from gifted students (Pfeiffer, 2002).

The Current Study

Both performance and nonperformance gifted identification methods have their strengths and weaknesses and choosing one over another would be questionable in terms of identification outcomes (Pfeiffer & Blei, 2008; Worrell, 2003, 2009). The multiple criteria approach to gifted identification is more pluralist and more inclusive of diverse abilities and backgrounds (Ford & Grantham, 2003; Richert, 1987; Tannenbaum, 2003). The concurrent use of various instruments or methods is consistent with the multiple criteria approach if the different assessments have a weak or moderate correlation (Bélanger & Gagné, 2006).

In some states, school districts, and countries, performance and nonperformance gifted education methods are used interchangeably. For example, alternative methods falling under the nonperformance method heading are used for gifted identification in some states as a substitute for standardized performance tests (e.g., Georgia Department of Education, 2014; Krisel & Brown, 1997; Tennessee Department of Education, 2010). The interchangeable use of the performance and nonperformance gifted identification methods assumes that they largely converge on the same identification outcomes. Is that really the case? Can performance and nonperformance gifted identification methods be used interchangeably? Can or should they replace one another? A very high correlation between the two methods or higher values of diagnostic accuracy (e.g., sensitivity, specificity, and efficiency) such as .70 or higher (Swets, 1988) can justify this practice.

The availability of multiple identification instruments poses a challenge to practitioners in choosing the optimum combination of instruments and strategies for the identification of the gifted. McBee, Peters, and Waterman (2014) tested the usefulness of different combinations of identification strategies and provided useful tips. However, a deliberate and research-based approach to determining the best gifted identification strategies would require a comprehensive review of the literature analyzed systematically. Previous literature reviews (e.g., Hoge & Cudmore, 1986) on gifted identification strategies are dated and lack a systematic analysis. The recent studies are qualitative in nature (e.g., Worrell & Erwin, 2011) and quantitative syntheses are highly needed. This is the first meta-analytic study comparing various gifted identification methods and the degree to which they overlap.

A crucial advantage of systematic reviews is the consideration of the impact of potential moderators. The eligible studies within this meta-analytic study included three major moderators: *grade level of the students* who have gone through the identification process (i.e., kindergarten, elementary, middle, and high school levels), *type of nonperformance gifted identification methods* (e.g., self-ratings, teacher ratings, teacher nomination, and parent nomination), and *type of performance gifted identification methods* (e.g., tests of intelligence, tests of aptitude and achievement, and tests of creativity). Other potential moderators such as gender, target group (dominant vs. cultural or economically diverse groups), and type of performance test (verbal vs. nonverbal) were not included within this meta-analysis due to lack of sufficient studies providing information on these variables.

Comparison of the grade levels is important because researchers often site difficulties in the identification of younger students (Fatouros, 1986; Johnson, 1983; Shaklee, 1992). We also compared the consistency with the performance methods when self-ratings, teacher nominations, and teacher ratings, and teacher nomination were used. This way, we would be able to know which of the nonperformance methods deserves more attention for the practitioners. Finally, consistency was compared for intelligence, achievement, and creativity tests. We expected consistency to be lower for creativity because teachers' conceptions of creativity are questionable (Aljughaiman & Mowrer-Reynolds, 2005; Kampylis, Berki, & Saariluoma, 2009) and typical characteristics of the creative individuals can be seen as unacceptable in a classroom (see Westby & Dawson, 1995) and may overshadow the students' strengths.

The current study investigated the consistency of performance and nonperformance identification methods of giftedness in two different studies. Study 1 focused on studies reporting correlation values as effect sizes to investigate the relationship between performance and nonperformance methods. Study 2 focused on studies that provided values of diagnostic accuracy between performance and nonperformance assessment methods in order to clarify if one method could be substituted for the other.

Method

Study Variables

The current study examined the degree of consistency between performance and nonperformance methods that are typically used in gifted identification. The performance methods were comprised of tests of general ability and intelligence, tests of aptitude and achievement, and tests of creativity. The nonperformance methods consisted of rating scales, teacher and parent nominations, and self-nomination or self-reports. A list of instruments used in the study is provided in Table 1.

Table 1. List of Performance and Nonperformance Measures and Studies Using Them.

Performance instruments	Studies using the listed scales
General ability	
1. California Test of Mental Maturity (Traxler, 1939)	Alexander (1953)
2. Culture Fair Intelligence Test Scale I (Weiss & Osterland, 1997)	Spinath and Spinath (2005)
3. IPAT Culture Fair Intelligence Test (Cattell, 1958)	Chambers, Barron, and Sprecher (1980)
4. Leiter International Performance Scale (Leiter, 1952)	Ryan (1983)
5. Lorge-Thorndike Test (Lorge & Thorndike, 1962)	Cornish (1968)
6. Naglieri Nonverbal Ability Test (NNAT) (Naglieri, 1991)	Bracken and Brown (2008)
7. Otis-Lennon Group Test (Otis & Lennon, 1967)	Ashman and Vukelich (1983)
8. Raven's Advanced Progressive Matrices (RAPM; Jordanian version) (Allyan & Smadi, 1988)	Subhi (1997)
9. Shipley Institute of Living Scale (Zachary, 1991)	Carman (2011)
10. Slosson Intelligence Test (Slosson, 1963)	Ciha, Harris, Hoffman, and Potter (1974), Hartsough, Elias, and Wheeler (1983), Rust and Lose (1971)
11. Stanford-Binet (Terman & Merrill, 1937, 1960)	Kirk (1966), Pegnato and Birch (1959), Silverman, Chitwood, and Waters (1986)
12. Stanford-Binet, Form L-M (Terman & Merrill 1962)	J. W. Baldwin (1962), Cornish (1968), Ryan (1983)
13. Wechsler Intelligence Scale for Children (WISC-R) (Wechsler, 1949, 1974)	Ciha et al. (1974), Hunter and Lowe (1978), Lowenstein (1982) Rust and Lose (1971)
14. Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV) (Wechsler, 2003)	Pfeiffer and Jarosewich (2007)
15. Wechsler Preschool and Primary Scale of Intelligence (WPPSI) (Wechsler, 1967)	Jacobs (1971)
Achievement and aptitude	
1. American College Testing (ACT)	Lee and Olszewski-Kubilius (2006)
2. Bracken Basic Concept Scale—Revised (BBCS-R) (Bracken, 1998)	Bracken and Brown (2006, 2008)
3. California Achievement Test (CAT)	Ashman and Vukelich (1983), Gallagher (1985)
4. Cognitive Ability Test (CAT) (Heller, 2000; Heller, Gaedike, & Weinläder, 1987; Heller & Perleth, 2000)	Neber (2004)
5. Iowa Tests of Basic Skills (ITBS)	Hunter and Lowe (1978)
6. Mathematical Skills Assessment (Ministry of Education & United Nations Relief and Works Agency, 1990)	Subhi (1997)
7. Scholastic Aptitude Test (SAT)	Lee and Olszewski-Kubilius (2006)
8. Stanford Achievement Test (SAT) (Madden, Gardner, Rudman, Karlsen, & Merwin, 1973)	Elliott, Argulewicz, and Turco (1986)
9. SRA Primary Mental Abilities Test (Thurstone & Thurstone, 1949)	Cornish (1968), C. D. Wilson (1963)
10. Woodcock-Johnson Psycho-Educational Battery—Revised: Tests of Achievement (WJ-RACH) (Woodcock & Mather, 1990)	Crosby and French (2002)
Creativity	
1. Barron Symbol Equivalence Test (Barron, 1976a, 1976b)	Chambers et al. (1980)
2. Divergent Thinking (Wallach & Kogan, 1965; Ward, 1968)	Chambers et al. (1980), Renzulli, Hartman, and Callahan (1971), Harrington, Block, and Block (1983)
3. Imaginative Writing (Dewing, 1970)	Dewing (1970)
4. Torrance Tests of Creative Thinking (Torrance, 1962, 1974)	Gallagher (1985), Swenson (1978)
Nonperformance Instruments	
Rating scales	
1. California Child Q-Set (Block & Block, 1981)	Harrington et al. (1983)
2. Children's ability self-perceptions (Spinath & Spinath, 2005)	Spinath and Spinath (2005)
3. Classroom Performance Profile (CPP) (Crosby & French, 2002)	Crosby and French (2002)
4. Clinical Assessment of Behavior—Teachers (CAB-T) (Bracken & Keith, 2004)	Bracken and Brown (2008)
5. Creative Behavior Scale (Swenson, 1978)	Swenson (1978)
6. Gifted Rating Scales—Preschool/Kindergarten Form (Pfeiffer & Jarosewich, 2003)	Pfeiffer and Petscher (2008)

(continued)

Table 1. (continued)

Nonperformance Instruments	Studies using the listed scales
7. Gifted Rating Scales—School Form (Pfeiffer & Jarosewich, 2003)	Pfeiffer and Jarosewich (2007)
8. Gifted and Talented Evaluation Scales (Gilliam, Carpenter, & Christensen, 1996; Gilliam & Jerman, 2015)	Gilliam et al. (1996), Gilliam and Jerman (2015)
9. Golan Creative Motivation Scale	Dewing (1970)
10. Overexcitability Questionnaire (Lysy & Piechowski, 1983)	Gallagher (1985)
11. Overexcitability Questionnaire II (OEQII) (Bouchet & Falk, 2001)	Carman (2011)
12. Parents' Perceptions of Child Ability—Derived from Absolute School-Based Ability Perceptions (Stiensmeier-Pelster, Spinath, Schone, & Dickhauser, 2002)	Spinath and Spinath (2005)
13. Renzulli-Hartmann Scale for Rating for Rating Behavioral Characteristics of Superior Students (SRBCSS) (Renzulli & Hartman, 1971)	Ashman and Vukelich (1983), Elliott et al. (1986), Hunter and Lowe (1978), Renzulli, Hartman, and Callahan (1971), Rust and Lose (1980), Subhi (1997)
14. Self-efficacy scale (subscale of Motivated Strategies for Learning Questionnaire [MSLQ]) (Garcia & Pintrich, 1996)	Neber (2004)
15. Teacher rating scale (Neber & Heller, 1995)	Neber (2004)
16. Torrance Creative Leisure Interests Checklist	Dewing (1970)

Data Sources and Search Strategies

Combinations of several keywords were used to locate the articles such as “gifted identification,” “parent nomination,” “teacher nomination,” “teacher judgments,” “parent judgments,” “rating scales,” “performance measures,” “subjective methods,” and “objective methods.” The initial search was conducted through Academic Search Complete, EBSCOhost EJS, Psychological & Behavioral Sciences Collection, PsychARTICLES, and Google Scholar. A secondary search was conducted by a review of the references of the relevant articles located in the initial search. Last, the test manuals of the instruments used in gifted identification were reviewed.

Inclusion and Exclusion Criteria

Studies were included in this research if they met the following criteria:

1. Studies using only quantitative methods of analyses. Journal articles, thesis and dissertations, and test manuals were part of the search process.
2. Articles must include at least one type of performance and nonperformance method for the identification of the gifted.
3. The relationship between the two methods must be reported through necessary statistics that allow calculation of effect size (i.e., mean and standard deviation, Pearson r , t , and F values along with degrees of freedom) or one of the three diagnostic criteria of efficiency, effectiveness/sensitivity, and/or specificity.
4. Studies published in English after 1950.

After excluding the qualitative works from the pool of potential studies, each study was reviewed for eligibility

based on the other criteria. Out of 83 quantitative studies, 32 studies did not have a research design that provided statistics between a performance and nonperformance method. Sixteen others failed to provide necessary statistics to obtain an index of effect size. The final data set included 35 studies. The articles used for this meta-analysis are marked with an asterisk in the reference list.

Data Coding

Thirty-five articles were coded by the first author through iterative process. First, a worksheet was developed by considering variables and moderators, which included grade level (i.e., kindergarten and elementary, middle school, elementary and middle school, and high school), type of non-performance methods (i.e., self-report, teacher ratings, parental evaluation or rating, teacher nomination, and other), and type of performance methods (i.e., tests of intelligence or IQ, achievement, and creativity). After the completion of the initial coding, all levels of variables and moderators were coded numerically. Some categories were merged to obtain a larger sample size (number of effect sizes) and the final data set included the following categories for Study 1:

Grade level: Younger students (kindergarten and elementary) and older students (middle school and above)

Type of nonperformance methods: Self-reports, teacher rating scales, and all other methods

Type of performance methods: Tests of intelligence or IQ, achievement or aptitude, and creativity

Study 2 had further limitations in terms of the number of effect sizes of the moderator categories. Thus, efficiency and sensitivity/effectiveness analyses were conducted with the

variables of grade level and type of nonperformance methods. Specificity analyses had no moderators.

Rater Reliability

Reliability of the coding process was assessed by examining the coding of the data set. First and second authors independently coded a randomly selected section (25%) of the final set of studies. An agreement index was calculated between codings of the two authors and disagreements in coding were resolved through investigation of the original values reported in the article. A two-way (effects and raters) intraclass correlation coefficient (ICC) for absolute agreement was .92.

The Calculation of Effect Sizes

The studies reported different type of statistics leading to various types of effect sizes. To increase the number of effect sizes included in the synthesis, we grouped effect sizes in two categories and analyzed them under two different studies (i.e., Study 1 and Study 2). The first category included studies reporting the values of Pearson r , t test, and Cohen's d . Because of the nature of the research question (i.e., Are nonperformance and performance methods consistent?), t test and Cohen's d were converted to Pearson r in Study 1 by following the procedures described by Lipsey and Wilson (2001; see Equations 1 and 2). This operation generated 166 effect sizes, which were analyzed in Study 1. Borenstein, Hedges, Higgins, and Rothstein (2009) suggested using a Fisher transformed correlation (zr) instead of Pearson correlation (r) because Pearson's product-moment correlation (r) is not normally distributed. Thus, the Pearson correlation effect sizes were converted to Fisher's zr s to combine the effect sizes properly in Study 1 (Hedges & Olkin, 1985; Rosenthal, 1994). After performing analyses with z -transformed correlation, these resulting Fisher-based correlations were back-transformed to Pearson correlations using an inverse Fisher transformation to make the interpretation easier (Lipsey & Wilson, 2001).

Transformation from Cohen's d to r was made using following equation:

$$r = \frac{d}{\sqrt{d^2 + a}}, \quad (1)$$

where a is a correction factor which is calculated as $(n_1 + n_2)^2 / n_1 n_2$. We converted from a t -test statistics value to an r value using

$$r = \sqrt{\frac{t^2}{t^2 + df}}, \quad (2)$$

where df represents degrees of freedom.

The articles used in Study 2 reported diagnostic statistics as proportions. The studies reported four diagnostic statistics: effectiveness, efficiency, sensitivity, and specificity. The formula for these four statistics is provided in Figure 1.

		Performance Methods	
		Gifted	Not gifted
Non-performance Methods	Positive	True positives (TP)	False positives (FP)
	Negative	False negatives (FN)	True negatives (TN)

Effectiveness = TP/(TP+FN) Sensitivity = TP/(TP+FN)
 Efficiency = TP/(TP+FP) Specificity = TN/(TN+FP)

Figure 1. Diagnostic accuracy applied to gifted identification.

Effectiveness and sensitivity were calculated via the same formula and refers to the proportion of true positives to the sum of true positives and false negatives. In the context of this study, effectiveness and sensitivity reflect the percentage of accurately identified gifted students to the total number of all gifted students (those who were and were not identified as gifted). Efficiency refers to the proportion of all true positives to the sum of true and false positives. For this study, it is the percentage of accurately identified gifted students to the total number of all students who were nominated or considered gifted. Specificity refers to the proportion of true negatives to the sum of true negatives and false positives. In this study, it is the percentage of nongifted students who were neither nominated for gifted nor considered gifted to the total number of all nongifted students. Effectiveness and sensitivity are the same, so only three different proportional indices were used in the analyses.

We applied the diagnostic accuracy framework to the field of gifted identification where nonperformance methods are compared with performance-based methods. It should also be noted that there was no single cutoff score for the performance methods used in gifted identification (Kirschenbaum, 1983). Use of performance methods as criteria, of course, does not mean that they are the best method for the gifted identification. The main reason for using performance-based methods as a criterion for gifted identification is its historical priority to other methods although these methods are widely used currently. Furthermore, alternative nonperformance methods are sometimes presented as alternatives to performance methods for gifted identification.

Three different data sets were prepared: (a) effectiveness/sensitivity, (b) efficiency, and (c) specificity in Study 2 because the three proportional values represent three distinct approaches to diagnostic statistics. Lipsey and Wilson (2001) indicated that proportions are a form or metric of effect sizes. Therefore, three separate meta-analyses were performed in Study 2 following the appropriate procedures for meta-analysis described by Lipsey and Wilson (2001). According to Lipsey and Wilson (2001), meta-analysis of research findings with proportions can be accomplished either using

proportions directly or using their logit transformation. They recommended using the logit method if the variation around the mean proportion is important. Based on this recommendation, we used the logit method in Study 2. The following calculations were used to obtain logit effect size (ES_l), standard error (SE_l), and variance weight (w_l):

$$ES_l = \log_e \left[\frac{p}{1-p} \right], \quad (3)$$

$$SE_l = \sqrt{\frac{1}{np} + \frac{1}{n(1-p)}}, \quad (4)$$

$$w_l = \frac{1}{SE^2} = np(1-p), \quad (5)$$

where n is the total number of the subjects and p is the proportion of subjects in the category of interest (efficiency, sensitivity/effectiveness, and specificity).

The Assessment of Potential Publication Bias

A number of statistical procedures have been proposed for assessment of publication bias in meta-analyses (Begg & Mazumdar, 1994; Duval & Tweedie, 2000; Egger, Smith, Schneider, & Minder, 1997; Light & Pillemer, 1984; Rosenthal, 1979; Rothstein, Sutton, & Borenstein, 2005). A funnel plot (Light & Pillemer, 1984) was created and examined visually to see if there was a possible publication bias effect in the current data sets. A funnel-shaped distribution of a scatterplot between treatment effect and size of study indicates a lack of publication bias. In addition to visual examination of the funnel plot, the test proposed by Egger et al. (1997) was used to test for funnel plot asymmetry. Egger's regression test was used to examine the small-study effect in this meta-analysis. In addition, Begg's rank correlation method (Begg & Mazumdar, 1994) was also used to test for funnel-plot asymmetry.

Statistical Analyses

Meta-analytic studies can be conducted using fixed or random effects models based on heterogeneity in the data. Thus, the heterogeneity of the effect sizes should be determined before conducting meta-analysis. Independence of the effect sizes is another issue that should be examined before meta-analyses. Dependent effect sizes may be observed when some studies contain multiple effect sizes. As this was the case, traditional meta-analysis approaches were not appropriate. Alternative approaches such as multilevel meta-analysis were suggested as a solution to the problem of dependency (Cheung, 2014; Hox, 2002; Konstantopoulos, 2011; Scammacca, Roberts, & Stuebing, 2014; Stevens & Taylor, 2009; Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013).

These approaches were proposed to handle various kinds of dependency including dependence over studies, dependence due to use of multiple treatment groups with a control group in the same study, and dependence due to multiple effect sizes within the same study. A three-level multilevel meta-analysis was applied due to the use of multiple effect sizes from the same study in our meta-analytic data.

We followed procedures described by Van den Noortgate, López-López, Marín-Martínez, and Sánchez-Meca (2014) to handle dependency among the effect sizes. In this meta-analysis, the first level involves a within-effect size model, the second level a between effect size within-publication model, and the third level a between-publication model. The data are summarized in Table 2. The parameterization of the three-level model follows (see also Konstantopoulos, 2011).

First level (i.e., within-effect size) can be represented with Equation 6, in which T_{ig} is independently and normally distributed with the mean of π_{ig} and variance of v_i , which is assumed to be known (Hox, 2002),

$$T_{ig} = \pi_{ig} + e_{ig}, \quad (6)$$

In the second level (the between-effect size within-publication model) the unknown effect-size parameter π varies around a Level 3 unit g mean, namely,

$$\pi_{ig} = \beta_{0g} + r_{0g}, \quad (7)$$

where $g = 1, \dots, n$ represents the Level 3 units (e.g., publications). Finally, at the third level the Level 3 unit means vary around an overall mean γ_{00} ,

$$\beta_{0g} = \gamma_{00} + u_{0g}, \quad (8)$$

where u_{0g} is normally distributed with a mean of zero and between publication variance. The three-level model can be represented in the single following model (this single model equation was also presented in Van den Noortgate et al., 2014)

$$Y_{ig} = \gamma_{00} + u_{0g} + r_{0g} + e_{ig}, \quad (9)$$

when there is any additional predictor, the model can be extended adding estimates for each covariate as follows:

$$\pi_{ig} = \beta_{0g} + \beta_{1g}X_{1ig} + \dots + \beta_{pg}X_{pig} + r_{0g}. \quad (10)$$

When p predictors are included at the second level, the model is

$$\pi_{ig} = \beta_{0g} + \beta_{1g}X_{1ig} + \dots + \beta_{pg}X_{pig} + r_{0g}, \quad (\text{Equation 11})$$

where X_{1ig}, \dots, X_{pig} are study-specific covariates (e.g., Grade level), $\beta_{0g}, \beta_{1g}, \dots, \beta_{pg}$ are unknown regression coefficients that need to be estimated (Konstantopoulos, 2011). In this study, the data set has three second-level variables that can be examined in subgroups: (a) grade level (kindergarten and

Table 2. Descriptive Statistics Descriptive Statistics of Transformed Effect Sizes (ES).

Study ID	Number of ES	N	Mean ES (<i>r</i>)	SD	Minimum	Maximum
1	10	183	.358	.054	.236	.420
2	12	24	.062	.116	0.00	.321
3	10	20	.694	.143	.489	.902
4	3	60	.216	.258	-.030	.485
5	8	132	.016	.142	-.210	.185
6	2	13,074	.130	.007	.125	.135
7	4	416	.292	.140	.161	.472
8	1	90	.080	—	.080	.080
9	11	40	.422	.141	.245	.709
10	9	61	.644	.151	.436	.929
11	4	394	.246	.119	.151	.414
12	14	308	.512	.096	.343	.678
13	6	330	.243	.038	.192	.288
14	26	245	.080	.203	-.299	.758
15	2	90	.548	.025	.531	.566
16	31	217	.069	.279	-.419	.523
17	2	318	.410	.142	.310	.510
18	6	34	.305	.227	.070	.604
19	5	53	.443	.146	.255	.604

elementary vs. middle school or higher), (b) nonperformance methods (self-report, teacher ratings, and other methods), and (c) performance methods (tests of intelligence or IQ, achievement, and creativity). These variables can be used as Level 2 covariates to model the heterogeneity. The variables were entered to the multilevel model as second level covariates when a sufficient number of effect sizes were available. Therefore, Study 1 and Study 2 had differing numbers of covariates involved.

The Homogeneity Test

We calculated Cochran's heterogeneity statistic (Q) to estimate the homogeneity of the effect size distribution used in the traditional meta-analysis. As Cheung (2014) suggested, this test can also be used for multilevel meta-analysis studies. The Q test is simply a weighted sum of squared differences between the individual study effects and the pooled effect across studies. The Q test was distributed as a chi-square statistic with $N - 1$ (N : number of studies) degrees of freedom. A significant Q test indicates the presence of heterogeneity, in which the variability among effect sizes stems from between-study errors in addition to sampling error (Lipsey & Wilson, 2001). In addition, we also calculated the I^2 statistic (Higgins, Thompson, Deeks, & Altman, 2003), which quantifies the degree of heterogeneity in meta-analysis. The I^2 percentage ranges from 0 (no heterogeneity) to 100 (high-level heterogeneity; Higgins et al., 2003). In this study, an I^2 percentage was calculated using the values of Q statistics and the degrees of freedom as described in Higgins et al.'s (2003) work. I^2 is simply calculated by dividing the difference between Q statistics and the df by the Q statistics

itself. The obtained number is multiplied by 100 to be expressed in percentages.

Results

Study 1

Before proceeding to the main analyses, we examined the publication bias using the funnel plot (see Figure 2). The funnel plot was created with a scatterplot showing estimated treatment effects in individual studies against standard error of estimated treatment effects. As shown in Figure 2, the funnel is not visually symmetric. Slight deviations from funnel-shaped distribution can be observed in the upper part of the plot. We also fitted a regression model to investigate the small-study effects seen in this study (Rothstein et al., 2005). An Egger test was also performed to test for funnel plot asymmetry, $t(164) = 3.49, p = .001$. The funnel plot appears asymmetric, and there is evidence of publication bias using the Egger test but not using the Begg's rank correlation method ($p = .808$).

A total of 166 effect sizes (Pearson correlations: r_s) from 19 published studies were available for the analyses. Figure 3 shows a stem-and-leaf plot of the resulting 166 effect sizes (r_s) with two decimal places. As can be seen in Figure 3, the smallest effect size was $r = -.39$, the largest was $r = .73$. The descriptive statistics for mean effect size values of each publication are summarized in Table 2.

Homogeneity Test. According to MEAN ES analyses (using SPSS MEAN ES macro created by D. Wilson, 2001), the homogeneity-of-variance statistic (Q test) was 1,863.38 with

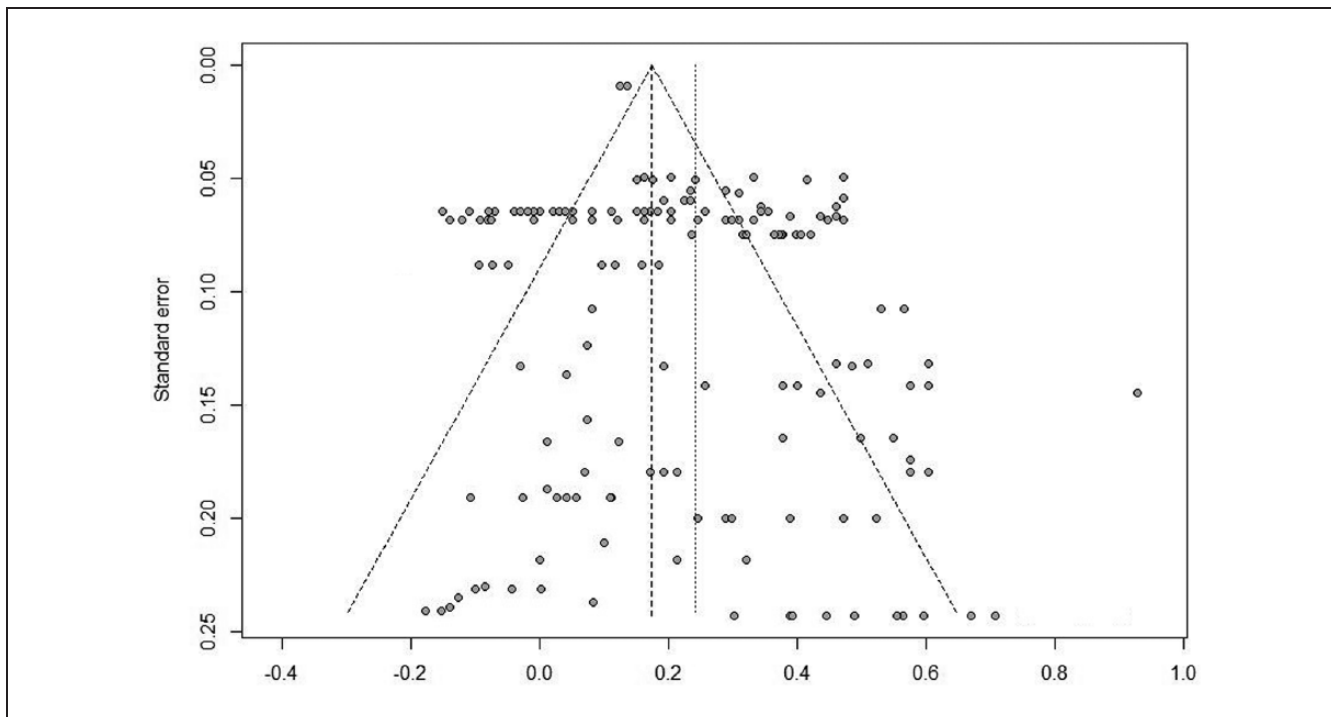


Figure 2. Funnel plot of standard error by Fisher's z.

Frequency	Stem & Leaf
2	-3 . 29
6	-2 . 046679
6	-1 . 123599
15	-0 . 012233447778899
20	0 . 0000000002345578889
19	1 . 1122355566677788999
24	2 . 000001112333344558888999
23	3 . 00011223344556667778899
21	4 . 112333344444556678889
20	5 . 00011222233344467899
8	6 . 01144489
2	7 . 13

Figure 3. Stem-and-leaf plot of all effect sizes (rs).

165 degrees of freedom. The *Q* test, which was found to be significant ($p < .001$), indicated the variance of the effect size was not homogeneous and larger than the effect of the standard error only. In addition, an I^2 percentage was calculated as 94.61, which also indicated a high-level of heterogeneity in this meta-analysis. Thus, we continued with a model that accounts for this heterogeneity (i.e., random effects model).

Multilevel Model Building. The multiple effect sizes used in the above results can be considered dependent because there were multiple effect sizes obtained from some of the studies. We used multilevel meta-analysis (Hox, 2002; Van den Noortgate et al., 2013) to handle the dependence in this

study. All of the multilevel models were estimated with restricted maximum likelihood (REML) using the proc SAS command. Results of the unconditional model were used to estimate an overall mean as a random effect, and the variances at the second and third levels. Preliminary analyses with the *Level 1 Variance Known model* (Bryk & Raudenbush, 1992) that had no predictors showed that the mean correlation was .302 (95% confidence interval [CI] = [.26, .35], $p < .0001$). This is a moderate correlation according to Cohen (1988). Second- and third-level variances were .033 (95% CI [.020, .046]) and .029 (95% CI [.029, .040]) respectively, and found to be significant ($p < .05$). Significant Level 2 and Level 3 variances indicated that the effect sizes varied across publications.

Following the multilevel research tradition, we continued with a three-level random-effects model. When three predictors were included in the unconditional model (i.e., main model) the overall effect size estimate was .159 (95% CI [.143, .174]), which was significant ($p < .001$). The results of the three-level main model indicated that only one of the moderators (i.e., type of nonperformance methods) explained a significant amount variation in effect sizes (see Table 3). Correlation values were significantly higher ($\beta = .31$, $SE = .05$, $p < .001$) between performance measures and teacher ratings ($r = .35$, 95% CI [.31, .38]) than with the other forms of nonperformance methods of identification ($r = -.01$, 95% CI [-.08, .05]) such as teacher or parent nomination and parent ratings. Correlations with self-reports were also smaller ($r = .12$, 95% CI [.05, .18]) than those with teacher ratings

Table 3. Parameter Estimates for Main Model.

	Estimates	SE
Fixed effects		
Intercept	0.159	0.008
Grade Level 0	-0.096	0.080
Nonperformance Type 0	0.049	0.078
Nonperformance Type I	0.309*	0.049
Performance Type 0	0.050	0.049
Performance Type I	0.031	0.054
Variance components		
Second level	0.021	0.011
Third level	0.027	0.003

Note: Grade Level 0 = younger students (kindergarten-elementary); Grade Level 1 = older students (middle or above); Nonperformance Type 0 = self-reports; Nonperformance Type I = teacher rating scales; Performance Type 0 = intelligence/general ability; Performance Type I = achievement/aptitude. * $p < .001$.

Table 4. Descriptive Statistics for the Moderator Categories.

Levels of moderators	k	Mean ES (<i>r</i>)	95%CI	Q
Grade level				
Kindergarten-elementary	104	.29	.24, .33	1157.12*
Middle or higher	52	.16	.12, .20	426.13*
Nonperformance methods				
Self-report	49	.12	.05, .18	356.06*
Teacher ratings	99	.35	.31, .38	594.98*
Parental nominations or ratings	18	-.01	-.08, .05	321.94*
Performance measures				
Intelligence or general ability	70	.23	.18, .29	676.14*
Achievement and aptitude	46	.30	.25, .34	691.98*
Creativity	50	.19	.10, .27	491.22*

Note. ES = effect size; CI = confidence interval. $dfs = k - 1$. * $p < .001$.

and self-reports, but this difference was not significant ($\beta = .01$, $SE = .08$, $p < .87$). Mean effect size values for each category are provided in Table 4.

Two other moderators (i.e., type of performance methods and grade level) were not significant although the mean effect size values of different categories were seemingly varied. For example, correlations with nonperformance methods were higher with achievement and aptitude tests ($r = .30$, 95% CI [.25, .36]) than intelligence or general ability tests ($r = .23$, 95% CI [.18, .29]) and creativity tests ($r = .19$, 95% CI [.10, .27]). Likewise, correlation values were higher among younger students (i.e., kindergarten-elementary; $r = .29$,

95% CI [.24, .33]) than older students (middle school or higher; $r = .16$, 95% CI [.12, .20]).

Study 2

Proportional effect sizes from diagnostic statistics were provided from 16 studies that reported 42 efficiency, 73 sensitivity/effectiveness, and 33 specificity values as proportions. We performed three different sets of multilevel meta-analyses with REML estimation using the proc SAS command because these were distinct qualities. Finally, descriptive statistics for the different levels of moderators were also provided.

As mentioned above, a funnel plot was created and examined for publication bias for each individual proportion metric (see Figure 4) before proceeding to the meta-analysis of the proportions. The Egger test (Egger et al., 1997) and Begg's rank correlation (Begg & Mazumdar, 1994) method were also applied as in Study 1 (see Figure 4 for the three funnel plots). We also fitted a regression model to investigate the small-study effects seen in each of the data sets through the Egger tests. The Egger tests were also performed to test for funnel plot asymmetry. Based on the Egger test results, efficiency funnel plot was found to be asymmetric, $t(40) = -4.40$, $p < .001$, while the effectiveness-sensitivity funnel plot, $t(71) = -1.00$, $p = .32$, and the specificity funnel plot, $t(31) = 1.85$, $p = .07$, were statistically symmetric. In addition, all of the funnel plots were found to be symmetric based on the Begg rank correlation method.

As suggested by Lipsey and Wilson (2001), analyses were performed with logit transformed values (see Equation 3) and then the mean values were back transformed to proportions in the text for ease of interpretation. The results of the three-level models were presented without transformation (as logit values) as the main purpose was to observe the moderators within the model.

Efficiency. A total of 42 effect sizes from 11 published studies were available for the analyses. Homogeneity statistics indicated the data set was heterogeneous, $Q_r(41) = 3318.23$, $p < .01$, $I^2 = 98.8\%$. Thus, we conducted a random effects model using the logit method for the efficiency proportions. As mentioned above, there were originally three moderators in this study. Two moderators (i.e., grade level and nonperformance method) had a sufficient number of effect sizes for efficiency and the third moderator had only one category, which is tests of intelligence, so these two moderators were dummy coded and entered into the random effects regression model.

To take dependency in account, we continued with a multilevel model. First, an unconditional multilevel meta-analysis was conducted with no predictors. The unconditional model results yielded a mean effect size value of 0.39 (95% CI = [0.24, 0.55], $p = .21$). Second- and third-level variances were 1.14 ($p = .03$) and 0.21 ($p = .001$), respectively. The results of the main three-level meta-analysis are presented in Table 5. The three-level model indicated that grade level ($\beta = -0.23$,

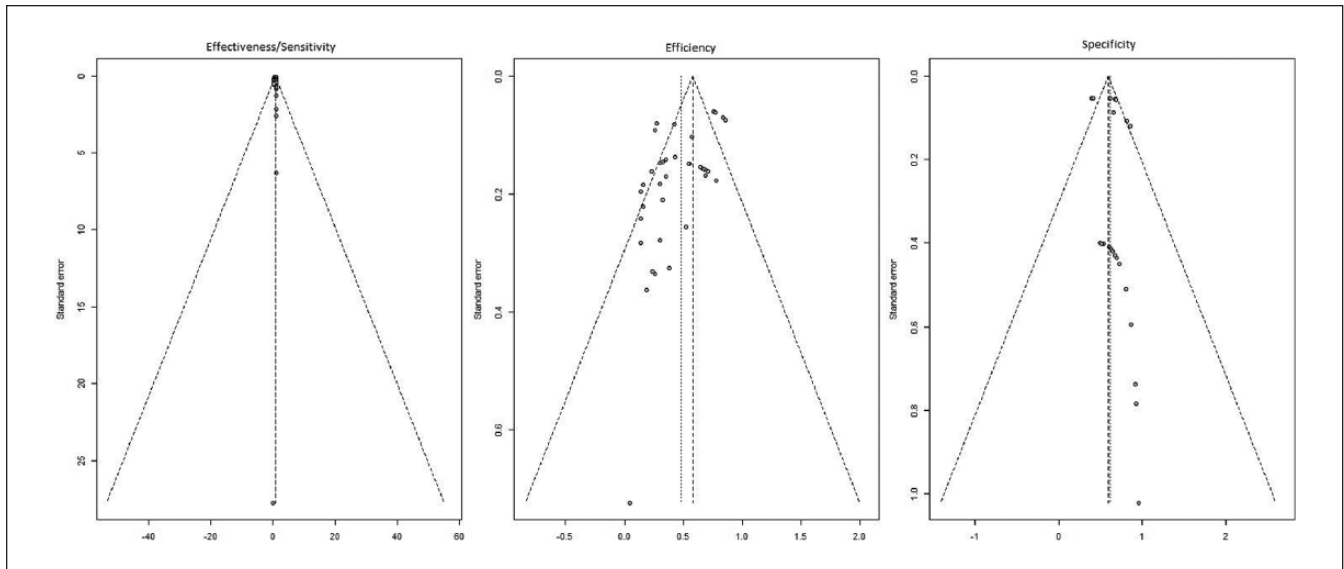


Figure 4. Funnel plots for Study 2 (standard error by logit transformation values).

Table 5. Main Models for Effectiveness/Sensitivity and Efficiency.

		Estimates	SE
Effectiveness/ Sensitivity	Fixed effects		
	Intercept	0.567	0.630
	Grade Level 0	-1.232	0.742
	Nonperformance Type 0	0.467	0.465
	Variance components		
	Second level	0.816	0.649
	Third level	1.094*	0.279
Efficiency	Fixed effects		
	Intercept	-0.553	0.429
	Grade Level 0	-0.235	0.505
	Nonperformance Type 0	-0.333	0.237
	Variance components		
	Second level	0.648	0.526
	Third Level	0.215*	0.074

Note: Values on the table are based on logit data. Grade Level 0 = younger students (kindergarten-elementary); Grade Level 1 = older students (middle or above); Nonperformance Type 0 = self-reports; Nonperformance Type 1 = teacher rating scales; Performance Type 0 = intelligence/general ability; Performance Type 1 = achievement/aptitude. * $p < .001$.

$SE = 0.51, p = .65$) and nonperformance methods ($\beta = -0.33, SE = 0.24, p = .17$) were not significant.

The mean proportion values (i.e., \bar{p}) and 95% CIs for each group in each moderator are presented in Table 6. Efficiency was higher in the kindergarten-elementary group ($\bar{p} = 0.51, 95\% \text{ CI } [0.42, 0.59]$) than the middle or higher school group ($\bar{p} = 0.26, 95\% \text{ CI } [0.22, 0.31]$), and teacher ratings had higher efficiency values ($\bar{p} = 0.57, 95\% \text{ CI } [0.46, 0.68]$) than other methods ($\bar{p} = 0.28, 95\% \text{ CI } [0.20, 0.37]$) albeit both were nonsignificant.

Sensitivity/Effectiveness. A total of 73 effect sizes from 15 published studies were available for the analyses. Homogeneity statistics indicated the data set was heterogeneous, $Q_T(41) = 2642.60, p < .01, I^2 = 97.3\%$. Thus, we conducted a random effects model using the logit method for efficiency proportions. Again, because two moderators (i.e., grade level and nonperformance method) had sufficient numbers of effect sizes for efficiency and the third had only one category (i.e., tests of intelligence or general ability), they were dummy coded and entered into the random effects regression model.

Again, an unconditional multilevel meta-analysis was conducted with no predictors. The unconditional model results yielded a mean effect size value of 0.59 (95% CI = 0.42; 0.74, $p = .33$). Second- and third-level variances were 1.19 ($p = .04$) and 1.21 ($p < .001$), respectively. The results of main three-level meta-analysis are presented in Table 5. Similar to the results in efficiency analyses, grade level ($\beta = -1.23, SE = 0.69, p < .11$) and nonperformance methods ($\beta = .47, SE = 0.44, p = .29$) were not significant.

Based on mean proportion values (i.e., \bar{p}) presented in Table 6, sensitivity/effectiveness was higher, though not significant, in the middle school or higher group ($\bar{p} = 0.71, 95\% \text{ CI } [0.63, 0.78]$) than the kindergarten-elementary group ($\bar{p} = 0.49, 95\% \text{ CI } [0.40, 0.57]$). Teacher ratings had similar sensitivity/effectiveness values ($\bar{p} = 0.60, 95\% \text{ CI } [0.53, 0.67]$) to other methods ($\bar{p} = 0.64, 95\% \text{ CI } [0.54, 0.73]$).

Specificity. A total of 33 effect sizes from three published studies were available for the analyses. Homogeneity statistics indicated the data set was heterogeneous, $Q_T(41) = 863.14, p < .01, I^2 = 96.3\%$. We only conducted an unconditional multilevel meta-analysis for the specificity data set because none of the moderators had a sufficient sample and

Table 6. Mean Effect Size Values for Efficiency, Effectiveness/Sensitivity, and Specificity Based on Random Effects Model.

Levels of moderators	Efficiency			
	<i>k</i>	Mean ES (\bar{p})*	95% CI	<i>Q</i>
Grade level				
Kindergarten-elementary	15	.51	.42, .59	172.30*
Middle school or higher	18	.26	.22, .31	167.82*
Nonperformance methods				
Teacher ratings	22	.57	.46, .68	2152.74*
Other	20	.28	.20, .37	454.52*
Levels of moderators	Effectiveness/sensitivity			
	<i>k</i>	ES	95% CI	<i>Q</i>
Grade level				
Kindergarten-elementary	42	.49	.40, .57	506.00*
Middle school or higher	18	.71	.63, .78	346.44*
Nonperformance methods				
Teacher ratings	51	.60	.53, .67	2216.77*
Other	22	.64	.54, .73	422.02*
Levels of moderators	Specificity			
	<i>k</i>	ES	95% CI	<i>Q</i>
Grade level				
Kindergarten-elementary	24	.73	.66, .80	68.00*
Middle school or higher				
Nonperformance methods				
Teacher ratings	33	.71	.65, .76	863.14*
Other				

Note. ES = effect size; CI = confidence interval. ES values were presented as proportions (percentages) from random effects model. The only type of performance measure was intelligence or general ability as criterion. $dfs = k - 1$ where *k* represents the number of ES.

* $p < .001$.

the data set consisted of proportion values obtained from the studies that employed tests of intelligence as the performance method, teacher rating scales as the nonperformance method, and elementary-kindergarten students. The unconditional model results yielded a mean effect size value of .70 (95% CI = [.59, .80], $p = .11$). Second- and third-level variances were 0.13 ($p < .28$) and 0.46 ($p = .007$), respectively.

Discussion

Our analyses from two different studies provided interesting and, to some degree, overlapping results about the consistency between the nonperformance and performance methods of gifted identification. Analyses with Pearson r as the effect size in Study 1 indicate there is a moderate relationship ($r = .30$) between the two methods based on Cohen's (1988) criteria on the magnitude of effect sizes. On the other hand, heterogeneity of the effect sizes required an analysis of the moderators, allowing for the observation of the influence of the grade level, type of nonperformance method, and type of performance method. This analysis has revealed the conditions

under which consistency between the nonperformance and performance methods are remarkably higher or lower than the overall mean effect size.

Analyses of the moderators indicate the consistency between the nonperformance and performance methods tends to be higher when teacher ratings are used as a method of nonperformance identification versus other methods such as self-reports and parent or teacher nominations. This finding is not surprising. Teachers observe and interact with students frequently and on a regular basis, which put them in a special position in terms of the recognition of the skills and abilities that help students with academic excellence and outperforming on the performance tests (J. Borland, 1978; Bracken & Brown, 2006; Jarosewich, Pfeiffer, & Morris, 2002). Therefore, our results support the previous findings indicating the usefulness of teacher ratings even in the identification of those who are gifted beyond stereotypical conceptions (Achenbach, 1997; Rohrer, 1995).

Previous findings favoring parents' over teachers' ratings (e.g., Ciha et al., 1974; Jacobs, 1971) are not supported by our results when performance measures are used as the basis

of comparison. The lower correlations with nominations could be related to a narrow conceptualization of giftedness (Bishofberger, 2012; Brighton, Moon, Jarvis, & Hockett, 2007) accompanied by lack of a structure for the nomination process that is overcome with the teacher rating scales. Concerns around the scientific rigor of the nomination process (Moon & Brighton, 2008; Neumeister et al., 2007) are worthy of attention with our findings. However, inclusion of these methods can still provide valuable information when the goal is to assess the aspects of giftedness not measured through performance measures. This approach would also be consistent with the suggestion to be inclusive in identification practices (Tannenbaum, 2003). A better approach would be to train teachers about gifted characteristics and effective identification (Bégin & Gagné, 1994; McCoach & Siegle, 2007). The way nomination is used in gifted identification may also need to be reconsidered. Erwin and Worrell (2012) argued that teachers should be asked to nominate students who do “the best academic work” because they have the opportunity to observe and track students’ academic work and performance on a continuing basis. This suggestion can be useful because it does not require involving teacher conceptualizations of giftedness in the nomination process.

Based on our findings in this study, gifted identification seems to be more uncertain for parents and students rather than teachers when performance measures are the criteria. Hunsaker, Finley, and Frank (1997) found teacher nominations could be useful for gifted identification when appropriate instruments (i.e., rating scales) are provided to the teachers. With a medium correlation, teacher ratings should be used with the intent of supporting the identification process as additional data instead of replacing a performance gifted identification methods. Teacher ratings, in that sense, would be useful as additional evidence rather than as the single point of evidence for a final decision regarding identification (Haroutounian, 1995; Jarosewich, Pfeiffer, & Morris, 2002; Pfeiffer, 2002).

Some researchers (e.g., Gear, 1976; Sattler, 1982) criticized teacher judgments for lacking accuracy in the identification process. Although a medium correlation seems to give some credit to Gear’s (1976) conclusion, teachers seem to be more consistent with the performance related identification methods when they use rating scales. But, which rating scale should be used? The selection of the teacher rating scales is a critical decision as they vary in the way they measure giftedness. Besides general psychometric qualities of the instruments (e.g., reliability and validity), this decision should be made on the basis of the way giftedness is defined in the local school system because states typically define giftedness differently (National Association for Gifted Children, 2015). Teacher rating scales that are consistent with the components included in the state gifted definition would be more useful. Another major issue of consideration is related to the goal of identification and programs to be offered. Identification methods should take into consideration the

curriculum and the programming for gifted programs (Feldhusen, Asher, & Hoover, 1984; VanTassel-Baska, 2006). Therefore, selection of the teacher rating scales should be guided by the program goals.

Given that teacher ratings scales are more consistent with performance measures and higher correlations among various measures increase the accuracy of identification (McBee et al., 2014), the selection of the instruments that have the highest correlation with other instruments is more defensible unless a special emphasis is given to areas such as leadership, communication, or athletic ability, which are not typically measured in traditional performance measures. Instruments with higher correlations and reliability could assist performance measures better in identifying the gifted students. As Jarosewich, Pfeiffer, and Morris (2002) did in their review, the current empirical evidence regarding these criteria should be compared to make an informed decision.

It is also important to note that teachers are not a uniform group and certain teacher characteristics such as experience can influence their conceptions of giftedness (Megay-Nespoli, 2001). Endepohls-Ulpe and Ruf (2006) found that experienced teachers have much more precise notions of giftedness than inexperienced teachers in the cognitive arena, which can influence their nominations, and therefore, using a rating scale can guide them to look for more specific criteria.

The students’ age group was not significant, which indicates the identification practices are not influenced by students’ age or grade level. While correlations are higher among younger students ($r = .29$) than the older students ($r = .16$) there was not a significant difference. Contrary to the expectation that correlation values would diminish because younger students are difficult to identify (Fatouros, 1986; Johnson, 1983; Roedell, 1989), teacher ratings and nomination procedures seem to be able to similarly match the performance test results, especially for students of younger age. This finding may be explained by the amount of time a teacher spends with younger students and the evaluations benefit from the longer periods of observation.

The type of performance methods are also not significant. The expectation that consistency could be lower with tests of creativity because of some undesirable characteristics of creative students (Bachtold, 1974; Dawson, 1997; Scott, 1999; Torrance, 1963; Westby & Dawson, 1995) is not supported although it does have a lower correlation ($r = .19$) than tests of intelligence ($r = .23$) and tests of achievement and aptitude ($r = .30$).

Results from Study 2 are based on the proportional values (percentages) reported as the values of efficiency, effectiveness/sensitivity, and specificity. When the mean proportions are calculated for these three indicators, specificity is remarkably higher (70%) than sensitivity/effectiveness (59%) and efficiency (39%). In other words, nonperformance methods perform better at identifying the nongifted students (true negatives) accurately based on the performance methods. The mean sensitivity/effectiveness value indicate that among

those who are gifted based on performance methods, 59% are identified as gifted based on nonperformance methods. In other words, 41% of the gifted students identified by the performance measures are not identified by the nonperformance methods. This percentage is slightly smaller than what McBee et al. (2014) reported in their simulation study estimating the sensitivity value within the range of 61% and 76%. McBee et al. also noted that the identification system in the state of Georgia might be producing more false negatives (unidentified gifted students) than false positives (nongifted students who were identified as gifted).

Efficiency values indicate that among those who are identified or nominated as gifted, only 39% were also identified as gifted based on the performance methods. In other words, nonperformance methods extend the pool of gifted candidates, but if a nonperformance method precedes a performance method and a final decision is made based on the performance method, the nonperformance methods will not help to diversify the pool. Although the effect size values are heterogeneous, moderators did not explain the variation.

These findings provide strong evidence for the usefulness of multiple criteria for gifted identification. Nonperformance methods seem to be extremely helpful in identifying nongifted students at the expense of nongifted students (30% false positives) that would actually be considered gifted when nonperformance methods are used as the only gifted identification source. Moreover, when nonperformance methods are used as the only identification method, they seem to be too liberal and not very efficient. The ratio of the false positives (61%) exceeded the ratio of true positives (39%).

Based on these results, multiple sources of data including both nonperformance and performance methods should ideally be collected and evaluated simultaneously (Ford, 1998; Siegle et al., 2010) rather than collecting and evaluating them consecutively (see Heller, 2004; Pfeiffer, 2002). When nonperformance methods are used as the only criteria, half of the potentially gifted students may be overlooked based on performance tests. Lohman and Gambrell (2012) cautioned that testing only those who were nominated by teachers could lead to dismissing gifted students who do not fit in with teachers' conceptions of giftedness. Chester (2003) indicated that simply lumping together various measures and scores together does not guarantee a successful outcome. Therefore, the way these multiple sources of identification methods (performance and nonperformance) are combined is quite critical.

McBee et al. (2014) compared three different approaches with combining multiple evidence. The conjunctive or "and" rule requires meeting the cutoff scores across all criteria, the disjunctive/complementary or "or" rule requires exceeding at least one of the cutoff scores, and the compensatory or "mean" rule uses the mean values from multiple assessments. The authors suggested using the "or" rule for larger programs such as enrichment, in which misidentification is less likely to be observed. This method is expected to be

more diverse and inclusive of low-income, minority students. As far as small programs with more precise programming, McBee et al. proposed using the "and" method because this method is more selective and minimizes the chances of failure. Overall, the authors favored the "mean" rule because of this method's capability of balancing the false positives of the "or" rule and false negatives of the "and" rule. This suggestion seems to be consistent with our findings because performance and nonperformance methods tend to identify different types of students although they do converge on some students; and they are also moderately related rather than strongly related.

As indicated in our findings, what should be done when performance and nonperformance test do not fully converge? Lohman and Lakin (2007) proposed a system for balancing and integrating these two different sources of evidence (i.e., performance and nonperformance) using two tests (e.g., CogAT and Renzulli-Hartmann Scale for Rating Behavioral Characteristics of Superior Students) as an example. They suggested that students in Category I, who had exceptionally high scores on both, are first to be accepted to these programs. Category II students are those who received exceptional scores on the CogAT but not as highly rated by teachers. They suggested monitoring the progress of these students because there may be an unwillingness not to eliminate these students from programming on the basis of teacher ratings. Category III represents students who got very high teacher ratings in spite of lower, but still high scores on CogAT. These students would be included in school-wide enrichment programs. Category IV consists of students who had high, not exceptional scores on the CogAT but are not rated highly by their teachers. These students would receive special services for the gifted only if they rank highly within their own underprivileged group. Lohman and Gambrell (2012) recommend using local norms more often and endorsed the proposition that teacher ratings should be used to be more inclusive and to provide opportunity rather than as exclusion criteria. Such guidelines are important because ability profiles of gifted students are more discrepant than average students (Lohman, Gambrell, & Lakin, 2008).

Another relevant finding in the McBee et al. (2014) study was the changes in false positives or incorrect identification (i.e., identification of the students that are not truly gifted) and false negatives (i.e., failure to identify truly gifted students) as a result of the reliability of the instruments and correlations among them. Their simulation study indicated that probability of false positives and false negatives increase as reliability and correlations decrease.

Some of the current practices of identification do not seem to align with the conclusions and suggestions in this study. For example, the State of New York started using both a performance measure (i.e., OLSAT) and nonperformance scale (i.e., GRS), but then they replaced GRS with Bracken School Readiness Assessment (BSRA; Bracken, 2002) leaning more toward performance assessment approach. The

Atlanta Public school system adopted the GRS just as a screening tool to determine qualification for automatic testing rather than using it concurrently with performance-based measures. Such practices are likely to continue because school districts tend to be slow in adopting alternative identification strategies (McClain & Pfeiffer, 2012; Reis & Renzulli, 2009).

There are limitations of the current meta-analysis. In spite of the voluminous research and controversy on the identification of the gifted students from minority and diverse groups (e.g., A. Y. Baldwin, 2005; Ford & Grantham, 2003; Lohman, 2005a; McBee, 2010; Naglieri & Ford, 2005), our data set did not include a satisfactory number of those studies and effect sizes to use in the analyses because they did not meet the inclusion criteria. Some other critical issues of identification such as comparison of gender (J. Borland, 1978; Siegle, 2001) as well as verbal and nonverbal tests (Lohman, 2005b) could also be explored in later studies when more research is conducted in the future. On a separate note, we did not present performance methods as the “truth” although it is a common terminology in diagnostic tests (Renzulli & Delcourt, 1986). Our purpose was rather to observe the magnitude of overlap and discrepancy across different methods.

One of the most frequently stated concerns about gifted identification is that some gifted students may be omitted (J. H. Borland & Wright, 2000; Worrell, 2003). Our findings indicate the use of nonperformance or performance methods alone would result in the identification of the different type of students with some overlap with the performance methods. Contrary to expectations, nonperformance methods do not include all students that would be identified as gifted by the performance methods. In other words, both nonperformance and performance methods seem to exclude some students who would be considered gifted according to the other method. Therefore, when both are used, performance and nonperformance methods should be collected and evaluated concurrently rather than successively.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

References marked with an asterisk were used in the meta-analysis. Achenbach, G. (1997). *The screening of gifted students in Pennsylvania: Do elementary teachers feel adequately prepared?* (Unpublished doctoral dissertation). Widener University, Chester, PA.

*Alexander, A. M. (1953). Teacher judgment of pupil intelligence and achievement is not enough. *Elementary School Journal*, 53, 396-401. doi:10.1086/458511

- Aljughaiman, A., & Mowrer-Reynolds, E. (2005). Teachers' conceptions of creativity and creative students. *Journal of Creative Behavior*, 39, 17-34. doi:10.1002/j.2162-6057.2005.tb01247.x
- Allyan, K., & Smadi, J. (1988). *The Jordanian version of the Advanced Raven's Progressive Matrices*. Amman, Jordan: University of Jordan.
- Alvidrez, J., & Weinstein, R. S. (1999). Early perceptions and later student academic achievement. *Journal of Educational Psychology*, 91, 731-746. doi:10.1037/0022-0663.91.4.731
- Alvino, J., McDonnel, R., & Richert, E. S. (1981). National survey of identification practices in gifted and talented education. *Exceptional Children*, 48, 124-132.
- Anderson, E. S., & Keith, T. Z. (1997). A longitudinal test of a model of academic success for at-risk high school students. *Journal of Educational Research*, 90, 259-268. doi:10.1080/00220671.1997.10544582
- Archambault, F., Westberg, K., Brown, S., Hallmark, B., Emmons, C., & Zhang, W. (1993). *Regular classroom practices with gifted students: Results of a national survey of classroom teachers* (RM93102). Storrs: The National Research Center on the Gifted and Talented, University of Connecticut. Retrieved from http://nrcgt.uconn.edu/research-based_resources/arch-west/
- *Ashman, S. S., & Vukelich, C. (1983). The effect of different types of nomination forms on teachers' identification of gifted children. *Psychology in the Schools*, 20, 518-527. doi:10.1002/1520-6807(198310)20:4<518::AID-PITS2310200421>3.0.CO;2-B
- Bachtold, L. (1974). The creative personality and the ideal pupil revisited. *Journal of Creative Behavior*, 8, 47-54. doi:10.1002/j.2162-6057.1974.tb01108.x
- Baldwin, A. Y. (2005). Identification concerns and promises for gifted students of diverse populations. *Theory into Practice*, 44, 105-114. doi:10.1207/s15430421tip4402_5
- *Baldwin, J. W. (1962). The relationship between teacher-judged giftedness, a group intelligence test and an individual intelligence test with possible gifted kindergarten pupils. *Gifted Child Quarterly*, 6, 153-156.
- Barron, F. (1976a, March). *New directions for gifted education* (Brief No. 3). Los Angeles, CA: National/State Leadership Training Institute on the Gifted and Talented. Retrieved from <http://files.eric.ed.gov/fulltext/ED131620.pdf>
- Barron, F. (1976b, October). *Symbolic scope as a predictor of ability in art*. Paper presented at 8th Western Symposium on Learning, Bellingham, WA.
- Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50, 1088-1101. doi:10.2307/2533446
- Bégin, J., & Gagné, F. (1994). Predictors of attitudes toward gifted education: A review of the literature and blueprints for future research. *Journal for the Education of the Gifted*, 17, 161-179. doi:10.1177/016235329401700206
- Bélanger, J., & Gagné, F. (2006). Estimating the size of the gifted/talented population from multiple identification criteria. *Journal for the Education of the Gifted*, 30, 131-163.
- Bianco, M., Harris, B., Garrison-Wade, D., & Leech, N. (2011). Gifted girls: Gender bias in gifted referrals. *Roeper Review*, 33, 170-181. doi:10.1080/02783193.2011.580500
- Bishopberger, S. D. (2012). *Elementary teachers' perceptions of giftedness: An examination of the relationship between teacher*

- background and gifted identification (Unpublished doctoral dissertation). University of Tennessee, Knoxville.
- Block, J., & Block, J. H. (1981). *The California Child Q-Set*. Palo Alto, CA: Consulting Psychologists Press.
- Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley. doi:10.1002/9780470743386
- Borland, J. (1978). Teacher identification of the gifted: A new look. *Journal for the Education of the Gifted*, 2, 22-32.
- Borland, J. H. (2009). Myth 2: The gifted constitute 3% to 5% of the population. Moreover, giftedness equals high IQ, which is a stable measure of aptitude: Spinal tap psychometrics in gifted education. *Gifted Child Quarterly*, 53, 236-238. doi:10.1177/0016986209346825
- Borland, J. H., & Wright, L. (2000). *Identifying and educating poor and under-represented gifted students*. In K. A. Heller, F. J. Monks, R. J. Sternberg, & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (2nd ed., pp. 587-594). Oxford, England: Pergamon.
- Bouchet, N., & Falk, R. F. (2001). The relationship among giftedness, gender, and overexcitability. *Gifted Child Quarterly*, 45, 260-267. doi:10.1177/001698620104500404
- Bracken, B. A. (1998). *Bracken Basic Concept Scale-Revised*. San Antonio, TX: Harcourt Assessments.
- Bracken, B. A. (2002). *Bracken School Readiness Assessment*. San Antonio, TX: Psychological Corporation.
- *Bracken, B. A., & Brown, E. F. (2006). Behavioral identification and assessment of gifted and talented students. *Journal of Psychoeducational Assessment*, 24, 112-122. doi:10.1177/0734282905285246
- *Bracken, B. A., & Brown, E. F. (2008). Early identification of high-ability students: Clinical assessment of behavior. *Journal for the Education of the Gifted*, 31, 403-426.
- Bracken, B. A., & Keith, L. K. (2004). *Professional manual for the Clinical Assessment of Behavior*. Lutz, FL: Psychological Assessment Resources.
- Brighton, C. M., Moon, T. R., Jarvis, J. M., & Hockett, J. A. (2007). *Primary grade teachers' conceptions of giftedness and talent: A case-based investigation* (RM07232). Storrs: The National Research Center on the Gifted and Talented, University of Connecticut. Retrieved from http://nrcgt.uconn.edu/research-based_resources/brigmoon/
- Brown, S. W., Renzulli, J. S., Gubbins, E. J., Siegle, D., Zhang, W., & Chen, C. (2005). Assumptions underlying the identification of gifted and talented students. *Gifted Child Quarterly*, 49, 68-79. doi:10.1177/001698620504900107
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- *Carman, C. A. (2011). Adding personality to gifted identification: Relationships among traditional and personality-based constructs. *Journal of Advanced Academics*, 22, 412-446. doi:10.1177/1932202X1102200303
- Cattell, R. B. (1958). *IPAT Culture Fair Intelligence Test*. Champagne, IL: Institute for Personality and Ability Testing.
- *Chambers, J. A., Barron, F., & Sprecher, J. W. (1980). Identifying gifted Mexican-American students. *Gifted Child Quarterly*, 24, 123-128. doi:10.1177/001698628002400306
- Chan, D. W. (2000). Exploring identification procedures of gifted students by teacher ratings, parent ratings and student self-reports in Hong Kong. *High Ability Studies*, 11, 69-82. doi:10.1080/713669176
- Chester, M. D. (2003). Multiple measures and high-stakes decisions: A framework for combining measures. *Educational Measurement: Issues and Practice*, 22, 32-41. doi:10.1111/j.1745-3992.2003.tb00126.x
- Cheung, M. W.-L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, 19, 211-229. doi:10.1037/a0032968
- *Ciha, T. E., Harris, R., Hoffman, C., & Potter, M. W. (1974). Parents as identifiers of giftedness, ignored but accurate. *Gifted Child Quarterly*, 18, 191-195.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Coleman, M. R., & Gallagher, J. J. (1995). State identification policies: Gifted students from special populations. *Roeper Review*, 17, 268-275. doi:10.1080/02783199509553681
- *Cornish, R. L. (1968). Parents', teachers', and pupils' perception of the gifted child's ability. *Gifted Child Quarterly*, 12, 14-17.
- Council of State Directors of Programs for the Gifted. (1999). *The 1998-99 state of the states gifted and talented report*. Longmont, CO: Author.
- *Crosby, E. G., & French, J. L. (2002). Psychometric data for teacher judgments regarding the learning behaviors of primary grade children. *Psychology in the Schools*, 39, 235-244. doi:10.1002/pits.10034
- Davis, G. A., & Rimm, S. B. (2003). *Education of the gifted and talented* (5th ed.). Boston, MA: Allyn & Bacon.
- Dawson, V. L. (1997). In search of the wild bohemian: Challenges in the identification of the creatively gifted. *Roeper Review*, 19, 148-152. doi:10.1080/02783199709553811
- *Dewing, K. (1970). The reliability and validity of selected tests of creative thinking in a sample of seventh-grade West Australian children. *British Journal of Educational Psychology*, 40, 35-42. doi:10.1111/j.2044-8279.1970.tb02096.x
- Duval, S. J., & Tweedie, R. L. (2000). A non-parametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89-98.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629-634. doi:10.1136/bmj.315.7109.629
- Elhoweris, H. (2008). Teacher judgment in identifying gifted/talented students. *Multicultural Education*, 15, 35-38.
- Elhoweris, H., Mutua, K., Alsheikh, N., & Holloway, P. (2005). The effect of the child's ethnicity on teachers' referral and recommendations decisions in the gifted/talented programs. *Remedial and Special Education*, 26, 25-31. doi:10.1177/07419325050260010401
- *Elliott, S. N., Argulewicz, E. N., & Turco, T. L. (1986). Predictive validity of the Scales for Rating the Behavioral Characteristics of Superior Students for gifted children from three sociocultural groups. *Journal of Experimental Education*, 55, 27-32. doi:10.1080/00220973.1986.10806431
- Endepohls-Ulpe, M., & Ruf, H. (2006). Primary school teachers' criteria for the identification of gifted pupils. *High Ability Studies*, 16, 219-228. doi:10.1080/13598130600618140
- Erwin, J. O., & Worrell, F. C. (2012). Assessment practices and the underrepresentation of minority students in gifted and talented education. *Journal of Psychoeducational Assessment*, 30, 74-87. doi:10.1177/0734282911428197

- Evans, K. (1996). Policy for the identification of students for gifted programs. *Journal of Secondary Gifted Education, 8*, 74-87. doi:10.1177/1932202x9600800204
- Fatouros, C. (1986). Early identification of gifted children is crucial . . . but how should we go about it? *Gifted Education International, 4*, 24-28. doi:10.1177/026142948600400107
- Feldhusen, J. F., Asher, J. W., & Hoover, S. M. (1984). Problems in the identification of giftedness, talent, or ability. *Gifted Child Quarterly, 28*, 149-151. doi:10.1177/001698628402800402
- Ford, D. Y. (1998). The underrepresentation of minority students in gifted education problems and promises in recruitment and retention. *Journal of Special Education, 32*, 4-14. doi:10.1177/002246699803200102
- Ford, D. Y., & Grantham, T. C. (2003). Providing access for culturally diverse gifted students: From deficit to dynamic thinking. *Theory into Practice, 42*, 217-225. doi:10.1207/s15430421tip4203_8
- Frasier, M. M. (1997). Multiple criteria: The mandate and the challenge. *Roeper Review, 20*, 2-4. doi:10.1080/02783199709553868
- Frasier, M. M., García, J. H., & Passow, A. H. (1995). *A review of assessment issues in gifted education and their implications for identifying gifted minority students* (RM95204). Storrs: The National Research Center on the Gifted and Talented, University of Connecticut. Retrieved from http://nrcgt.uconn.edu/research-based_resources/frasgarc/
- Friedman-Nimz, R. (2009). Myth 6: Cosmetic use of multiple selection criteria. *Gifted Child Quarterly, 53*, 248-250. doi:10.1177/0016986209346925
- *Gallagher, S. A. (1985). A comparison of the concept of over-excitabilities with measures of creativity and school achievement in sixth-grade students. *Roeper Review, 8*, 115-119. doi:10.1080/02783198509552950
- Garcia, T., & Pintrich, P. R. (1996). Assessing students' motivation and learning strategies in the classroom context: The Motivated Strategies for Learning Questionnaire. In M. Birenbaum & J. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 319-339). Boston, MA: Kluwer. doi:10.1007/978-94-011-0657-3_12
- Gear, G. H. (1976). Accuracy of teacher judgment in identifying intellectually gifted children: A review of the literature. *Gifted Child Quarterly, 20*, 478-490.
- Gentry, M., & Mann, R. L. (2008). *Total school cluster grouping and differentiation: A comprehensive, research-based plan for raising student achievement and improving teacher practices*. Mansfield Center, CT: Creative Learning Press.
- George, W. C. (1979). The talent-search concept: An identification strategy for the intellectually gifted. *Journal of Special Education, 13*, 221-237. doi:10.1177/002246697901300303
- Georgia Department of Education. (2014). *Georgia resource manual for gifted education services*. Retrieved from <http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Curriculum-and-Instruction/Documents/2012%202013%20GA%20Gifted%20%20Resource%20Manual.pdf>
- *Gilliam, J. E., Carpenter, B. O., & Christensen, J. R. (1996). *Gifted and Talented Evaluation Scales*. Waco, TX: Prufrock Press.
- *Gilliam, J. E., & Jerman, O. (2015). *Gifted and Talented Evaluation Scales: Examiner's manual* (2nd ed.). Austin, TX: Pro-Ed.
- Haroutounian, J. (1995). Talent identification and development in the arts: An artistic/educational dialogue. *Roeper Review, 18*, 112-117. doi:10.1080/02783199509553710
- *Harrington, D. M., Block, J., & Block, J. H. (1983). Predicting creativity in preadolescence from divergent thinking in early childhood. *Journal of Personality and Social Psychology, 45*, 609-623. doi:10.1037/0022-3514.45.3.609
- *Hartsough, C. S., Elias, P., & Wheeler, P. (1983). Evaluation of a nonintellectual assessment procedure for the early screening of exceptionality. *Journal of School Psychology, 21*, 133-142. doi:10.1016/0022-4405(83)90038-9
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic.
- Heller, K. A. (2000). Hochbegabungsdiagnose [Identification]. In K. A. Heller (Hrsg.), *Begabungsdiagnostik in der Schul- und Erziehungsberatung. Lehrbuch* [Aptitude diagnostics in the school and educational counseling textbook] (2nd ed., pp. S241-S258). Bern, Switzerland: Huber.
- Heller, K. A. (2004). Identification of gifted and talented students. *Psychology Science, 46*, 302-323.
- Heller, K. A., Gaedike, A.-K., & Weinläder, H. (1987). *Kognitiver Fähigkeits Test (KFT 4-13+ (2. Aufl.)* [Cognitive Abilities Tests—KFT 4-13+] (2nd ed.). Weinheim, Germany: Beltz Testgesellschaft.
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeits Test 4-12. Klassen* [Cognitive Abilities Tests 4-12. Classes] (revised version). Göttingen, Germany: Hogrefe.
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557-560. doi:10.1136/bmj.327.7414.557
- Hoge, R. D., & Cudmore, L. (1986). The use of teacher-judgment measures in the identification of gifted pupils. *Teaching & Teacher Education, 2*, 181-196. doi:10.1016/0742-051X(86)90016-8
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Hunsaker, S. L., & Callahan, C. M. (1995). Creativity and giftedness: Published instrument uses and abuses. *Gifted Child Quarterly, 39*, 110-114. doi:10.1177/001698629503900207
- Hunsaker, S. L., Finley, V. S., & Frank, E. L. (1997). An analysis of teacher nominations and student performance in gifted programs. *Gifted Child Quarterly, 41*, 19-24. doi:10.1177/001698629704100203
- *Hunter, J. A., Jr., & Lowe, J. D., Jr. (1978). The use of the WISC-R, Otis, Iowa, and SRBCSS in identifying gifted elementary students. *Southern Journal of Educational Research, 12*(1), 59-65.
- *Jacobs, J. C. (1971). Effectiveness of teacher and parent identification of gifted children as a function of school level. *Psychology in the Schools, 8*, 140-142. doi:10.1002/1520-6807(197104)8:2<140::AID-PITS2310080210>3.0.CO;2-K
- 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. doi:10.1177/073428290202000401
- Johnson, L. G. (1983). Giftedness in preschool: A better time for development than identification. *Roeper Review, 5*, 13-15. doi:10.1080/02783198309552715
- Kampylis, P., Berki, E., & Saarioluoma, P. (2009). In-service and prospective teachers' conceptions of creativity. *Thinking Skills and Creativity, 4*, 15-29. doi:10.1016/j.tsc.2008.10.001
- Kaufman, A. S., & Harrison, P. L. (1986). Intelligence tests and gifted assessment: What are the positives? *Roeper Review, 8*, 154-159. doi:10.1080/02783198609552961

- *Kirk, W. D. (1966). A tentative screening procedure for selecting bright and slow children in kindergarten. *Exceptional Children*, 33, 235-241.
- Kirschenbaum, R. J. (1983). Let's cut out the cut-off score in the identification of the gifted. *Roeper Review*, 5, 6-10. doi:10.1080/02783198309552713
- Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis? *Research Synthesis Methods*, 2, 61-76. doi:10.1002/jrsm.35
- Krisel, S. C., & Brown, R. S. (1997). Georgia's journey toward multiple-criteria identification of gifted students. *Roeper Review*, 20, A1-A3. doi:10.1080/02783199709553867
- Lee, L. (1999). Teachers' conceptions of gifted and talented young children. *High Ability Studies*, 10, 183-196. doi:10.1080/1359813990100205
- *Lee, S. Y., & Olszewski-Kubilius, P. (2006). Comparisons between talent search students qualifying via scores on standardized tests and via parent nomination. *Roeper Review*, 28, 157-166. doi:10.1080/02783190609554355
- Leiter, R. G. (1952). *The Leiter International Performance Scale—Manual* (Vol. 2). Washington, DC: Psychological Services Center Press.
- Light, R., & Pillemer, D. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49). Thousand Oaks, CA: Sage.
- Lohman, D. F. (2005a). An aptitude perspective on talent: Implications for identification of academically gifted minority students. *Journal for the Education of the Gifted*, 28, 333-360.
- Lohman, D. F. (2005b). The role of nonverbal ability tests in identifying academically gifted students: An aptitude perspective. *Gifted Child Quarterly*, 49, 111-138. doi:10.1177/001698620504900203
- Lohman, D. F., & Gambrell, J. (2012). Use of nonverbal measures in gifted identification. *Journal of Psychoeducational Assessment*, 30, 25-44. doi:10.1177/0734282911428194
- Lohman, D. F., Gambrell, J., & Lakin, J. (2008). The commonality of extreme discrepancies in the ability profiles of academically gifted students. *Psychology Science*, 50, 269-282.
- Lohman, D. F., & Lakin, J. (2007). Nonverbal test scores as one component of an identification system: Integrating ability, achievement, and teacher ratings. In J. VanTassel-Baska (Ed.), *Alternative assessments for identifying gifted and talented students* (pp. 41-66). Austin, TX: Prufrock Press.
- Lorge, I., & Thorndike, R. L. (1962). *The Lorge-Thorndike Intelligence Tests*. Boston, MA: Houghton, Mifflin.
- *Lowenstein, L. (1982). Teachers' effectiveness in identifying gifted children. *Gifted Educational International*, 1, 33-35. doi:10.1177/026142948200100121
- Lysy, K. Z., & Piechowski, M. M. (1983). Personal growth: An empirical study using Jungian and Dabrowskian measures. *Genetic Psychology Monographs*, 108, 267-320.
- Madden, R., Gardner, E., Rudman, H., Karlson, B., & Merwin, J. (1973). *Stanford Achievement Test norm booklet, Form A, Primary Level I to Intermediate Level I battery*. New York, NY: Harcourt Brace Jovanovich.
- Marland, S. P., Jr. (1972). *Education of the gifted and talented: Report to the Congress of the United States by the U.S. Commissioner of Education*. Washington, DC: Government Printing Office.
- Martinson, R. A. (1974). *The identification of the gifted and talented: An instructional syllabus for the national summer leadership training institute on the education of the gifted and talented*. Ventura, CA: Office of the Ventura County Superintendent of Schools.
- Masten, W. G., & Plata, M. (2000). Acculturation and teacher ratings of Hispanic and Anglo-American students. *Roeper Review*, 23, 45-46. doi:10.1080/02783190009554061
- Masten, W. G., Plata, M., Wenglar, K., & Thedford, J. (1999). Acculturation and teacher ratings of Hispanic and Anglo-American students. *Roeper Review*, 22, 64-65. doi:10.1080/02783199909554001
- McBee, M. T. (2006). A descriptive analysis of referral sources for gifted identification screening by race and socioeconomic status. *Journal of Secondary Gifted Education*, 17, 103-111.
- McBee, M. T. (2010). Examining the probability of identification for gifted programs for students in Georgia elementary schools: A multilevel path analysis study. *Gifted Child Quarterly*, 54, 283-297. doi:10.1177/0016986210377927
- McBee, M. T., Peters, S. J., & Waterman, C. (2014). Combining scores in multiple-criteria assessment systems: The impact of combination rule. *Gifted Child Quarterly*, 58, 69-89. doi:10.1177/0016986213513794
- McClain, M. C., & Pfeiffer, S. (2012). Identification of gifted students in the United States today: A look at state definitions, policies, and practices. *Journal of Applied School Psychology*, 28, 59-88. doi:10.1080/15377903.2012.643757
- McCoach, D. B., & Siegle, D. (2007). What predicts teachers' attitudes toward the gifted? *Gifted Child Quarterly*, 51, 246-254. doi:10.1177/0016986207302719
- Megay-Nespoli, K. (2001). Beliefs and attitudes of novice teachers regarding instruction of academically talented learners. *Roeper Review*, 23, 178-182. doi:10.1080/02783190109554092
- Ministry of Education & United Nations Relief and Works Agency for Palestine Refugees in the Near East. (1990). *Mathematical Skills Assessment*. Amman, Jordan: Author.
- Moon, T. R., & Brighton, C. M. (2008). Primary teachers' conceptions of giftedness. *Journal for the Education of the Gifted*, 31, 447-480.
- Naglieri, J. A. (1991). *Naglieri Nonverbal Ability Test*. San Antonio, TX: Harcourt Assessments.
- Naglieri, J. A., & Ford, D. Y. (2005). Increasing minority children's participation in gifted classes using the NNAT: A response to Lohman. *Gifted Child Quarterly*, 49, 29-36. doi:10.1177/001698620504900104
- National Association for Gifted Children. (2009). *State of the states in gifted education: 2008-2009*. Washington, DC: Author.
- National Association for Gifted Children. (2013). *State of the states in gifted education: 2012-2013*. Washington, DC: Author.
- National Association for Gifted Children. (2015). *State of the states in gifted education: 2014-2015*. Washington, DC: Author. Retrieved from <http://www.nagc.org/sites/default/files/key%20reports/2014-2015%20State%20of%20the%20States%20%28final%29.pdf>
- *Neber, H. (2004). Teacher identification of students for gifted programs: Nominations to a summer school for highly-gifted students. *Psychology Science*, 46, 348-362.

- Neber, H., & Heller, K. A. (1995). *Untersuchungen zur Nomination von Teilnehmern für die Deutsche Schulerakademie. Forschungsbericht* [Investigations on the nomination of participants for the German Pupils Academy. Research paper]. München, Germany: Ludwig-Maximilians-Universität.
- Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77-101. doi:10.1037/0003-066X.51.2.77
- Neumeister, K. L. S., Adams, C. M., Pierce, R. L., Cassady, J. C., & Dixon, F. A. (2007). Fourth-grade teachers' perceptions of giftedness: Implications for identifying and serving diverse gifted students. *Journal for the Education of the Gifted, 30*, 479-499.
- Otis, A. S., & Lennon, R. T. (1967). *Otis-Lennon Mental Ability Test, Elementary I level*. New York, NY: Harcourt, Brace & World.
- *Pegnato, C. V., & Birch, J. W. (1959). Locating gifted children in junior high schools: A comparison of methods. *Exceptional Children, 25*, 300-304.
- Peterson, J. S. (1999). Gifted—through whose cultural lens? An application of the postpositivistic mode of inquiry. *Journal for the Education of the Gifted, 22*, 254-283. doi:10.1177/016235329902200403
- Pfeiffer, S. I. (2002). Identifying gifted and talented students: Recurring issues and promising solutions. *Journal of Applied School Psychology, 19*, 31-50. doi:10.1300/J008v19n01_03
- Pfeiffer, S. I. (2003). Challenges and opportunities for students who are gifted: What the experts say. *Gifted Child Quarterly, 47*, 161-169. doi:10.1177/001698620304700207
- Pfeiffer, S. I. (2015). *Essentials of gifted assessment*. Hoboken, NJ: John Wiley.
- Pfeiffer, S. I., & Blei, S. (2008). Gifted identification beyond the IQ test: Rating scales and other assessment procedures. In S. I. Pfeiffer (Ed.), *Handbook of giftedness in children: Psychoeducational theory, research, and best practices* (pp. 177-198). New York, NY: Springer. doi:10.1007/978-0-387-74401-8_10
- Pfeiffer, S. I., & Jarosewich, T. (2003). *Gifted Rating Scales*. San Antonio, TX: Harcourt Assessment.
- *Pfeiffer, S. I., & Jarosewich, T. (2007). The gifted rating scales—school form an analysis of the standardization sample based on age, gender, race, and diagnostic efficiency. *Gifted Child Quarterly, 51*, 39-50. doi:10.1177/0016986206296658
- *Pfeiffer, S. I., & Petscher, Y. (2008). Identifying young gifted children using the gifted rating scales: Preschool/ kindergarten form. *Gifted Child Quarterly, 52*, 19-29. doi:10.1177/0016986207311055
- Reis, S. M., & Renzulli, J. S. (2009). Myth No. 1: The gifted and talented constitute one single homogeneous group and giftedness is a way of being that stays in the person over time and experiences. *Gifted Child Quarterly, 53*, 233-235. doi:10.1177/0016986209346824
- Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan, 60*, 180-184.
- Renzulli, J. S. (1990). A practical system for identifying gifted and talented students. *Early Child Development and Care, 63*, 9-18. doi:10.1080/0300443900630103
- Renzulli, J. S. (2011). What makes giftedness? Reexamining a definition: Giftedness needs to be redefined to include three elements: Above-average intelligence, high levels of task commitment, and high levels of creativity. *Phi Delta Kappan, 92*, 81-88. doi:10.1177/003172171109200821
- Renzulli, J. S., & Delcourt, M. A. (1986). The legacy and logic of research on the identification of gifted persons. *Gifted Child Quarterly, 30*, 20-23. doi:10.1177/001698628603000104
- Renzulli, J. S., & Hartman, R. K. (1971). Out of the classroom: Scale for Rating Behavioral Characteristics of Superior Students. *Exceptional Children, 38*, 243-248.
- Renzulli, J. S., & Reis, S. M. (Eds.). (2004). *Identification of students for gifted and talented programs*. Thousand Oaks, CA: Corwin Press.
- *Renzulli, J. S., Hartman, R. K., & Callahan, C. M. (1971). Teacher identification of superior students. *Exceptional Children, 38*, 211-214.
- Richert, E. S. (1987). Rampant problems and promising practices in the identification of disadvantaged gifted students. *Gifted Child Quarterly, 31*, 149-154. doi:10.1177/001698628703100403
- Roedell, W. C. (1989). *Early development of gifted children*. In J. L. VanTassel-Baska & P. Olszewski-Kubilius (Eds.), *Patterns of influence on gifted learners: The home, the self, and the school* (pp. 13-28). New York, NY: Teachers College.
- Rohrer, J. C. (1995). Primary teacher conceptions of giftedness: Image, evidence, and non-evidence. *Journal for the Education of the Gifted, 18*, 269-283. doi:10.1177/016235329501800304
- Rothstein, H., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis—Prevention, assessment and adjustments*. Chichester, England: Wiley. doi:10.1002/0470870168
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin, 85*, 638-641. doi:10.1037/0033-2909.86.3.638
- Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *Handbook of research synthesis* (pp. 231-244). New York, NY: Russell Sage Foundation.
- *Rust, J. O., & Lose, B. D. (1980). Screening for giftedness with the Slosson and the Scale for Rating Behavioral Characteristics of Superior Students. *Psychology in the Schools, 17*, 446-451. doi:10.1002/1520-6807(198010)17:4<446::AID-PITS2310170405>3.0.CO;2-7
- *Ryan, J. S. (1983). Identifying intellectually superior black children. *Journal of Educational Research, 76*, 153-156. doi:10.1080/00220671.1983.10885441
- Sattler, J. M. (1982). *Assessment of children's intelligence and special abilities*. Boston, MA: Allyn & Bacon.
- Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research, 84*, 328-364. doi:10.3102/0034654313500826
- Scott, C. L. (1999). Teachers' biases toward creative children. *Creativity Research Journal, 12*, 321-328. doi:10.1207/s15326934crj1204_10
- Shaklee, B. D. (1992). Identification of young gifted students. *Journal for the Education of the Gifted, 15*, 134-144. doi:10.1177/016235329201500203
- Siegle, D. (2001, April). *Teacher bias in identifying gifted and talented students*. Paper presented at the 80th Annual Meeting of the Council for Exceptional Children, Kansas City, MO.
- Siegle, D., & Powell, T. (2004). Exploring teacher biases when nominating students for gifted programs. *Gifted Child Quarterly, 48*, 21-29. doi:10.1177/001698620404800103

- Siegle, D., Moore, M., Mann, R. L., & Wilson, H. E. (2010). Factors that influence in-service and preservice teachers' nominations of students for gifted and talented programs. *Journal for the Education of the Gifted*, 33, 337-360.
- *Silverman, L. K., Chitwood, D. G., & Waters, J. L. (1986). Young gifted children: Can parents identify giftedness? *Topics in Early Childhood Special Education*, 6, 23-38. doi:10.1177/027112148600600106
- Slosson, R. L. (1963). *The Slosson Intelligence Test for Children and Adults*. East Aurora, NY: Slosson Educational Publications.
- *Spinath, B., & Spinath, F. M. (2005). Development of self-perceived ability in elementary school: The role of parents' perceptions, teacher evaluations, and intelligence. *Cognitive Development*, 20, 190-204. doi:10.1016/j.cogdev.2005.01.001
- Sternberg, R. J. (1986). Identifying the gifted through IQ: Why a little bit of knowledge is a dangerous thing. *Roepers Review*, 8, 143-147. doi:10.1080/02783198609552958
- Sternberg, R. J., Ferrari, M., Clinkenbeard, P., & Grigorenko, E. L. (1996). Identification, instruction, and assessment of gifted children: A construct validation of a triarchic model. *Gifted Child Quarterly*, 40, 129-137. doi:10.1177/001698629604000303
- Stevens, J. R., & Taylor, A. M. (2009). Hierarchical dependence in meta-analysis. *Journal of Educational and Behavioral Statistics*, 34, 46-73. doi:10.3102/1076998607309080
- Stiensmeier-Pelster, J., Spinath, B., Schone, C., & Dickhauser, O. (2002). *Elternform der Skalen zur Erfassung des schulischen Selbstkonzepts* [Parent version of the School-based Self-concept Scales] (Unpublished test material). University of Giessen, Germany.
- *Subhi, T. (1997). Who is gifted? A computerized identification procedure. *High Ability Studies*, 8, 189-211. doi:10.1080/1359813970080205
- *Swenson, E. V. (1978). Teacher-assessment of creative behavior in disadvantaged children. *Gifted Child Quarterly*, 22, 338-343.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293. doi:10.1126/science.3287615
- Tannenbaum, A. J. (2003). Nature and nurture of giftedness. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 45-59). New York, NY: Allyn & Bacon.
- Tennessee Department of Education. (2010). *Tennessee state plan for the education of intellectually gifted students*. Retrieved from https://tn.gov/assets/entities/education/attachments/se_eligibility_gifted_manual.pdf
- Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence: A guide to the administration of the New Revised Stanford-Binet Tests of Intelligence*. Oxford, England: Houghton Mifflin.
- Terman, L. M., & Merrill, M. A. (1960). *Stanford-Binet Intelligence Scale, Form LM*. Boston, MA: Houghton Mifflin.
- Terman, L. M., & Merrill, M. A. (1962). *Stanford-Binet Intelligence Scale—Third Revision. Form LM*. Boston, MA: Houghton Mifflin.
- Thurstone, L. L., & Thurstone, T. G. (1949). *Examiner manual for the SRA Primary Mental Abilities Test (Form 10414)*. Chicago, IL: Science Research.
- Torrance, E. P. (1962). *The Minnesota Tests of Creative Thinking*. In E. P. Torrance (Ed.), *Guiding creative talent* (pp. 44-64). Englewood Cliffs, NJ: Prentice Hall. doi:10.1037/13134-003
- Torrance, E. P. (1963). The creative personality and the ideal pupil. *Teachers College Record*, 65, 220-226.
- Torrance, E. P. (1984). The role of creativity in identification of the gifted and talented. *Gifted Child Quarterly*, 28, 153-156. doi:10.1177/001698628402800403
- Torrance, E. P. (1974). *The Torrance Tests of Creative Thinking: Norms—Technical manual*. Princeton, NJ: Personnel Press.
- Traxler, A. E. (1939). A study of the California Test of Mental Maturity: Advanced battery. *Journal of Educational Research*, 32, 329-335. doi:10.1080/00220671.1939.10880841
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three level meta-analyses of dependent effect sizes. *Behavior Research Methods*, 45, 576-594. doi:10.3758/s13428-012-0261-6
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2014). Meta-analysis of multilevel outcomes: A multilevel approach. *Behavior Research Methods*, 47, 1274-1294. doi:10.3758/s13428-014-0527-2
- VanTassel-Baska, J. (2006). A content analysis of evaluation findings across 20 gifted programs: A clarion call for enhanced gifted program development. *Gifted Child Quarterly*, 50, 199-215. doi:10.1177/001698620605000302
- VanTassel-Baska, J., Feng, A. X., & Evans, B. L. (2007). Patterns of identification and performance among gifted students identified through performance tasks: A three-year analysis. *Gifted Child Quarterly*, 51, 218-231. doi:10.1177/0016986207302717
- Wallach, M., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction*. New York, NY: Holt, Rinehart & Winston.
- Ward, W. C. (1968). Creativity in young children. *Child Development*, 39, 737-754. doi:10.2307/1126980
- Wechsler, D. (1949). *Intelligence Scale for Children—Manual*. Oxford, England: Psychological Corporation.
- Wechsler, D. (1967). *Wechsler Preschool and Primary Scale of Intelligence—WPPSI*. Oxford, England: Psychological Corporation.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised*. Oxford, England: Psychological Corporation.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV)*. San Antonio, TX: Psychological Corporation.
- Weiss, R., & Osterland, J. (1997). *Grundintelligenztest Skala 1 CFT 1* [Culture Fair Intelligence Test Scale 1] (5th rev. ed.). Göttingen, Germany: Hogrefe.
- Westby, E. L., & Dawson, V. L. (1995). Creativity: Asset or burden in the classroom? *Creativity Research Journal*, 8, 1-10. doi:10.1207/s15326934crj0801_1
- *Wilson, C. D. (1963). Using test results and teacher evaluation in identifying gifted pupils. *Personnel and Guidance Journal*, 41, 720-721. doi:10.1002/j.2164-4918.1963.tb02381.x
- Wilson, D. (2001). *Effect size determination program*. College Park: University of Maryland.
- Woodcock, R. W., & Mather, N. (1990). *Woodcock-Johnson Psycho-Educational Battery—Revised*. Allen, TX: DLM Teaching Resources.
- Worrell, F. C. (2003). *Why are there so few African Americans in gifted programs?* In C. C. Yeakey & R. D. Henderson (Eds.), *Surmounting the odds: Education, opportunity, and society in the new millennium* (pp. 423-454). Greenwich, CT: Information Age.

- Worrell, F. C. (2009). Myth 4: A single test score or indicator tells us all we need to know about giftedness. *Gifted Child Quarterly*, 53, 242-244. doi:10.1177/0016986209346828
- Worrell, F. C., & Erwin, J. O. (2011). Best practices in identifying students for gifted and talented education programs. *Journal of Applied School Psychology*, 27, 319-340. doi:10.1080/15377903.2011.615817
- Zachary, R. A. (1991). *Shipley Institute of Living Scale—Revised manual*. Los Angeles, CA: Western Psychological Services.

Author Biographies

Selcuk Acar is an assistant professor at the International Center for Studies in Creativity, SUNY Buffalo State. He earned his PhD from University of Georgia in educational psychology with an emphasis in gifted and creative education. His research interests include assessment of creativity, divergent thinking, and identification of the gifted and talented. He presented his research at the major conferences and published articles in the peer-reviewed journals such as *Psychology of Aesthetics, Creativity, and the Arts*; *Creativity Research Journal*; and *Journal of Creative Behavior*. He has also contributed to the literature through book chapters and encyclo-

pedia entries. He is currently teaching graduate- and undergraduate-level courses on creativity and leadership.

Sedat Sen is an assistant professor at Harran University, Sanliurfa, Turkey. He earned his PhD in 2014 from University of Georgia. His research interests focus on quantitative methods, applied statistics, and psychometrics. His recent publications appeared in *Applied Psychological Measurement* and *Psychology of Creativity, Aesthetics and the Arts*. He teaches courses about research methods, educational measurement, and statistics at both undergraduate and graduate levels.

Nur Cayirdag is an assistant professor at İstanbul Sabahattin Zaim University. She received her PhD in counseling psychology. She also earned her second MA degree in gifted and creative education. Her most recent works appeared in the *Journal of Adolescence*, *Encyclopedia of Creativity*, *Handbook of Research on the Education of Young Children*, and *Researching Creative Learning: Methods and Approaches*. She teaches courses and delivers trainings on creativity and gifted education. She contributed to several projects organized by Torrance Center for Creativity and Talent Development and Counseling and Human Development Services at University of Georgia as well as International Center for Studies in Creativity at Buffalo State, State University of New York.