



The identification of gifted underachievement: Validity evidence for the commonly used methods

Rahmi Luke Jackson*  and Jae Yup Jung 

School of Education, The University of New South Wales, Sydney, New South Wales, Australia

Background. Much confusion exists about the underachievement of gifted students due to significant variations in how the phenomenon has been identified. From a review of the literature, five methods were found to be commonly used to identify gifted underachievement.

Aims. The purpose of the study was to assess the equivalence of the commonly used methods to identify gifted underachievement, and to determine which of these methods may be optimal.

Sample. Data were collected from a school in Sydney, Australia.

Method. Three measures of convergence (i.e., difference in proportions, phi association, and kappa agreement) were used to assess the equivalence of the identification methods, while latent class analysis was used to determine the optimal identification method.

Results. The convergence evidence suggested that the commonly used identification methods may not be considered convergent, while the criterion evidence indicated that one of the five identification methods may have strong levels of criterion validity.

Conclusions. A conclusion was reached that the simple difference method may be the most valid method to identify gifted underachievement.

Although giftedness may be perceived to provide advantage in the modern world (Jung, 2014), gifted students have in fact been reported to be suffering from an ‘epidemic’ of underachievement (Rimm, 2003). This disturbing claim is supported by evidence which indicates that up to, or even greater than, half the population of gifted students exhibit significant academic underachievement (although the most commonly reported prevalence rates are between 10% and 20%; Hsieh, Sullivan, & Guerra, 2007; Rimm, 2003; Steenbergen-Hu, Olszewski-Kubilius, & Calvert, 2020; White, Graham, & Blaas, 2018). Of note, the disappointing educational outcomes for many gifted students do not appear to be fully explainable by the commonly recognized sources of disadvantage, including socio-economic status and ethnic background. The various socio-emotional factors associated with the learning needs of these students (i.e., self-perceptions,

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

*Correspondence should be addressed to Rahmi Luke Jackson, School of Education, The University of New South Wales, Sydney NSW 2052 Australia (email: rahmijackson@hotmail.com).

[Correction added on 10 May 2022, after first online publication: CAUL funding statement has been added.]

self-concept, self-efficacy, self-regulation, attitudes towards teachers/school, anxiety, emotional engagement, and goal orientations) also appear to give rise to underachievement (Coleman & Cross, 2014; Desmet, Pereira, & Peterson, 2020; Gilar-Corbi, Veas, Miñano, & Castejón, 2019; Masden, Leung, Shore, Schneider, & Udvari, 2015; Rimm, 2003; Steenbergen-Hu et al., 2020; White et al., 2018).

When educators attempt to respond to the problem of underachievement in gifted students, they may not always find clear guidance from the literature due to the lack of consensus among scholars (Gorard & Smith, 2004; Steenbergen-Hu et al., 2020). One of the major reasons for such confusion may lie in the significant variations in how gifted underachievement has been identified (Mofield & Parker Peters, 2019; White et al., 2018). Of note, Dowdall and Colangelo (1982) indicated that the use of different identification methods by researchers may lead to the study of populations that are, effectively, *different*, while White et al. (2018) proposed that '(m) easurement inconsistencies . . . (may) lead to variation in the number and type of students participating in gifted education research, which affects the comparability, validity and applicability of research findings' (p. 56).

To address the problem, Reis and McCoach (2000) proposed 'an imperfect, yet workable operational definition' of gifted underachievement to be 'a severe discrepancy between expected achievement (as measured by cognitive/intellectual ability assessments or standardized achievement test scores) and actual achievement (as measured by class grades and teacher evaluations)' (p. 157). It is noted that this discrepancy must not be due to a disability (e.g., specific learning disability) and must last for an extended period of time (Reis & McCoach, 2000; Ridgley, Rubenstein, & Callan, 2020). The definition appears to have achieved wide acceptance in the identification of gifted underachievement today (Gilar-Corbi et al., 2019; Mofield & Parker Peters, 2019; Steenbergen-Hu et al., 2020). Nevertheless, substantial inconsistencies continue to exist in the manner in which expected and actual achievement are compared to establish gifted underachievement.

At present, four types of methods appear to be commonly used to identify underachievement in gifted students (McCoach & Siegle, 2014; Steenbergen-Hu et al., 2020). In order of popularity, they are as follows:

- a. The absolute split ('ABS') method (e.g., Vlahovic-Stetic, Vidovic, & Arambasic, 1999): Gifted underachievement is deemed when the level of expected achievement is above the threshold for giftedness (e.g., top 10%), and the level of actual achievement is below the threshold for poor achievement (e.g., below average);
- b. The nomination ('NOM') method (e.g., Cavilla, 2015): Gifted underachievement is deemed when a student is nominated as exhibiting gifted underachievement on the basis of the observations of nominees (e.g., teachers);
- c. The simple difference ('SDF') method (e.g., Lau & Chan, 2001): Gifted underachievement is deemed when the standardized level of actual achievement is substantially below the standardized level of expected achievement (i.e., commonly a difference of at least one standard deviation), for a student identified as being gifted (e.g., the level of expected achievement is above the threshold for giftedness); and
- d. The regression ('REG') method (e.g., Redding, 1990): Gifted underachievement is deemed when the level of expected achievement, as predicted using the simple linear relationship between the expected and actual achievement scores for the cohort of interest (i.e., identified through regression analysis), is substantially below the level of actual achievement (i.e., commonly a difference of at least one standard error) for a

student identified as being gifted (e.g., the level of expected achievement is above the threshold for giftedness).

These methods of identification of gifted underachievement may be classified according to whether they rely on statistical techniques that compare quantitative assessments of expected and actual achievement (i.e., the absolute split method, the simple difference method, and the regression method), or qualitative judgements of individuals (i.e., the nomination method). Furthermore, as two clear variants of the type of thresholds used with the absolute split method were identified in the literature (i.e., a threshold based on student rank and a threshold based on student raw achievement scores), the absolute split method may be further classified into two separate identification methods (i.e., the absolute split I method [a threshold based on achievement rank] and the absolute split II method [a threshold based on raw achievement scores]).

To date, three empirical attempts have been made to compare the various methods used to identify gifted underachievement. First, Annesley, Odhner, Madoff, and Chansky (1970) investigated a number of these methods (i.e., the simple difference method, the regression method, and the nomination method) with a group of 157 first grade students, and concluded that the different methods do not identify the same group of students. Thereafter, Lau and Chan (2001) investigated these methods (i.e., the simple difference method, the regression method, the absolute split I method, and the nomination method) with a group of 126 Grade 7 students. In contrast to Annesley et al. (1970), Lau and Chan (2001) reached the conclusion that the three statistical methods were in agreement. Most recently, Gilar-Corbi et al. (2019) compared the rates of identification of gifted underachievement among 164 gifted seventh and eighth grade students using two of these methods (i.e., the simple difference method and the regression method), along with the less common Rasch method, and found significant differences in the rates of identification according to the method that was used. A further study by Cheung and Rudowicz (2003), that included both gifted and non-gifted students, investigated underachievement among 2,720 students in Grades 8 and 9 by comparing six variations of the commonly used identification methods to reach the conclusion that these identification methods 'were far from identical' (p. 310). Some possible reasons for the inconsistent findings across these studies may lie in the relatively small sample sizes that were generally utilized, the recruitment of participants from different grade levels, the recruitment of participants from only one or two grade levels, the employment of a restricted combination of expected and actual achievement measurements, and the differing use of statistical techniques to assess convergence among the different identification methods.

It is noteworthy that the problems associated with the identification of underachievement extend beyond the field of gifted education. For example, Smith (2010), in reference to the broad field of education, suggested that 'underachievement is a term over which there is little consensus' (p. 41). Multiple fields outside of education may also be affected, as some methods that have been used to identify underachievement have also been used in the fields of sociology (Smith, 2010), psychology (Preckel, Holling, & Vock, 2006; Tuss, Zimmer, & Ho, 1995), law (Maki, Floyd, & Roberson, 2015; Zirkel & Thomas, 2010), and policy (Sikora & Saha, 2011) on topics including disability (Mayes, Waschbusch, Calhoun, & Mattison, 2019; Van den Broeck, 2002), delinquency (Hoffmann, 2020; Timmermans, vanLier, & Koot, 2009), motivation, child development (Maynard, Waters, & Clement,

2013; McCall, Beach, & Lau, 2000), resilience, ethnicity, social class, and gender gaps (Strand, 2014; Wood, 2003).

Research questions

The main purpose of this study was to produce guidance for researchers and practitioners in the selection of the best method(s) to identify gifted underachievement, to resolve a long-standing concern first expressed over five decades ago (Farquhar & Payne, 1964), and which continues to this day (Desmet et al., 2020; Ridgley et al., 2020). To this end, a rigorous review and assessment of the various methods that are commonly used to identify gifted underachievement was undertaken, with a focus on the validity of the use of these methods (Kane, 2013). The following research questions directed the investigation:

1. Are the different methods commonly used to identify gifted underachievement equivalent?
2. What is the optimal method to identify gifted underachievement?

Method

Selection of identification methods for assessment

Each of the commonly used methods to identify gifted underachievement, with the two common variations of the absolute split method treated as separate methods (i.e., the absolute split I, absolute split II, nomination, simple difference, and regression methods), were assessed in this study. For the statistical identification methods, the thresholds that were adopted to discriminate between achievement and underachievement were those that were commonly adopted in the literature (i.e., a difference of one standard deviation between expected and actual achievement for the simple difference method [Lau & Chan, 2001; White et al., 2018], a difference of one standard error between expected achievement and actual achievement for the regression method [Redding, 1990], a 50th percentile rank as the threshold for 'poor' achievement in the absolute split I method [Staudt & Neubauer, 2006], and a raw score of 80% as the threshold for 'poor' achievement in the absolute split II method). Table 1 provides details on the five identification methods that were assessed.

Table 1. Methods used to identify gifted underachievement

Method of identification	Threshold	Type of method
Absolute split I	Actual achievement < 50 th percentile	Statistical
Absolute split II	Actual achievement < 80% raw score	Statistical
Simple difference	Actual achievement < expected achievement by 1 standard deviation	Statistical
Regression	Actual achievement < expected achievement by 1 standard error of estimate	Statistical
Nomination	Personal judgement of nominator	Nomination

Sample

In recognition of the relatively small sample sizes that were utilized in the three previous studies on the topic with gifted students (Annesley et al., 1970; Gilar-Corbi et al., 2019; Lau & Chan, 2001), a large sample size was targeted for this study. Specifically, archive data were obtained from a co-educational K-12 Independent school located in the south-western suburbs of Sydney, Australia. At the time of data collection, the school, which had been in operation for over three decades, had a total enrolment of over 1,300 students, of which, 53% were male students and 41% had a language background other than English. The school's Index of Community Socio-Educational Advantage, which is a measure of the socio-economic status of the students who attend the school, placed the school on the 70th percentile of Australian schools (ACARA, 2013). In accordance with Gagné (2004, 2009, 2013), whose model of giftedness acknowledges underachievement (Jung, 2022), only data relating to current or former students at the school whose expected achievement level placed them within the top 10% of age peers (and were therefore classifiable as gifted students) were selected for inclusion.

Instruments

Statistical methods

Over the history of the school from which the data were obtained, many different instruments were used, in different periods, and for different grade groups, to assess expected and actual achievement of students. These instruments include the:

- a. Otis–Lennon School Ability Test (OLSAT): A group test of cognitive abilities that produces scores for verbal ability (i.e., Verbal Score [VS]), non-verbal ability (i.e., Non-verbal Score [NV]), and composite ability (i.e., School Ability Index [SAI]).
- b. Higher School Certificate (HSC): The highest educational credential that a student can receive in the state of New South Wales after 13 years of education (i.e., K-12). Students who are awarded the HSC are assessed in multiple self-selected subject areas. Nevertheless, for the purposes of this study, only the results for English and mathematics were utilized, as almost all students complete these subjects.
- c. School Certificate (SC): A recently retired educational credential awarded to students in the state of New South Wales after the completion of Grade 10, and prior to the commencement of the HSC. Students who are awarded the SC complete external examinations in English and mathematics.
- d. National Assessment Program - Literacy and Numeracy (NAPLAN): A series of tests that students sit in Grades 3, 5, 7, and 9, which monitor skills, and provide measurements, in literacy and numeracy.
- e. School assessments (SA): The internal assessments used by the school to determine student progress in mastering the content of courses. For comparability to the other available data, a focus was given to school assessment results in English and mathematics.

The archive data obtained comprised all data for Grade 7 to 12 students at the school for more than ten years. The school has systematically tested all students and has maintained consistent records of multiple measurements of achievement for every student. Specifically, the archives contained OLSAT results for 2,501 students, HSC results for 1,267 students, SC results for 1,207 students, NAPLAN results for 2,599 students, and 367,567 individual SA results. For the purposes of this study, the individual SA results were

combined using school-specified weights to arrive at an overall score for each student in English and mathematics across the junior high school years (Grades 7 to 10; i.e., ‘Junior SA’ for English and mathematics) and senior high school years (Grades 11 to 12; i.e., ‘Senior SA’ for English and mathematics; a total of 6,511 records). Nevertheless, there were some limitations to the data, as each instrument was not administered to all students, some instruments were only administered over a limited number of years (e.g., 2006–2014 for the OLSAT), while other instruments were only administered to students at certain grade levels (e.g., Grades 3, 5, 7, and 9 for the NAPLAN). Consequently, there were differing sample sizes for each combination of instruments used to identify gifted underachievement. Table 2 provides details on the calendar years and grade levels for which the archives contained data for each instrument (e.g., the OLSAT data relates to the performance of Grade 6 students at the school who were administered the instrument from 2006 to 2014).

Nomination method

As the archives did not include any nomination data, such data were newly collected by requesting current teachers at the school from which archive data were accessed to complete a survey. Although nominations may be made by a number of parties (e.g., teachers, parents, peers, and students), data from teachers were collected in this study, as teacher nominations appear to be the most commonly utilized by researchers to identify gifted underachievement (for some exceptions, see Flint, 2002; Lau & Chan, 2001).

No published teacher nomination instruments for the identification of gifted underachievement could be located in the literature. Nevertheless, most researchers appear to require teachers to classify each student in their classes as exhibiting achievement or underachievement on the basis of their observations (Annesley et al., 1970; Cavilla, 2015; Jones & Myhill, 2004; White et al., 2018), with some also requiring reasons for the nominations (Dunne & Gazeley, 2008; Lau & Chan, 2001). In addition, some researchers appear to provide guidance to teachers on gifted underachievement before they are asked to nominate students (Kanevsky & Keighley, 2003). All these elements of teacher nominations were incorporated into the teacher nomination survey that was developed for this study. Specifically, a presentation was made to Grade 7 to 12 teachers at the school on gifted underachievement, following which, the teachers were asked to reflect on their current students (for whom they had just prepared academic reports) and: (a) nominate students in their classes who exhibit gifted underachievement (i.e., defined as ‘achievement significantly below their potential in the past semester’), (b) indicate the classes the students were enrolled in, and (c) provide reasons for their nominations. The nomination instrument was presented as a self-administered online

Table 2. Outline of the administration of instruments used

Instrument	Calendar years of administration (inclusive)	Grades administered
OLSAT	2006–2014	Grade 6
HSC	2002–2014	Grade 12
SC	2001–2011	Grade 10
NAPLAN	2008–2013	Grades 3, 5, 7, 9
SA	2002–2014	Grades 7 to 12

form, which was emailed to all Grade 7 to 12 teachers at the school at the conclusion of the presentation.

The teachers nominated students in each grade across 79 Grade 7 to 12 classes in all subject areas, for an average of 1.4 nominations in each class. None of the 122 teacher nominations that were received needed to be removed from consideration due to the non-provision of reasons for nomination or the provision of inappropriate reasons (Dunne & Gazeley, 2008; Lau & Chan, 2001). Generally, students were nominated due to the observation of some common characteristics of underachieving gifted students (79% of responses) and/or comparisons of the expected and actual achievement of the students (29% of responses). Some examples of student characteristics cited by the participating teachers as indicative of underachievement included a substantial difference in the quality of work produced and verbal contributions in class, procrastination, rushed work, and argumentative task avoidance (among students placed in gifted classes). If the assumption is made that the students who were not nominated are classifiable as exhibiting achievement, the nomination process resulted in a total of 1,505 teacher classifications of gifted students as exhibiting achievement or underachievement.

Data preparation

One hundred and ten possible combinations of expected and actual achievement data existed in the archive data. Nevertheless, to ensure that only meaningful comparisons of expected and actual achievement data were made, and to reduce the total number of data combinations studied to a more manageable size, only those data combinations relating to the same type of 'content' were investigated in the study. For example, combinations of expected achievement in numeracy and actual achievement in mathematics were included, while combinations of expected achievement in numeracy with actual achievement in English were excluded. Moreover, any archive data were only considered to assess expected achievement if the assessment took place prior to the assessment of actual achievement. For example, data from the Higher School Certificate (HSC) could not be used to assess expected achievement as it is the final assessment of a student's achievement before they leave high school. Collectively, this process reduced the number of expected/actual data combinations to 41 (nevertheless, it is noted that NAPLAN data, which could not be converted into raw percentage scores due to the non-publication of maximum possible scores by the administering organization, could not be used to assess the absolute split II method). All such data were converted into standardized units, with a mean of zero and a standard deviation of one.

It is noted that for analysis relating to the nomination method, as the collected data related to multiple subject areas (e.g., a nomination could be made in science rather than English/literacy or mathematics/numeracy), meaningful comparisons to the other identification methods were only possible when the general School Ability Index of the OLSAT was used to assess expected achievement, and a weighted school assessment score in the relevant subject area for the immediately preceding semester (i.e., 'SAR') was used to assess actual achievement, for the other identification methods.

Analytic strategy

Different types of analyses were undertaken to address the two research questions that guided the study.

Equivalence of common methods

To address the first research question on the equivalence of the common identification methods, an assessment was made of the *convergence* of the results obtained from these methods. Following Zaki, Bulgiba, Ismail, and Ismail (2012), who proposed that comparisons of proportions, association, and agreement should all be used as convergence evidence, all three of these approaches were used in this study:

- a. Proportions: An examination of whether similar *percentages* of classifications are made (Ho et al., 2014). If the difference in percentage classifications is large and statistically significant, convergence is not supported.
- b. Association: A measurement of the degree to which two variables are *related* (Lau & Chan, 2001). If the association between variables is small or non-significant, convergence is not supported.
- c. Agreement: A measurement of the degree to which two variables are *equal* (Agresti, 2013; Bland & Altman, 1999; Hanneman, 2008). If the level of agreement is weak, convergence is not supported.

The collective consideration of the three approaches not only allowed for different perspectives to be gained on the convergence of the various identification methods, but also accommodated for the possible limitations associated with making an assessment using only one of the approaches (Bland & Altman, 1999; Zaki et al., 2012). Specifically, comparisons of only the *proportions* of gifted underachievement may be limited, because even if similar proportions of gifted underachievement are obtained for different identification methods, these methods may in fact identify different students as exhibiting gifted underachievement. Similarly, while assessments of *association* may suggest that the classifications obtained from two different identification methods may have a *perfect* relationship, this only indicates that the classification of gifted underachievement obtained using one identification method may be used to *predict* the classification that may be obtained using another identification method (i.e., a high association may not be used to infer that the classification using two different identification methods are the same, or even similar).

An optimal method

To address the second research question on the optimal identification method, the results obtained from the commonly used identification methods were compared to criterion values (i.e., values obtained from another source with more established validity, or a method that more thoroughly assesses the relevant variables to determine the value of interest; Kane, 2006, 2013), using the three above-mentioned approaches. Alonzo and Pepe (1999) referred to the process of obtaining such evidence (i.e., criterion evidence) by making comparisons to criterion values, as the process of making an assessment of *accuracy*.

The methods that are traditionally used to assess the relative merits of identification methods, such as receiver operating characteristic (ROC) curves, could not be used in this study, as not all of the common identification methods under investigation rely on a single continuous variable, and due to the difficulty in arriving at a suitable criterion from which sensitivity and specificity may be directly determined. As a result, latent class analysis was utilized (Collins & Huynh, 2014). Latent class analysis is undertaken in situations where multiple classification methods are used to classify the same variable, the true classification of which is not directly observable (Collins & Huynh, 2014). The procedure

assumes that a true underlying condition (i.e., gifted underachievement) exists, which influences each of the different classification methods, and involves the calculation of an estimate of the most likely 'true' classification (i.e. the criterion estimate) by combining the classifications from each of the different methods. Scholars including Mammadov, Ward, Cross, and Cross (2016) have noted the possible benefits of this methodological tool with high ability populations.

Meta-analyses

The statistical techniques used in this investigation produced results for each of the 41 combinations of expected achievement and actual achievement data, and for each possible pair of methods used to identify gifted underachievement. As such a large volume of results may lead to difficulty in the extraction of a single meaningful conclusion about each identification method, meta-analyses were conducted to produce an overall weighted average estimate of an effect size for each identification method. Specifically, the meta-analysis procedures outlined by Borenstein, Hedges, Higgins, and Rothstein (2009) were followed, whereby a weighted average of 41 raw effect sizes (from the 41 data combinations) was calculated, using the inverse of the variances of each effect size as weights. Lin and Sullivan (2009, p. 870) has recognized that such use of 'the usual weights based on inverse variances' is appropriate in meta-analyses of summary results of this nature involving overlapping subjects. The use of these procedures is consistent with the recommendation by Steenbergen-Hu and Olszewski-Kubilius (2016) to use meta-analytical techniques to increase the precision and reliability of research findings.

Results

Equivalence of identification methods

Convergence evidence was collected, using three approaches (i.e., comparisons of the proportions, association, and agreement of the commonly used identification methods) to ascertain the equivalence of the commonly used methods to identify gifted underachievement.

Proportions

As a first step, the proportion of students exhibiting gifted underachievement, as assessed using the 41 different combinations of expected/actual achievement data for each identification method, was determined. Thereafter, the differences in each of these proportions of gifted underachievement, for all possible pairs of the five identification methods, were calculated. A test that is commonly used to assess whether any difference between two sets of dichotomous classifications of the same group of individuals is statistically significant (i.e., McNemar's test; Agresti, 2013; Roberts, Sheffield, McIntire, & Alexander, 2011; Tang, He, & Tu, 2012) was used to establish whether any of these differences in proportions were statistically significant. The results of the analyses are outlined in Table 3 (e.g., in the first row, the differences in proportions between each pair of identification methods were calculated using OLSAT School Ability Index [SAI] data as the measure of expected achievement, and NAPLAN Literacy results data as the measure of actual achievement). Positive values in the table indicate that the first specified method

Table 3. Difference in proportions of gifted underachievement and mcnemar test results

Data Combination (Expected–Actual Achievement)	ABSII/ ABSII	ABSII/ SDF	ABSII/ REG	ABSII/ NOM	ABSII/ SDF	ABSII/ REG	ABSII/ NOM	SDF/ REG	SDF/ NOM	REG/ NOM
OLSAT SAI–NAPLAN Lit	-	-0.24*	-0.10*	-	-	-	-	0.12*	-	-
OLSAT SAI–NAPLAN Num	-	-0.09*	-0.11*	-	-	-	-	-0.02*	-	-
OLSAT VS–NAPLAN Lit	-	-0.20*	-0.10*	-	-	-	-	0.06*	-	-
OLSAT NV–NAPLAN Num	-	-0.16*	-0.10*	-	-	-	-	0.06*	-	-
OLSAT SAI–SC E	-0.10	-0.15*	-0.13*	-	0.05	0.03	-	-0.03	-	-
OLSAT SAI–SC M	-0.13*	-0.05	-0.08	-	-0.08	-0.05	-	0.03	-	-
OLSAT VS–SC E	-0.04	-0.13	-0.09	-	0.09	0.04	-	0.04	-	-
OLSAT NV–SC M	-0.13*	-0.13*	-0.07	-	0.00	-0.07	-	-0.07	-	-
OLSAT SAI–HSC E	0.02	-0.20*	0.45*	-	-0.22*	0.44*	-	-0.65*	-	-
OLSAT SAI–HSC M	0.04	-0.25*	0.50*	-	-0.29*	0.46*	-	0.75*	-	-
OLSAT VS–HSC E	0.03	-0.22*	0.39*	-	-0.25*	0.36*	-	0.61*	-	-
OLSAT NV–HSC M	0.10	-0.31*	0.59*	-	-0.41*	0.48*	-	0.90*	-	-
NAPLAN Lit–SC E	-0.05	-0.13*	-0.18*	-	-0.08	-0.13*	-	-0.05	-	-
NAPLAN Num–SC M	-0.03	-0.08*	-0.07*	-	-0.05	-0.04	-	0.01	-	-
NAPLAN Lit–HSC E	0.00	-0.41*	0.34*	-	-0.41*	0.34*	-	0.76*	-	-
NAPLAN Num–HSC M	0.07	-0.24*	0.56*	-	-0.31*	0.49	-	0.80*	-	-
SC E–HSC E	-0.01	-0.26*	0.25*	-	-0.25*	0.26*	-	0.51*	-	-
SC M–HSC M	-0.01	-0.30*	0.47*	-	-0.29*	0.48*	-	0.77*	-	-
OLSAT SAI–Junior SA E	-0.64*	-0.25*	-0.01	-	0.39*	0.63*	-	0.24*	-	-
OLSAT SAI–Junior SA M	-0.28*	-0.30*	-0.08*	-	-0.01	0.20*	-	0.22*	-	-
OLSAT VS–Junior SA E	-0.61*	-0.20*	-0.02	-	0.41*	0.59*	-	0.17*	-	-
OLSAT NV–Junior SA M	-0.31*	-0.35*	-0.04*	-	-0.03	0.27*	-	0.30*	-	-
OLSAT SAI–Senior SA E	-0.49*	-0.28*	0.17*	-	0.21*	0.67*	-	0.45*	-	-
OLSAT SAI–Senior SA M	-0.40*	-0.35*	0.07*	-	0.05	0.47*	-	0.41*	-	-
OLSAT VS–Senior SA E	-0.51*	-0.24*	0.16*	-	0.27*	0.67*	-	0.40*	-	-
OLSAT NV–Senior SA M	-0.43*	-0.35*	0.08*	-	0.09	0.51*	-	0.42*	-	-
NAPLAN Lit–Junior SA E	-0.52*	-0.09*	-0.09*	-	0.43*	0.44*	-	0.01	-	-

Continued

Table 3. (Continued)

Data Combination (Expected–Actual Achievement)	ABS I/		ABS II/		ABS I/		ABS II/		ABS I/		ABS II/		SDF/		SDF/		REG/		REG/		
	ABS I/	ABS II/	REG	NOM	SDF	REG	NOM	SDF	REG	NOM	SDF	REG	NOM	REG	NOM	SDF	REG	NOM	SDF	REG	NOM
NAPLAN Num–Junior SA M	-0.31*	-0.44*	-0.14*	-	-0.13*	0.18*	-	-0.13*	0.18*	-	0.30*	-	-	0.30*	-	-	-	-	-	-	-
NAPLAN Lit–Senior SA E	-0.52*	-0.24*	0.08*	-	0.28*	0.60*	-	0.28*	0.60*	-	0.32*	-	-	0.32*	-	-	-	-	-	-	-
NAPLAN Num–Senior SA M	-0.40*	-0.41*	0.05*	-	-0.02	0.45*	-	-0.02	0.45*	-	0.47*	-	-	0.47*	-	-	-	-	-	-	-
Junior SA E–SC E	-0.01	0.00	-0.11*	-	0.01	-0.09*	-	0.01	-0.09*	-	-0.11*	-	-	-0.11*	-	-	-	-	-	-	-
Junior SA M–SC M	-0.06	-0.01	-0.06	-	0.04	0.00	-	0.04	0.00	-	-0.04	-	-	-0.04	-	-	-	-	-	-	-
SC E–Senior SA E	-0.32*	-0.20*	0.03	-	0.12*	0.34*	-	0.12*	0.34*	-	0.22*	-	-	0.22*	-	-	-	-	-	-	-
SC M–Senior SA M	-0.47*	-0.28*	-0.02	-	0.19*	0.46*	-	0.19*	0.46*	-	0.27*	-	-	0.27*	-	-	-	-	-	-	-
Junior SA E–HSC E	0.00	-0.33*	0.22*	-	-0.33*	0.22*	-	-0.33*	0.22*	-	0.54*	-	-	0.54*	-	-	-	-	-	-	-
Junior SA M–HSC M	0.04	-0.24*	0.47*	-	-0.29*	0.43*	-	-0.29*	0.43*	-	0.71*	-	-	0.71*	-	-	-	-	-	-	-
Senior SA E–HSC E	-0.02	-0.28*	-0.09	-	-0.26*	-0.07	-	-0.26*	-0.07	-	0.20*	-	-	0.20*	-	-	-	-	-	-	-
Senior SA M–HSC M	-0.05	-0.28*	0.15*	-	-0.23*	0.21*	-	-0.23*	0.21*	-	0.44*	-	-	0.44*	-	-	-	-	-	-	-
Junior SA E–Senior SA E	-0.46*	-0.10*	0.00	-	0.36*	0.45*	-	0.36*	0.45*	-	0.09*	-	-	0.09*	-	-	-	-	-	-	-
Junior SA M–Senior SA M	-0.40*	-0.10*	0.01	-	0.32*	0.42*	-	0.32*	0.42*	-	0.09*	-	-	0.09*	-	-	-	-	-	-	-
OLSAT SAI–SAR	-0.27*	-0.26*	0.06*	-0.21*	-0.01	0.34*	-0.02	-0.01	0.34*	-0.02	0.32*	0.03	-	0.32*	0.03	-	-	-	-	-	-
Weighted average	-0.29*	-0.23*	0.00	-0.21*	0.04*	0.34*	-0.02	0.04*	0.34*	-0.02	0.22*	0.03	-	0.22*	0.03	-	-	-	-	-	-
Weighted average of absolute values	0.29*	0.23*	0.12*	0.21*	0.18*	0.35*	0.02	0.18*	0.35*	0.02	0.23*	0.03	-	0.23*	0.03	-	-	-	-	-	-

Note. ABSI = Absolute split I; ABSII = Absolute split II; E = English; Junior SA = School Assessment in Grades 7 to 10; Lit = Literacy; M = Mathematics; NOM = Nomination; Num = Numeracy; NV = Non-Verbal Score; REG = Regression; SAI = School Ability Index; SAR = School Assessment for Relevant Subject; SDF = Simple difference; Senior SA = School Assessment in Grades 11 and 12; VS = Verbal Score.
*p < 0.05.

identified a higher proportion of gifted underachievement cases than the second specified method.

The observed differences in proportions of gifted underachievement classifications, reflecting the 41 combinations of expected and actual achievement data, varied from -0.65 to 0.90 . The McNemar test results indicated that most (73%) of these differences in proportions for each data combination were statistically significant ($p < 0.05$), and therefore non-supportive of convergence between the identification methods. The weighted averages of the differences in the proportions of gifted underachievement (range from -0.29 to 0.34), and the weighted averages of the absolute values of these differences in proportions of gifted underachievement (range from 0.02 to 0.35), as calculated using the approach of Borenstein et al. (2009), were *mostly* statistically significant, and therefore similarly non-supportive of convergence between the identification methods.

According to the weighted average calculations, the exception to these findings was *some* lack of statistically significant difference in the proportions of gifted underachievement cases for certain pairs of identification methods (i.e., absolute split II/nomination, simple difference/nomination, and absolute split I/regression). Therefore, some support was provided for the possibility of convergence between these identification methods. Nevertheless, for the two pairs of methods involving the nomination method, the support only appears to be tentative, as it was not derived from analyses relating to multiple expected achievement/ actual achievement data combinations.

Association

An assessment was made of the level of association between the results obtained from the various methods that identify gifted underachievement, by the calculation of phi coefficients (ϕ) which may be considered to be measures of the strength of the relationship (i.e., association) between two sets of categorical variables (Agresti, 2013; Tang et al., 2012). Phi coefficient values may be interpreted by making a comparison to threshold values. For example, Pett (1997) provides guidelines for weak ($0.00 - 0.29$), low ($0.30 - 0.49$), moderate ($0.50 - 0.69$), strong ($0.70 - 0.89$), and very strong ($0.90 - 1.00$) levels of association. In comparison, Park, Riddle, and Tekian (2014) suggest that a phi coefficient value of greater than 0.70 may be required to establish convergence.

The phi coefficient calculations for the results obtained from all the possible pairings of the identification methods are reported in Table 4. It is noted that only 38 of the 238 measurements of association (i.e., 16%) could be classified as *strong* or *very strong* according to Pett's (1997) criteria (i.e., $\phi > 0.70$). A further 59 (25%) were of *moderate strength* (i.e., $0.50 < \phi < 0.70$), 90 (38%) were of *low strength* (i.e., $0.30 < \phi < 0.50$), and 51 (21%) were of *weak strength* (i.e., $\phi < 0.30$). If Park et al.'s (2014) criteria are used ($\phi > 0.70$), only 16% of the associations appeared to be strong enough to support convergence. The findings were corroborated with the weighted average associations calculated using the approach of Borenstein et al. (2009), all of which were classifiable as being of *weak*, *low*, or *moderate* strength according to Pett's (1997) guidelines, and none of which met Park et al.'s (2014) criteria for convergence.

Generally, the association results did not provide support for the convergence of the results of the five methods that identify gifted underachievement. Nevertheless, two pairs of identification methods (i.e., absolute split I/regression and simple difference/regression) which had the highest weighted average phi coefficient values (i.e., 0.59 and 0.58 , respectively) may be considered to be approaching convergence. The weakest

Table 4. Phi coefficients of association between identification methods

Data Combination (Expected-Actual Achievement)	ABSII		ABSII/		ABSII/		ABSII/		ABSII/		SDF/		SDF/		REG/	
	ABSII	ABSII	SDF	REG	NOM	REG	NOM	SDF	REG	NOM	SDF	REG	NOM	SDF	REG	NOM
OLSAT SAI-NAPLAN Lit	-	0.34	0.34	0.49	-	-	-	-	-	-	0.70	-	-	-	-	-
OLSAT SAI-NAPLAN Num	-	0.41	0.41	0.38	-	-	-	-	-	-	0.92	-	-	-	-	-
OLSAT VS-NAPLAN Lit	-	0.34	0.34	0.43	-	-	-	-	-	-	0.79	-	-	-	-	-
OLSAT NV-NAPLAN Num	-	0.36	0.36	0.45	-	-	-	-	-	-	0.78	-	-	-	-	-
OLSAT SAI-SC E	0.00	0.00	0.00	0.00	-	0.79	0.88	-	-	-	0.90	-	-	-	-	-
OLSAT SAI-SC M	0.50	0.69	0.69	0.61	-	0.72	0.82	-	-	-	0.88	-	-	-	-	-
OLSAT VS-SC E	0.00	0.00	0.00	0.00	-	0.55	0.69	-	-	-	0.34	-	-	-	-	-
OLSAT NV-SC M	0.46	0.46	0.46	0.61	-	0.39	0.76	-	-	-	0.58	-	-	-	-	-
OLSAT SAI-HSC E	0.96	0.60	0.60	0.35	-	0.58	0.36	-	-	-	0.21	-	-	-	-	-
OLSAT SAI-HSC M	0.92	0.49	0.49	0.29	-	0.45	0.32	-	-	-	0.14	-	-	-	-	-
OLSAT VS-HSC E	0.94	0.58	0.58	0.42	-	0.55	0.44	-	-	-	0.24	-	-	-	-	-
OLSAT NV-HSC M	0.80	0.26	0.26	0.20	-	0.21	0.25	-	-	-	0.05	-	-	-	-	-
NAPLAN Lit-SC E	0.00	0.00	0.00	0.00	-	0.61	0.50	-	-	-	0.82	-	-	-	-	-
NAPLAN Num-SC M	0.70	0.48	0.48	0.52	-	0.50	0.54	-	-	-	0.93	-	-	-	-	-
NAPLAN Lit-HSC E	1.00	0.37	0.37	0.37	-	0.37	0.37	-	-	-	0.14	-	-	-	-	-
NAPLAN Num-HSC M	0.86	0.40	0.40	0.26	-	0.34	0.31	-	-	-	0.10	-	-	-	-	-
SC E-HSC E	0.91	0.58	0.58	0.54	-	0.60	0.52	-	-	-	0.31	-	-	-	-	-
SC M-HSC M	0.86	0.40	0.40	0.32	-	0.41	0.31	-	-	-	0.13	-	-	-	-	-
OLSAT SAI-Junior SA E	0.21	0.48	0.48	0.94	-	0.43	0.22	-	-	-	0.51	-	-	-	-	-
OLSAT SAI-Junior SA M	0.37	0.36	0.36	0.65	-	0.64	0.53	-	-	-	0.56	-	-	-	-	-
OLSAT VS-Junior SA E	0.21	0.50	0.50	0.78	-	0.37	0.24	-	-	-	0.57	-	-	-	-	-
OLSAT NV-Junior SA M	0.39	0.36	0.36	0.67	-	0.51	0.48	-	-	-	0.45	-	-	-	-	-
OLSAT SAI-Senior SA E	0.32	0.56	0.56	0.62	-	0.43	0.20	-	-	-	0.35	-	-	-	-	-
OLSAT SAI-Senior SA M	0.42	0.47	0.47	0.81	-	0.56	0.34	-	-	-	0.38	-	-	-	-	-
OLSAT VS-Senior SA E	0.32	0.61	0.61	0.60	-	0.44	0.19	-	-	-	0.36	-	-	-	-	-
OLSAT NV-Senior SA M	0.39	0.47	0.47	0.79	-	0.40	0.31	-	-	-	0.38	-	-	-	-	-
NAPLAN Lit-Junior SA E	0.18	0.54	0.54	0.55	-	0.34	0.33	-	-	-	0.92	-	-	-	-	-

Continued

Table 4. (Continued)

Data Combination (Expected-Actual Achievement)	ABSII/ ABSII	ABSII/ SDF	ABSII/ REG	ABSII/ NOM	ABSII/ SDF	ABSII/ REG	ABSII/ NOM	ABSII/ SDF	ABSII/ REG	ABSII/ NOM	SDF/ NOM	REG/ NOM
NAPLAN Num-Junior SA M	0.23	0.17	0.37	-	0.49	0.49	-	0.47	0.47	-	-	-
NAPLAN Lit-Senior SA E	0.30	0.53	0.68	-	0.33	0.20	-	0.36	0.36	-	-	-
NAPLAN Num-Senior SA M	0.42	0.40	0.79	-	0.45	0.35	-	0.34	0.34	-	-	-
Junior SA E-SC E	0.00	0.00	0.00	-	0.00	0.34	-	0.00	0.00	-	-	-
Junior SA M-SC M	0.00	0.00	0.00	-	0.49	0.47	-	0.49	0.49	-	-	-
SC E-Senior SA E	0.48	0.61	0.91	-	0.68	0.43	-	0.55	0.55	-	-	-
SC M-Senior SA M	0.34	0.50	0.77	-	0.35	0.27	-	0.53	0.53	-	-	-
Junior SA E-HSC E	0.94	0.50	0.45	-	0.50	0.45	-	0.22	0.22	-	-	-
Junior SA M-HSC M	0.84	0.56	0.27	-	0.52	0.29	-	0.15	0.15	-	-	-
Senior SA E-HSC E	0.88	0.40	0.67	-	0.46	0.58	-	0.60	0.60	-	-	-
Senior SA M-HSC M	0.75	0.54	0.38	-	0.62	0.51	-	0.31	0.31	-	-	-
Junior SA E-Senior SA E	0.19	0.81	0.86	-	0.40	0.22	-	0.55	0.55	-	-	-
Junior SA M-Senior SA M	0.28	0.65	0.89	-	0.44	0.25	-	0.57	0.57	-	-	-
OLSAT SAI-SAR	0.56	0.58	0.82	0.18	0.67	0.48	0.33	0.49	0.49	0.34	0.13	0.13
Weighted average	0.50	0.42	0.59	0.18	0.48	0.39	0.33	0.58	0.58	0.34	0.13	0.13

Note. ABSI = Absolute split I; ABSII = Absolute split II; E = English; Junior SA = School Assessment in Grades 7 to 10; Lit = Literacy; M = Mathematics; NOM = Nomination; Num = Numeracy; NV = Non-Verbal Score; REG = Regression; SAI = School Ability Index; SAR = School Assessment for Relevant Subject; SDF = Simple difference; Senior SA = School Assessment in Grades 11 and 12; VS = Verbal Score.

levels of convergence appeared to exist between the results obtained from the nomination method and the statistical identification methods (i.e., phi coefficient values ranged from 0.13 to 0.34).

Agreement

An assessment was made of the level of agreement between the results obtained from the methods that identify gifted underachievement (i.e., the identification of the same students as exhibiting gifted underachievement) by making use of a statistical tool that is commonly adopted in the health fields to assess agreement between multiple methods that diagnose a health condition (Ewe et al., 2013). The Cohen's kappa (κ) statistic is the most popular tool used to measure the degree of agreement between classification tasks and has an advantage over other tools in that it accounts for agreement by chance factors (Agresti, 2013; Sim & Wright, 2005). It is noted that multiple scholars have proposed varying guidelines for the interpretation of Cohen's kappa values (Kundel & Polansky, 2003; Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009; Walts et al., 2011). Of these, the guidelines proposed by Walts et al. (2011) were selected, as they appeared to represent a useful, 'intermediate' position among the range of existing guidelines. Consequently, a Cohen's kappa statistic of 0.75 was deemed to be necessary for convergence to be supported between any pair of identification methods.

The Cohen's kappa values for the results obtained from all possible pairings of the identification methods are reported in Table 5. Only 14% of the Cohen's kappa values were found to be greater than the Walts et al. (2011) threshold for excellent agreement ($\kappa = 0.75$), while *none* of the weighted average Cohen's kappa values calculated using the approach of Borenstein et al. (2009) were above this threshold. Therefore, strong support could not be found for convergence between the various methods that identify gifted underachievement from an evaluation of the agreement of the results obtained from the various identification methods. In particular, the level of agreement between the nomination method and the other statistical methods was consistently less than may be expected by chance alone ($\kappa < 0.00$). This suggested that the nomination method, consistent with the findings of Lau and Chan (2001), may have the least agreement with the other methods used to identify gifted underachievement.

For comparison with the Cohen's kappa values, Table 6 outlines the raw percentage agreement values for each possible pairing of the five investigated identification methods. Weighted average calculations indicated that raw percentage agreement ranged from 0.36 to 0.91 for different pairs of identification methods. Nevertheless, as raw percentage agreement values do not account for agreement by chance factors, these findings should not be given substantial weight in assessments of the agreement between identification methods.

Overall assessment of the convergence evidence

While convergence did appear to exist for some of the identification methods with some specific data combinations, the collected evidence indicated that convergence was generally *not* supported. Indeed, consistent evidence across the multiple approaches used to assess convergence could not be found for *any* pair of identification methods.

Table 5. Cohen's kappa statistics between identification methods

Data Combination (Expected-Actual Achievement)	ABSII/ ABSII	ABSII/ SDF	ABSII/ REG	ABSII/ NOM	ABSII/ SDF	ABSII/ REG	ABSII/ NOM	SDF/ REG	SDF/ NOM	REG/ NOM
OLSAT SAI-NAPLAN Lit	-	0.21	0.38	-	-	-	-	0.66	-	-
OLSAT SAI-NAPLAN Num	-	0.29	0.25	-	-	-	-	0.91	-	-
OLSAT VS-NAPLAN Lit	-	0.20	0.31	-	-	-	-	0.77	-	-
OLSAT NV-NAPLAN Num	-	0.23	0.33	-	-	-	-	0.76	-	-
OLSAT SAI-SC E	0.00	0.00	0.00	-	0.77	0.87	-	0.89	-	-
OLSAT SAI-SC M	0.40	0.64	0.54	-	0.69	0.80	-	0.87	-	-
OLSAT VS-SC E	0.00	0.00	0.00	-	0.47	0.65	-	0.33	-	-
OLSAT NV-SC M	0.36	0.36	0.54	-	0.39	0.73	-	0.56	-	-
OLSAT SAI-HSC E	0.96	0.53	0.22	-	0.50	0.23	-	0.08	-	-
OLSAT SAI-HSC M	0.91	0.38	0.16	-	0.33	0.19	-	0.04	-	-
OLSAT VS-HSC E	0.94	0.51	0.29	-	0.46	0.32	-	0.11	-	-
OLSAT NV-HSC M	0.79	0.13	0.08	-	0.08	0.11	-	0.01	-	-
NAPLAN Lit-SC E	0.00	0.00	0.00	-	0.54	0.40	-	0.80	-	-
NAPLAN Num-SC M	0.65	0.37	0.42	-	0.46	0.51	-	0.93	-	-
NAPLAN Lit-HSC E	1.00	0.24	0.24	-	0.24	0.24	-	0.04	-	-
NAPLAN Num-HSC M	0.85	0.27	0.13	-	0.21	0.17	-	0.02	-	-
SC E-HSC E	0.91	0.51	0.45	-	0.53	0.43	-	0.18	-	-
SC M-HSC M	0.86	0.27	0.18	-	0.28	0.18	-	0.03	-	-
OLSAT SAI-Junior SA E	0.08	0.37	0.94	-	0.32	0.09	-	0.41	-	-
OLSAT SAI-Junior SA M	0.24	0.23	0.59	-	0.64	0.46	-	0.47	-	-
OLSAT VS-Junior SA E	0.08	0.40	0.78	-	0.26	0.11	-	0.49	-	-
OLSAT NV-Junior SA M	0.26	0.23	0.66	-	0.51	0.37	-	0.33	-	-
OLSAT SAI-Senior SA E	0.19	0.48	0.55	-	0.37	0.08	-	0.21	-	-
OLSAT SAI-Senior SA M	0.30	0.36	0.79	-	0.56	0.20	-	0.25	-	-
OLSAT VS-Senior SA E	0.19	0.54	0.52	-	0.36	0.07	-	0.23	-	-
OLSAT NV-Senior SA M	0.27	0.37	0.77	-	0.39	0.18	-	0.25	-	-
NAPLAN Lit-Junior SA E	0.07	0.45	0.47	-	0.21	0.20	-	0.92	-	-

Continued

Table 5. (Continued)

Data Combination (Expected–Actual Achievement)	ABSII		ABSII/		ABSII/		ABSII/		ABSII/		SDF/		SDF/		REG/		
	ABSII	ABSII	SDF	REG	NOM	SDF	REG	NOM	SDF	REG	NOM	SDF	REG	NOM	SDF	REG	NOM
NAPLAN Num–Junior SA M	0.10	0.06	0.06	0.24	-	0.47	0.43	-	0.36	-	-	-	-	-	-	-	-
NAPLAN Lit–Senior SA E	0.16	0.44	0.44	0.63	-	0.28	0.08	-	0.23	-	-	-	-	-	-	-	-
NAPLAN Num–Senior SA M	0.30	0.28	0.28	0.78	-	0.45	0.22	-	0.20	-	-	-	-	-	-	-	-
Junior SA E–SC E	0.00	N	0.00	0.00	-	0.00	0.20	-	0.00	-	-	-	-	-	-	-	-
Junior SA M–SC M	0.00	0.00	0.00	0.00	-	0.39	0.47	-	0.39	-	-	-	-	-	-	-	-
SC E–Senior SA E	0.37	0.54	0.54	0.91	-	0.66	0.32	-	0.47	-	-	-	-	-	-	-	-
SC M–Senior SA M	0.21	0.40	0.40	0.77	-	0.33	0.17	-	0.43	-	-	-	-	-	-	-	-
Junior SA E–HSC E	0.94	0.40	0.40	0.33	-	0.40	0.33	-	0.10	-	-	-	-	-	-	-	-
Junior SA M–HSC M	0.84	0.48	0.48	0.14	-	0.42	0.16	-	0.04	-	-	-	-	-	-	-	-
Senior SA E–HSC E	0.88	0.28	0.28	0.62	-	0.34	0.56	-	0.53	-	-	-	-	-	-	-	-
Senior SA M–HSC M	0.75	0.46	0.46	0.33	-	0.55	0.41	-	0.18	-	-	-	-	-	-	-	-
Junior SA E–Senior SA E	0.07	0.37	0.37	0.85	-	0.27	0.09	-	0.46	-	-	-	-	-	-	-	-
Junior SA M–Senior SA M	0.15	0.59	0.59	0.88	-	0.32	0.12	-	0.49	-	-	-	-	-	-	-	-
OLSAT SAI–SAR	0.48	0.50	0.50	0.81	-0.16	0.67	0.37	-0.32	0.36	-0.10	-0.34	-0.10	-0.34	-0.10	-0.34	-0.34	-0.34
Weighted average	0.33	0.30	0.30	0.46	-0.16	0.40	0.17	-0.32	0.11	-0.10	-0.34	-0.10	-0.34	-0.10	-0.34	-0.34	-0.34

Note. ABSI = Absolute split I; ABSII = Absolute split II; E = English; Junior SA = School Assessment in Grades 7 to 10; Lit = Literacy; M = Mathematics; NOM = Nomination; Num = Numeracy; NV = Non-Verbal Score; REG = Regression; SAI = School Ability Index; SAR = School Assessment for Relevant Subject; SDF = Simple difference; Senior SA = School Assessment in Grades 11 and 12; VS = Verbal Score; κ could not be calculated as no cases of gifted underachievement were identified by either method.

Table 6. Percentage Agreement between Identification Methods

Data Combination (Expected–Actual Achievement)	ABSII/ ABSII	ABSII/ SDF	ABSII/ REG	ABSII/ NOM	ABSII/ SDF	ABSII/ REG	ABSII/ NOM	SDF/ REG	SDF/ NOM	REG/ NOM
OLSAT SAI–NAPLAN Lit	-	0.76	0.88	-	-	-	-	0.88	-	-
OLSAT SAI–NAPLAN Num	-	0.91	0.89	-	-	-	-	0.98	-	-
OLSAT VS–NAPLAN Lit	-	0.83	0.89	-	-	-	-	0.94	-	-
OLSAT NV–NAPLAN Num	-	0.84	0.90	-	-	-	-	0.94	-	-
OLSAT SAI–SC E	0.90	0.85	0.87	-	0.95	0.97	-	0.97	-	-
OLSAT SAI–SC M	0.87	0.95	0.92	-	0.92	0.95	-	0.97	-	-
OLSAT VS–SC E	0.96	0.87	0.91	-	0.91	0.96	-	0.87	-	-
OLSAT NV–SC M	0.87	0.87	0.93	-	0.83	0.93	-	0.89	-	-
OLSAT SAI–HSC E	0.98	0.80	0.55	-	0.78	0.56	-	0.35	-	-
OLSAT SAI–HSC M	0.96	0.75	0.50	-	0.71	0.54	-	0.25	-	-
OLSAT VS–HSC E	0.97	0.78	0.61	-	0.75	0.64	-	0.39	-	-
OLSAT NV–HSC M	0.90	0.69	0.41	-	0.59	0.52	-	0.10	-	-
NAPLAN Lit–SC E	0.95	0.87	0.82	-	0.92	0.87	-	0.95	-	-
NAPLAN Num–SC M	0.97	0.92	0.93	-	0.92	0.93	-	0.99	-	-
NAPLAN Lit–HSC E	1.00	0.59	0.66	-	0.59	0.66	-	0.24	-	-
NAPLAN Num–HSC M	0.93	0.76	0.44	-	0.69	0.51	-	0.20	-	-
SC E–HSC E	0.96	0.74	0.75	-	0.75	0.74	-	0.49	-	-
SC M–HSC M	0.93	0.70	0.53	-	0.71	0.52	-	0.23	-	-
OLSAT SAI–Junior SA E	0.36	0.75	0.99	-	0.61	0.37	-	0.76	-	-
OLSAT SAI–Junior SA M	0.72	0.70	0.92	-	0.84	0.78	-	0.78	-	-
OLSAT VS–Junior SA E	0.39	0.80	0.96	-	0.57	0.41	-	0.83	-	-
OLSAT NV–Junior SA M	0.69	0.65	0.93	-	0.76	0.73	-	0.70	-	-
OLSAT SAI–Senior SA E	0.51	0.72	0.83	-	0.73	0.33	-	0.55	-	-
OLSAT SAI–Senior SA M	0.60	0.65	0.93	-	0.79	0.53	-	0.59	-	-
OLSAT VS–Senior SA E	0.49	0.76	0.84	-	0.69	0.33	-	0.60	-	-
OLSAT NV–Senior SA M	0.57	0.65	0.92	-	0.72	0.49	-	0.58	-	-
NAPLAN Lit–Junior SA E	0.48	0.91	0.91	-	0.57	0.56	-	0.98	-	-

Continued

Table 6. (Continued)

Data Combination (Expected–Actual Achievement)	ABSII		ABSII/		ABSII/		ABSII/		ABSII/		ABSII/		ABSII/		ABSII/		ABSII/			
	ABSII	SDF	REG	NOM	SDF	REG	NOM	SDF	REG	NOM	SDF	REG	NOM	SDF	REG	NOM	SDF	REG	NOM	
NAPLAN Num–Junior SA M	0.69	0.56	0.86	-	0.74	0.78	-	0.70	-	-	-	-	-	-	-	-	-	-	-	
NAPLAN Lit–Senior SA E	0.48	0.76	0.92	-	0.61	0.40	-	0.68	-	-	-	-	-	-	-	-	-	-	-	
NAPLAN Num–Senior SA M	0.60	0.59	0.93	-	0.74	0.55	-	0.53	-	-	-	-	-	-	-	-	-	-	-	
Junior SA E–SC E	0.99	1.00	0.89	-	0.99	0.91	-	0.89	-	-	-	-	-	-	-	-	-	-	-	
Junior SA M–SC M	0.94	0.99	0.94	-	0.96	0.94	-	0.96	-	-	-	-	-	-	-	-	-	-	-	
SC E–Senior SA E	0.68	0.80	0.97	-	0.83	0.66	-	0.78	-	-	-	-	-	-	-	-	-	-	-	
SC M–Senior SA M	0.53	0.72	0.93	-	0.65	0.51	-	0.73	-	-	-	-	-	-	-	-	-	-	-	
Junior SA E–HSC E	0.97	0.67	0.78	-	0.67	0.78	-	0.46	-	-	-	-	-	-	-	-	-	-	-	
Junior SA M–HSC M	0.92	0.76	0.53	-	0.71	0.57	-	0.29	-	-	-	-	-	-	-	-	-	-	-	
Senior SA E–HSC E	0.98	0.72	0.91	-	0.74	0.89	-	0.80	-	-	-	-	-	-	-	-	-	-	-	
Senior SA M–HSC M	0.90	0.72	0.79	-	0.77	0.79	-	0.56	-	-	-	-	-	-	-	-	-	-	-	
Junior SA E–Senior SA E	0.54	0.90	0.99	-	0.64	0.55	-	0.91	-	-	-	-	-	-	-	-	-	-	-	
Junior SA M–Senior SA M	0.60	0.92	0.99	-	0.68	0.58	-	0.91	-	-	-	-	-	-	-	-	-	-	-	
OLSAT SAI–SAR	0.73	0.74	0.93	0.48	0.84	0.66	0.36	0.68	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.51	
Weighted average	0.79	0.81	0.91	0.48	0.77	0.69	0.36	0.85	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.51

Note. ABSI = Absolute split I; ABSII = Absolute split II; E = English; Junior SA = School Assessment in Grades 7 to 10; Lit = Literacy; M = Mathematics; NOM = Nomination; Num = Numeracy; NV = Non-Verbal Score; REG = Regression; SAI = School Ability Index; SAR = School Assessment for Relevant Subject; SDF = Simple difference; Senior SA = School Assessment in Grades 11 and 12; VS = Verbal Score.

An optimal method

After the collection of convergence evidence to assess whether the commonly used methods to identify gifted underachievement may be considered equivalent, criterion evidence was collected to make a determination on an optimal method for the identification of gifted underachievement. As a first step, latent class analysis was undertaken with the entire data set using the *poLCA* statistical package within the *R* software environment (Linzer & Lewis, 2011), to create a single latent class model that provided criterion value estimates (i.e., estimates of the most likely ‘true’ classification of each student as exhibiting gifted achievement or gifted underachievement; Collins & Huynh, 2014). Thereafter, the convergence of the identification results obtained from the various identification methods with the latent class model (the criterion value estimates) was assessed to determine the accuracy of the various identification methods. The three approaches used previously to assess convergence were again used. Table 7 outlines the results, including the difference in the proportions of gifted underachievement (δ), phi coefficient values (ϕ) that indicate the level of association, and Cohen’s kappa values (κ) that indicate the level of agreement.

The initial results demonstrated that the classifications of the simple difference method had the highest degree of convergence with the criterion value estimates. Specifically, the simple difference method produced: (a) the smallest difference in identified proportions of gifted underachievement with the criterion values, (b) association levels with the criterion value estimates that were well above the required threshold for convergence ($\phi > 0.70$; Park et al., 2014; Pett, 1997), and (c) a Cohen’s kappa value that may be classifiable as being in *almost perfect* agreement with the criterion value estimates (Kundel & Polansky, 2003). While the McNemar test did indicate that the difference in the proportions of gifted underachievement identified between the simple difference method and the criterion value estimates was statistically significant, this may not be of practical significance (Wasserstein & Lazar, 2016).

Among the other identification methods, the classifications of the absolute split II method showed the greatest convergence with the criterion value estimates, with a *strong* association equal to Park et al.’s (2014) threshold, and an agreement measure that

Table 7. Accuracy results using a latent class model

Latent class model	Method	Accuracy		
		δ	ϕ	κ
All data	SDF	0.08*	0.86	0.85
	REG	-0.23*	0.59	0.52
	ABSI	-0.23*	0.61	0.54
	ABSII	0.12*	0.70	0.68
	NOM	-0.31*	0.13	0.10
Nomination removed	SDF	0.08*	0.86	0.85
	REG	-0.23*	0.59	0.52
	ABSI	-0.23*	0.61	0.54
	ABSII	0.12*	0.70	0.68

Note. ABSI = Absolute split I; ABSII = Absolute split II; NOM = Nomination; REG = Regression; SDF = Simple difference; δ = difference in proportion; κ = kappa agreement coefficient; ϕ = phi association coefficient.

* $p < 0.05$ from McNemar’s test.

was only 0.07 below the Walts et al. (2011) threshold. The classifications of the absolute split I and regression methods showed more moderate levels of convergence. The least convergent with the criterion value estimates were the classifications of the nomination method. Not only did the nomination method produce the lowest Cohen's kappa and phi coefficient values, it also identified 31% less students as exhibiting gifted underachievement than the latent class model (i.e., the largest difference among all identification methods). Therefore, the initial latent class model provided the least support for the validity of the use of the nomination method.

In recognition of the possibility that the inclusion of any identification method in the latent class analysis may influence the latent class model, the analyses were repeated *without* the nomination method. Table 7 outlines the results of the refined analyses, which were identical to the initial latent class model, and verified that the classifications of the simple difference method had the highest level of convergence with the criterion value estimates.

Overall assessment of criterion evidence

The criterion results supported the validity of the use of only one of the investigated identification methods to identify gifted underachievement (i.e., the simple difference method). Differing levels of support were found for the validity of the use of the other methods. As such, the findings allowed for a ranking of the confidence in the validity of the use of the commonly used methods to identify gifted underachievement (i.e., from the most valid to the least valid: the simple difference, the absolute split II, the absolute split I, the regression, and the nomination methods).

Discussion

The empirical evidence gathered in this investigation suggested that the commonly used methods to identify (i.e., the absolute split I, absolute split II, nomination, regression, and simple difference methods) gifted underachievement may not be considered convergent, and therefore should not be used interchangeably. Indeed, the various identification methods were demonstrated in this study to have differing levels of validity. The strongest evidence for validity was found for the simple difference method, to suggest that this method may be the optimal method to identify gifted underachievement.

The level of empirical support for the simple difference method and the other identification methods should be considered alongside the possible theoretical issues associated with each method. First of all, it is noted that both absolute split methods rely on a fixed expected achievement threshold (i.e., rank or score), which may mean that all gifted students will be treated under these methods as having identical capabilities, possibly leading to the systematic non-identification of underachievement among the most highly gifted students (Reis & McCoach, 2000; Tze, Daniels, & Klassen, 2016). In comparison, concerns have been raised about the regression method as it may be 'logically inconsistent with the concept of underachievement' (Van den Broeck, 2002, p.197). Unlike the other identification methods which compare a student's expected achievement and actual achievement, the regression method compares a student's actual achievement to patterns of achievement of *other* students in a cohort of interest. Finally, scholars have suggested that despite the knowledge of teachers, they appear largely unable to correctly identify cases of gifted underachievement (Dunne & Gazeley, 2008;

Jones & Myhill, 2004; Lau & Chan, 2001), raising questions about the validity of the use of the nomination method.

The simple difference method is free of all of these concerns. Furthermore, it has the advantage of providing a measurement of the *degree* of underachievement of gifted students, that is not possible with either of the absolute split methods or the nomination method, and may represent valuable additional information that may inform decision-making on the most appropriate educational and related provisions for the identified students.

Implications for research and practice

This study may have a number of implications for consideration by researchers. First, as the commonly used methods that identify gifted underachievement have not been established to be equivalent, researchers will need to be consistent in their selection of identification methods, recognize the differences between these methods (and the meaninglessness of making simple comparisons between these methods), and refrain from aggregating the results obtained from these methods. Of the available methods, the simple difference method has been identified to be optimal. The wide adoption of this method may help to resolve one of the longest standing issues in the field of gifted education (Desmet et al., 2020; Dowdall & Colangelo, 1982; Farquhar & Payne, 1964; Reis & McCoach, 2000; Ridgley et al., 2020), and therefore to improve the comparability and validity of studies on gifted underachievement.

With respect to practice, the findings of the study provide clear guidelines on *how specifically* underachievement should be identified in gifted students. Of note, the study suggested that a reliance on teacher nominations may not be particularly valid or tenable. Instead, a systematic approach that utilizes the simple difference method, and involves the assessment of all gifted students with the increasingly sophisticated software packages that are now available in schools to manage student data, may be useful. The resulting information may serve as useful triggers for further investigation, and the provision of appropriate educational and related interventions for the affected students.

It is possible that, while this study was designed to investigate the identification of gifted underachievement, its findings may also have application to the general student population. Indeed, a significant discrepancy between expected and actual achievement, or underachievement, is likely to be a major problem that could affect any student (Smith, 2010; Veas, Gilar, Castejon, & Minano, 2016). Nevertheless, further research may be needed to determine whether any modifications to the identification process recommended in this study may be necessary for students who are not identified as gifted.

The findings of the study may also have application to the proposal by some scholars for a paradigm shift in the field of gifted education, whereby giftedness should not be considered to be a construct of an individual student, but rather a construct of the environment that students are in (Dai & Renzulli, 2008; Harder, Vialle, & Ziegler, 2014; Ziegler, 2005; Ziegler & Phillipson, 2012). Of note, Funk-Werblo (2003) has proposed that underachievement could be assessed at the level of the environment (e.g., schools). Such an approach to the evaluation of the educational 'health' of schools, or the effectiveness of teachers, may have significant advantages over some of the contemporary approaches that are used (e.g., standardized achievement test results), which may be subject to manipulation (e.g., the exclusive enrolment of high achievers), are likely impacted by socio-economic factors (Smith, 2010), and may systematically ignore the

underachievement of gifted students (who may nevertheless achieve highly; Kanevsky & Keighley, 2003).

Areas for further investigation

A number of areas exist for further investigation. First, of the various types of nominations, only teacher nominations were investigated in this study. Therefore, a more thorough examination of the nomination method, which includes data from different nominating parties (e.g., parents, peers, and students), and a greater total number of nominations, may allow for a more complete understanding of the validity of the use of this method. Second, one area that was not investigated in this study was the discrimination of underachievement from normal fluctuations in achievement. While the commonly adopted threshold of one standard deviation between expected and actual achievement levels to deem underachievement under the simple difference method *does* appear to be widely accepted, there is no strong argument for the optimality of this somewhat arbitrary value. Finally, the empirical evidence in this study was obtained using data obtained from gifted students attending a single school. Consequently, replication studies using samples from other schools, school sectors, geographical locations, and countries may be desirable.

Acknowledgements

Open Access Funding provided by University of New South Wales within the CAUL Agreement.

Conflicts of interest

All authors declare no conflict of interest.

Author contribution

Rahmi Luke Jackson: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Writing – original draft; Writing – review & editing. **Jae Yup Jung:** Supervision; Writing – review & editing.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

- Agresti, A. (2013). *Categorical data analysis* (3rd. ed.). Hoboken, NJ: John Wiley & Sons.
- Alonzo, T. A., & Pepe, M. S. (1999). Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Statistics in Medicine*, *18*, 2987–3003.
- Annesley, F., Odhner, F., Madoff, E., & Chansky, N. (1970). Identifying the first grade underachiever. *The Journal of Educational Research*, *63*, 459–462. <https://doi.org/10.1080/00220671.1970.10884065>
- Australian Curriculum, Assessment and reporting Authority (ACARA) (2013). *Guide to Understanding 2013 Index of Community Socio-educational Advantage (ICSEA) values*.

- Retrieved from http://www.acara.edu.au/_resources/Guide_to_understanding_2013_ICSEA_values.pdf
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8, 135–160. <https://doi.org/10.1191/096228099673819272>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, West Sussex, UK: Wiley.
- Cavilla, D. (2015). Observation and analysis of three gifted underachievers in an underserved, urban high school setting. *Gifted Education International*, 33(1), 1–14. <https://doi.org/10.1177/0261429414568181>
- Cheung, C.-K., & Rudowicz, E. (2003). Underachievement and attributions among students attending schools stratified by student ability. *Social Psychology of Education*, 6, 303–323. <https://doi.org/10.1080/00220670309598813>
- Coleman, L. J., & Cross, T. L. (2014). Is being gifted a social handicap? *Journal for the Education of the Gifted*, 37(1), 5–17. <https://doi.org/10.1177/016235328801100406>
- Collins, J., & Huynh, M. (2014). Estimation of diagnostic test accuracy without full verification: A review of latent class methods. *Statistics in Medicine*, 33, 4141–4169. <https://doi.org/10.1002/sim.6218>
- Dai, D. Y., & Renzulli, J. S. (2008). Snowflakes, living systems, and the mystery of giftedness. *Gifted Child Quarterly*, 52, 114–130. <https://doi.org/10.1177/0016986208315732>
- Desmet, O. A., Pereira, N., & Peterson, J. S. (2020). Telling a tale: How underachievement develops in gifted girls. *Gifted Child Quarterly*, 64(2), 85–99.
- Dowdall, C. B., & Colangelo, N. (1982). Underachieving gifted students: Review and implications. *Gifted Child Quarterly*, 26, 179–184. <https://doi.org/10.1177/001698628202600406>
- Dunne, M., & Gazeley, L. (2008). Teachers, social class and underachievement. *British Journal of Sociology of Education*, 29, 451–463. <https://doi.org/10.1080/01425690802263627>
- Ewe, S. H., Delgado, V., van der Geest, R., Westenber, J. J. M., Haeck, M. L. A., Witkowski, T. G., . . . Siebelink, H.-M. (2013). Accuracy of three-dimensional versus two-dimensional echocardiography for quantification of aortic regurgitation and validation by three-dimensional three-directional velocity-encoded magnetic resonance imaging. *The American Journal of Cardiology*, 112, 560–566. <https://doi.org/10.1016/j.amjcard.2013.04.025>
- Farquhar, W. W., & Payne, D. A. (1964). A classification and comparison of techniques used in selecting under- and over-achievers. *Personnel and Guidance Journal*, 42, 874–884. <https://doi.org/10.1002/j.2164-4918.1964.tb04746.x>
- Flint, L. J. (2002). *Self-interventions of gifted underachievers: Stories of success [PhD: Dissertation]*. University of Georgia. Retrieved July 12, 2012, from https://getd.libs.uga.edu/pdfs/flint_lori_j_200208_phd.pdf
- Flint, L. J. (2002). *Self-interventions of gifted underachievers: Stories of success [PhD: Dissertation]*. University of Georgia. Retrieved July 12, 2012, from https://getd.libs.uga.edu/pdfs/flint_lori_j_200208_phd.pdf
- Funk-Werblo, D. (2003). The invisible gifted child. In F. J. Smutny (Ed.), *Underserved gifted populations* (pp. 27–51). Cresskill, NJ: Hampton Press.
- Gagné, F. (2004). Transforming gifts into talents: The DMGT as a developmental theory. *High Ability Studies*, 15, 119–147. <https://doi.org/10.1080/1359813042000314682>
- Gagné, F. (2009). Building gifts into talents: Detailed overview of the DMGT 2.0. In B. MacFarlane & T. Stambaugh (Eds.), *Leading change in gifted education: The festschrift of Dr Joyce VanTassel-Baska* (pp. 61–80). Waco, TX: Prufrock Press.
- Gagné, F. (2013). The DMGT: Changes within, beneath, and beyond. *Talent Development and Excellence*, 5(1), 5–19.
- Gilar-Corbi, R., Veas, A., Miñano, P., & Castejón, J.-L. (2019). Differences in personal, familial, social, and school factors between underachieving and non-underachieving gifted secondary students. *Frontiers in Psychology*, 10, 1–10. <https://doi.org/10.3389/fpsyg.2019.02367>
- Gorard, S., & Smith, E. (2004). What is 'underachievement' at school? *School Leadership & Management*, 24, 205–225. <https://doi.org/10.1080/1363243041000695831>

- Hanneman, S. K. (2008). Design, analysis, and interpretation of method-comparison studies. *AACN Advanced Critical Care*, 19, 223–234. <https://doi.org/10.1097/01.AACN.0000318125.41512.a3>
- Harder, B., Vialle, W., & Ziegler, A. (2014). Conceptions of giftedness and expertise put to the empirical test. *High Ability Studies*, 25(2), 83–120. <https://doi.org/10.1080/13598139.2014.968462>
- Ho, P.-L., Leung, S.-M.-H., Tse, H., Chow, K.-H., Cheng, V.-C.-C., & Que, T.-L. (2014). Novel selective medium for isolation of *Staphylococcus lugdunensis* from wound specimens. *Journal of Clinical Microbiology*, 52, 2633–2636. <https://doi.org/10.1128/JCM.00706-14>
- Hoffmann, J. P. (2020). Academic underachievement and delinquent behavior. *Youth and Society*, 52, 728–755. <https://doi.org/10.1177/0044118X18767035>
- Hsieh, P., Sullivan, J. R., & Guerra, N. S. (2007). A closer look at college students: Self-efficacy and goal orientation. *Journal of Advanced Academics*, 18, 454–476. <https://doi.org/10.4219/jaa-2007-500>
- Jung, J. Y. (2014). Predictors of Attitudes to Gifted Programs/Provisions. *Gifted Child Quarterly*, 58(4), 247–258. <http://dx.doi.org/10.1177/0016986214547636>
- Jung, J. Y., Jackson, R. L., Townend, G., & McGregor, M. (2022). Equity in Gifted Education: The Importance of Definitions and a Focus on Underachieving Gifted Students. *Gifted Child Quarterly*, 66(2), 149–151. <http://dx.doi.org/10.1177/00169862211037945>
- Jones, S., & Myhill, D. (2004). Seeing things differently: Teachers' constructions of underachievement. *Gender and Education*, 16, 531–546. <https://doi.org/10.1080/09540250042000300411>
- Kane, M. T. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: Praeger.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Kanevsky, L., & Keighley, T. (2003). To produce or not to produce? Understanding boredom and the honor in underachievement. *Roeper Review*, 26, 20–28. <https://doi.org/10.1080/02783190309554235>
- Kundel, H. L., & Polansky, M. (2003). Measurement of observer agreement. *Radiology*, 228, 303–308. <https://doi.org/10.1148/radiol.2282011860>
- Lau, K., & Chan, D. W. (2001). Identification of underachievers in Hong Kong: Do different methods select different underachievers? *Educational Studies*, 27, 187–200. <https://doi.org/10.1080/03055690120050419>
- Lin, D.-Y., & Sullivan, P. F. (2009). Meta-analysis of genome-wide association studies with overlapping subjects. *The American Journal of Human Genetics*, 85, 862–872. <https://doi.org/10.1016/j.ajhg.2009.11.001>
- Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software*, 42, 1–29. Retrieved from <http://www.jstatsoft.org/v42/i10/>
- Maki, K. E., Floyd, R. G., & Roberson, T. (2015). State learning disability eligibility criteria: A comprehensive review. *School Psychology Quarterly*, 30, 457–469. <https://doi.org/10.1037/spq0000109>
- Mammadov, S., Ward, T. J., Cross, J. R., & Cross, T. L. (2016). Use of latent profile analysis in studies of gifted students. *Roeper Review*, 38, 175–184. <https://doi.org/10.1080/02783193.2016.1183739>
- Masden, C. A., Leung, O. N., Shore, B. M., Schneider, B. H., & Udvari, S. J. (2015). Social-perspective coordination and gifted adolescents' friendship quality. *High Ability Studies*, 26(1), 3–38. <https://doi.org/10.1080/13598139.2015.1028613>
- Mayes, S. D., Waschbusch, D. A., Calhoun, S. L., & Mattison, R. E. (2019). How common are academic overachievement and underachievement in children with autism or ADHD? *Journal of Developmental and Physical Disabilities*, 2020, 775–783. <https://doi.org/10.1007/s10882-019-09719-8>
- Maynard, T., Waters, J., & Clement, J. (2013). Child-initiated learning, the outdoor environment and the 'underachieving' child. *Early Years*, 33, 212–225. <https://doi.org/10.1080/09575146.2013.771152>

- McCall, R. B., Beach, S. R., & Lau, S. (2000). The nature and correlates of underachievement among elementary schoolchildren in Hong Kong. *Cbild Development*, *71*, 785–801. <https://doi.org/10.1111/1467-8624.00185>
- McCoach, D. B., & Siegle, D. (2014). Underachievers. In J. A. Plucker, & C. M. Callahan (Eds.). *Critical issues and practices in gifted education: What the research says* (2nd ed., pp. 691-706). Waco, TX: Prufrock Press.
- Mofield, E., & Parker, P. M. (2019). Understanding underachievement: Mindset, perfectionism, and achievement attitudes among gifted students. *Journal for the Education of the Gifted*, *42*(2), 107–134. <https://doi.org/10.1177/0162353219836737>
- Park, Y. S., Riddle, J., & Tekian, A. (2014). Validity evidence of resident competency ratings and the identification of problem residents. *Medical Education*, *48*, 614–622. <https://doi.org/10.1111/medu.12408>
- Pett, M. A. (1997). *Nonparametric statistics for health care research: Statistics for small samples and unusual distributions* (6th. ed.). Thousand Oaks, CA: Sage.
- Preckel, F., Holling, H., & Vock, M. (2006). Academic underachievement: Relationship with cognitive motivation, achievement motivation, and conscientiousness. *Psychology in the Schools*, *43*, 401–411. <https://doi.org/10.1002/pits.20154>
- Redding, R. E. (1990). Learning preferences and skill patterns among underachieving gifted adolescents. *Gifted Child Quarterly*, *34*(2), 72–75. <https://doi.org/10.1177/001698629003400204>
- Reis, S. M., & McCoach, D. B. (2000). The underachievement of gifted students: What do we know and where do we go? *Gifted Child Quarterly*, *44*, 152–170. <https://doi.org/10.1177/001698620004400302>
- Rettew, D. C., Lynch, A. D., Achenbach, T. M., Dumenci, L., & Ivanova, M. Y. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research*, *18*, 169–184. <https://doi.org/10.1002/mpr.289>
- Ridgley, L. M., Rubenstein, L. D., & Callan, G. L. (2020). Gifted underachievement within a self-regulated learning framework: Proposing a task-dependent model to guide early identification and intervention. *Psychology in the Schools*, *57*, 1365–1384. <https://doi.org/10.1002/pits.22408>
- Rimm, S. (2003). Underachievement: A national epidemic. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 424–443). Boston, MA: Allyn and Bacon.
- Roberts, S. W., Sheffield, J. S., McIntire, D. D., & Alexander, J. M. (2011). Urine screening for Chlamydia trachomatis during pregnancy. *Obstetrics & Gynecology*, *117*, 883–885. <https://doi.org/10.1097/AOG.0b013e3182107d47>
- Sikora, J., & Saha, L. J. (2011). *Lost talent? The occupational ambitions and attainments of young Australians*. Adelaide, SA: Commonwealth of Australia, Department of Education, Employment and Workplace Relations: National Centre for Vocational Education Research (NCVER).
- Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, *85*, 257–268. <https://doi.org/10.1093/ptj/85.3.257>
- Smith, E. (2010). Underachievement, failing youth and moral panics. *Evaluation and Research in Education*, *23*(1), 37–49. <https://doi.org/10.1080/09500791003605102>
- Staudt, B., & Neubauer, A. C. (2006). Achievement, underachievement and cortical activation: A comparative EEG study of achieving and underachieving adolescents of average and above-average intelligence. *High Ability Studies*, *17*(1), 3–16. <https://doi.org/10.1080/13598130600946855>
- Steenbergen-Hu, S., & Olszewski-Kubilius, P. (2016). How to conduct a good meta-analysis in gifted education. *Gifted Child Quarterly*, *60*, 134–154. <https://doi.org/10.1177/0016986216629545>
- Steenbergen-Hu, S., Olszewski-Kubilius, P., & Calvert, E. (2020). The effectiveness of current interventions to reverse the underachievement of gifted students: Findings of a meta-analysis and

- systematic review. *Gifted Child Quarterly*, 64, 132–165. <https://doi.org/10.1177/0016986220908601>
- Strand, S. (2014). Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'getting it' for the white working class. *Research Papers in Education*, 29, 131–171. <https://doi.org/10.1080/02671522.2013.767370>
- Tang, W., He, H., & Tu, X. M. (2012). *Applied categorical and count data analysis*. Boca Raton, FL: Taylor & Francis.
- Timmermans, M., vanLier, P. A., & Koot, H. M. (2009). Pathways of behavior problems from childhood to late adolescence leading to delinquency and academic underachievement. *Journal of Clinical Child & Adolescent Psychology*, 38, 630–638. <https://doi.org/10.1080/15374410903103502>
- Tuss, P., Zimmer, J., & Ho, H.-Z. (1995). Causal attributions of underachieving fourth-grade students in Cchina, Japan and the United States. *Journal of Cross-Cultural Psychology*, 26, 408–425. <https://doi.org/10.1177/0022022195264006>
- Tze, V. M., Daniels, L. M., & Klassen, R. M. (2016). Evaluating the relationship between boredom and academic outcomes: A meta-analysis. *Educational Psychology Review*, 28(1), 119–144. <https://doi.org/10.1007/s10648-015-9301-y>
- Van den Broeck, W. (2002). The misconception of the regression-based discrepancy operationalization in the definition and research of learning disabilities. *Journal of Learning Disabilities*, 35, 194–204. <https://doi.org/10.1177/002221940203500301>
- Veas, A., Gilar, R., Castejon, J. L., & Minano, P. (2016). Underachievement in Compulsory Secondary Education: a ~ comparison of statistical methods for identification in Spain. *European Journal of Investigation in Health, Psychology and Education*, 6(3), 133–149. <https://doi.org/10.30552/ejihpe.v6i3.175>
- Vlahovic-Stetic, V., Vidovic, V., & Arambasic, L. (1999). Motivational characteristics in mathematical achievement: A study of gifted high-achieving, gifted underachieving and non-gifted pupils. *High Ability Studies*, 10(1), 37–49. <https://doi.org/10.1080/1359813990100104>
- Walts, A. E., Bose, S., Fan, X., Frishberg, D., Scharre, K., de Peralta-Venturina, M., . . . Marchevsky, A. M. (2011). A simplified Bethesda system for reporting thyroid cytopathology using only four categories improves intra- and inter-observer diagnostic agreement and provides non-overlapping estimates of malignancy risks. *Diagnostic Cytopathology*, 40(S1), E62–E68. <https://doi.org/10.1002/dc.21697>
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129–133. <https://doi.org/10.1080/00031305.2016.1154108>
- White, S. L., Graham, L. J., & Blaas, S. (2018). Why do we know so little about the factors associated with gifted underachievement? A systematic literature review. *Educational Research Review*, 24, 55–66. <https://doi.org/10.1016/j.edurev.2018.03.001>
- Wood, E. (2003). The power of pupil perspectives in evidence-based practice: The case of gender and underachievement. *Research Papers in Education*, 18, 365–383. <https://doi.org/10.1080/0267152032000176864>
- Zaki, R., Bulgiba, A., Ismail, R., & Ismail, N. A. (2012). Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: A systematic review. *PLoS One*, 7, e37908. <https://doi.org/10.1371/journal.pone.0037908>
- Ziegler, A. (2005). The actiotope model of giftedness. In R. J. Sternberg, & J. E. Davidson (Eds.), *Conceptions of giftedness* (2nd ed., pp. 411–436). Cambridge, UK: Cambridge University Press.
- Ziegler, A., & Phillipson, S. N. (2012). Towards a systemic theory of gifted education. *High Ability Studies*, 23(1), 3–30. <https://doi.org/10.1080/13598139.2012.679085>
- Zirkel, P. A., & Thomas, L. B. (2010). State laws for RTI: An updated snapshot. *Teaching Exceptional Children*, 42(3), 56–63. <https://doi.org/10.1177/004005991004200306>