



# Assessing Giftedness and Twice-Exceptionality: A Critical Comparison of International Instruments and the Case for a Nationally Standardised CHC-Based Alternative

Sofia Tsakalidou

Aristotle University of Thessaloniki, Greece

## Abstract

*Educational systems seeking to identify gifted and twice-exceptional (2e) students typically rely on imported cognitive-ability instruments that were neither designed nor normed for their national context; Greece, like several other European systems, has no standardised identification pathway. This paper compares widely used instruments, including the WISC-V, CogAT, SB-5 and the adaptive AID 3, and identifies three limitations that recur, to different degrees, across them. First, composite-score fragmentation: global scores can obscure uneven ability profiles, while differences between subtest or index scores may be insufficiently reliable for individual-level interpretation. Second, ceiling effects: fixed-difficulty forms lose precision at the upper end of the distribution, where identification matters most. Third, 2e-blindness: ability-score-based identification may fail to detect high potential when co-occurring difficulties, such as ADHD, affect observed test performance. The paper argues that these limitations stem from a design logic shared across instruments, not from individual tests, and derives from this critique the architecture of METIS (Multi-aspect Excellence and Talent Identification Scale), a proposed digital battery for Greek students in Grades 2–9. Grounded in Cattell–Horn–Carroll (CHC) theory, METIS will comprise five computerised adaptive tests (verbal, quantitative and non-verbal fluid reasoning; verbal and visuospatial working memory), each calibrated on a single scale across grades, so that a strength in one domain can be measured at the upper end even when another domain is weak. The giftedness threshold will be applied to reasoning alone. Item-level response times will be modelled jointly with accuracy to distinguish reasoning ability from response-speed differences, reducing the risk that slow responding is misinterpreted as low ability. On this basis, the paper proposes a candidate masking signature, based on discrepancies among reasoning, working-memory performance and response-time patterns, while emphasising that its diagnostic validity and specificity require independent empirical validation. The paper thus establishes a conceptual and methodological foundation for developing a nationally calibrated, evidence-based identification framework for gifted and twice-exceptional students in Greece and comparable educational systems.*

**Keywords:** *giftedness identification; twice-exceptionality (2e); masking; Cattell–Horn–Carroll (CHC) theory; computerised adaptive testing; response-time modelling*

## 1. Introduction

Responding to the needs of all learners is a core principle of inclusive education, and it applies equally to students with high ability. Twice-exceptional (2e) students combine high intellectual potential with a learning, attentional or other difficulty. For these students, accurate identification is particularly critical, as misidentification can deprive them of both appropriate challenge and necessary support. A recent systematic review indicates that provision for gifted learners in Europe remains uneven and weakly institutionalised [1]. Moreover, many educational systems rely on cognitive-ability instruments developed in other national contexts, whose validity and suitability for local populations may vary. Greece provides an illustrative case. Although a public network of Interdisciplinary Assessment, Counselling and Support Centres (KEDASY) assesses special educational needs (SEN), no comparable national protocol exists for identifying high ability, leaving these students without an equivalent pathway within the public system of educational support. The implications of this gap are visible in everyday teaching, including in foreign-language (FL) classrooms. In a national survey of 152 FL teachers across 24 regional education directorates [2], only 27.6% reported being able to identify high-ability learners, compared with 92.6% for learners with SEN and 62.5% for learners with artistic or athletic talent (talent vs. high ability: McNemar's  $\chi^2(1) = 33.38, p < .001$ ). Teacher characteristics did not predict this disparity. Schools shared students' cognitive profiles with teachers in 65.9% of SEN cases, but in only 6.5% of high-potential cases. These findings suggest a systemic gap, consistent with the



absence of an established identification infrastructure. In this context, identification may depend disproportionately on parental initiative and families' ability to fund private assessment, while referral-based systems risk under-representing students from low-income and minority backgrounds [3]. This paper examines the limitations of widely used cognitive-ability instruments and derives design principles for a nationally calibrated, digital adaptive battery grounded in Cattell–Horn–Carroll (CHC) theory. It addresses three research questions:

**RQ1:** Which recurring limitations of widely used international instruments affect the identification of gifted and, in particular, twice-exceptional (2e) students?

**RQ2:** Do these limitations stem from features of individual tests or from design assumptions shared across instruments?

**RQ3:** Which design principles should guide a nationally calibrated, digital adaptive battery grounded in CHC theory?

## 2. Theoretical Framework

### 2.1 Giftedness, Twice-Exceptionality and the Masking Effect

Within contemporary developmental frameworks of talent development, giftedness can be understood as cognitive potential that may progressively translate into domain-specific achievement as talent develops [4]. The present paper addresses students in Grades 2–9, an age range in which identification should focus on cognitive potential rather than relying primarily on accumulated academic achievement. An instrument designed for this population should therefore assess reasoning abilities as independently as possible of curricular knowledge. This distinction is particularly important in twice-exceptionality. Foley-Nicpon et al. [5] review evidence that high ability can co-occur with specific learning difficulties (SLD), attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD), creating complex profiles in which strengths and difficulties may obscure one another. For example, gifted children with dyslexia may develop compensatory strategies that allow them to read adequately while still underperforming relative to their verbal potential [6]. For gifted students with ADHD, the first 2e profile targeted by the battery proposed in Section 5, masking may arise through three related mechanisms: (a) *score aggregation*, as composite indices can obscure uneven cognitive profiles by combining performance across domains in which ability differs [7], [8]; (b) *compensation*, whereby high cognitive ability may partly offset or conceal executive-functioning weaknesses in observed performance [9]; and (c) *response variability*, whereby attentional lapses may produce inconsistent accuracy, including errors on relatively easy items, and increased variability in response times [10]. These mechanisms may affect both overall scores and item-level response patterns, but their manifestations are not necessarily captured equally well by conventional composite scores. Masking therefore constitutes a distinct measurement challenge: identifying high potential requires examining not only the level of performance but also the configuration of cognitive strengths and weaknesses and the processes underlying observed responses.

### 2.2 CHC Theory as a Reference Taxonomy

The Cattell–Horn–Carroll (CHC) framework is a widely used, empirically supported taxonomy of human cognitive abilities. It organises cognitive abilities across three strata: general intelligence (*g*; Stratum III), broad cognitive abilities (Stratum II), and numerous narrow abilities (Stratum I) [11], [12]. Two features make CHC a suitable reference framework for giftedness and twice-exceptionality (2e) identification. First, it distinguishes fluid reasoning (*Gf*), which involves solving novel problems, from comprehension-knowledge (*Gc*), which reflects acquired knowledge shaped partly by educational and cultural experience. Second, it differentiates broad cognitive abilities associated with working memory (*Gwm*), visual processing (*Gv*) and, in relevant formulations, decision and reaction speed (*Gt*). These distinctions are pertinent to 2e identification because strong reasoning abilities may coexist with relative weaknesses in other cognitive domains [12]. The central question addressed in the following analysis is therefore not simply whether an instrument is CHC-based, but whether its assessment design and scoring procedures permit meaningful interpretation of individual ability profiles.

## 3. Method



The study adopts a critical narrative review design with a structured comparison of cognitive-ability instruments. Instruments were selected according to three criteria: (a) established use or recent development for cognitive assessment relevant to giftedness identification; (b) representation of the main design families, including comprehensive individually administered batteries, group-administered reasoning tests, digital assessment systems and adaptive tests; and (c) availability of independent psychometric evaluations or detailed technical documentation describing test structure and score interpretation. The final set comprised the WISC-V, the Cognitive Abilities Test (CogAT), the Stanford–Binet Intelligence Scales (SB-5), the Reynolds Intellectual Assessment Scales (RIAS-2), the Woodcock–Johnson Tests of Cognitive Abilities (WJ-IV/V), the Adaptive Intelligence Diagnosticum (AID 3) and the Invenio diagnostic system [13]. Each instrument was examined against five criteria derived from the theoretical framework: (a) the underlying model of cognitive abilities and the level at which scores are reported; (b) available evidence concerning the structural validity and interpretability of the reported scores; (c) measurement precision at the upper end of the ability distribution, including the role of adaptive item selection where applicable; (d) the extent to which the assessment design can represent heterogeneous 2e profiles and whether response-time information is incorporated into score interpretation; and (e) the availability of Greek national norms. Instrument selection was purposive and intended to cover major design families rather than provide an exhaustive inventory. Given their different purposes, constructs and administration formats, the instruments were evaluated against common design criteria rather than treated as interchangeable measures. The comparison drew on test manuals, technical documentation and published psychometric studies, with evidence recorded against the five criteria and synthesised qualitatively across instruments. Findings were synthesised across instruments to identify recurring design limitations and distinguish shared assumptions from instrument-specific characteristics.

#### **4. Findings: Three Recurring Limitations**

##### **4.1 Composite-Score Fragmentation**

The first limitation concerns the level at which cognitive ability is reported. Here, *composite-score fragmentation* refers to a recurring tension between a global score that may obscure an uneven ability profile and domain scores whose reliability may be insufficient for confident individual-level interpretation. For students with marked cognitive strengths and weaknesses, a global index such as the WISC-V Full Scale IQ (FSIQ) may obscure information relevant to identification. A student with exceptional verbal and fluid reasoning but relatively weaker working memory or processing speed may therefore obtain an FSIQ that does not adequately represent their strongest abilities [7]. Interpreting index or subtest scores instead of the composite does not necessarily resolve this problem. Research on the WISC-V has reported limited validity for personal strengths-and-weaknesses profiles [14], while difference scores within the RIAS-2 and WISC-V have shown low reliability [15]. Hierarchical factor analyses also raise questions about the distinctiveness of some broad-ability indices when the general factor accounts for a substantial proportion of reliable variance. In CHC-based analyses of the WISC-V, the Fluid Reasoning factor has been reported as largely indistinguishable from *g* [16], while analyses of the WJ-IV similarly highlight the predominance of the general factor and the resulting constraints on interpreting narrower indices [17]. The CogAT illustrates a related but distinct issue. Although it reports verbal, quantitative and non-verbal ability profiles, decisions based on composite cut-offs may fail to reflect substantial differences between these domains. Among 3,062 students scoring in the top stanine on at least two CogAT batteries, only 39% showed a flat profile, while approximately 15% showed an extreme weakness on the third battery [18]. These findings illustrate the extent of profile heterogeneity within a high-performing subgroup, although profile differences alone do not establish the reliability or diagnostic significance of those differences. Taken together, the instruments illustrate a recurring measurement challenge: global scores can conceal meaningful differences between cognitive domains, while interpreting separate scores requires evidence that those scores and their differences are sufficiently reliable and valid for individual-level use. Composite-score fragmentation is therefore best understood as a cross-instrument tension between summarising general ability and representing domain-specific strengths and weaknesses, rather than as an identical psychometric defect shared by every test.

##### **4.2 Ceiling Effects**



The second limitation concerns measurement precision at the upper end of the ability distribution. Fixed-form tests are most precise around their target ability range, with precision declining for examinees whose ability exceeds it substantially [19]. Minton and Pratt [20] found that highly gifted children obtained lower-than-expected SB-5 scores, raising concerns about its discriminative power at the upper end. Non-verbal matrix tests, often favoured for their perceived cultural fairness, raise a related concern: compared with the CogAT, they identified different proportions of gifted students, partly reflecting differences in norm quality, and predicted achievement less well than verbal and quantitative reasoning scores [21]. Adaptive branching, as implemented in the AID 3 [22], can mitigate ceiling effects but requires individual administration by a trained examiner. Moreover, adaptive testing cannot overcome the limits of its item bank [19] or address the other limitations examined here.

### 4.3 2e-Blindness

The third limitation concerns the implications of the first two for twice-exceptional (2e) students. *Masking* refers to a divergence between high reasoning ability and weaker performance in specific domains or conditions. Composite-based scoring may obscure this divergence, while interpreting score discrepancies can be difficult when their reliability is limited [15]. None of the instruments reviewed was designed primarily for 2e identification; recognition of such profiles therefore depends largely on post-hoc clinical judgement. Moreover, in the versions reviewed, none jointly models item-level response times and accuracy. Where speed affects scoring, it is incorporated through timed subtests or time bonuses, leaving the potential contribution of compensation and response variability (Section 2.1) unmodelled. The 2eCAB [23] is a recent related approach, using a decision-tree classifier for cognitive and reading tasks, but it targets dyslexia rather than ADHD and is based on only 15 2e cases. In Greece, a further limitation compounds these concerns. Although Greek adaptations exist for some clinical instruments, notably the WISC-V GR [24] ( $n = 880$ ), these are paper-and-pencil batteries normed and validated for general clinical assessment rather than specifically for identifying gifted and 2e students. Table 1 summarises the documented limitations of the reviewed instruments.

**Table 1.** Documented limitations of the reviewed instruments relevant to gifted and 2e identification

| Instrument              | Adaptive | RT model | 2e profile | Greek norms    | Documented limitation(s)                                                                                                          |
|-------------------------|----------|----------|------------|----------------|-----------------------------------------------------------------------------------------------------------------------------------|
| WISC-V /GR (2014 /2017) | X        | X        | X          | ✓ (paper only) | Limited validity of strength–weakness profiles [14]; Fluid Reasoning $\approx g$ [16]; validity evidence is version-specific [25] |
| Invenio (2022)          | X        | X        | X          | X              | No working-memory or adaptive testing [13]                                                                                        |
| CogAT (2017)            | X        | X        | X          | X              | Composite scores may obscure large discrepancies [18]                                                                             |
| RIAS-2 (2015)           | X        | X        | X          | X              | Low reliability of difference scores [15]                                                                                         |
| AID 3 (2023)            | ✓        | X        | X          | X              | Examiner-administered, one-to-one only [22]                                                                                       |
| SB-5 (2003)             | X        | X        | X          | X              | Lower-than-expected scores in highly gifted children [20]                                                                         |
| WJ-IV (2014)            | X        | X        | X          | X              | Strong general factor limits index interpretation [17]                                                                            |

**Note:** ✓ = documented and implemented; X = not available in the evaluated version, judged against the criteria of this review only. Adaptive: item selection based on the examinee's responses during the test (branched or computerised). RT model: Joint modelling of item-level response times and accuracy. 2e profile: support for 2e identification, not diagnosis. Independent evidence for the WJ V (2025) is not yet available.

## 5. Design Principles for a Nationally Calibrated CHC-Based Alternative

The limitations identified are not independent weaknesses that can be addressed one at a time, but consequences of shared design assumptions. The proposed METIS battery (Multi-aspect Excellence and Talent Identification Scale) will therefore be built on four design principles, each derived from the analysis above. Instead of assessing ability, response behaviour and specialist judgement in separate

steps, METIS will read ability and response times from the same items in two stages (Fig. 1). Adaptive testing, profile reporting and response-time modelling exist separately; METIS proposes their combination, whose added value remains to be demonstrated.

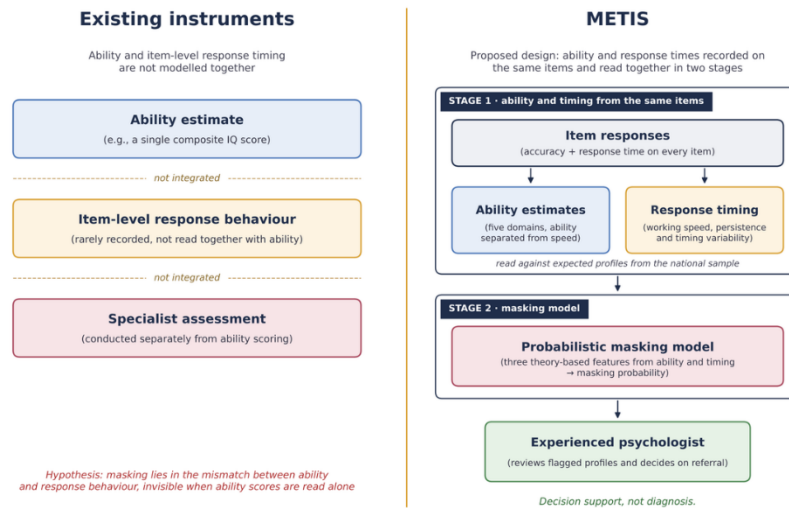


**Fig. 1.** Existing instruments (left) vs. the proposed two-stage METIS design (right).

### 5.1 Profile-Based Reporting Across Five Domains

METIS will assess five domains mapped onto CHC broad (Stratum II) and narrow (Stratum I) abilities (Fig. 2): verbal, non-verbal and quantitative fluid reasoning, since non-verbal tests alone may miss gifted students [21], and verbal and visuospatial working memory, capacities often affected in ADHD [26]. The reasoning domains assess higher-order cognitive potential [4], while the memory domains capture weaknesses that may mask reasoning ability in 2e students [27]. Processing speed will not be a separate domain, as speeded tasks could confound the slowness to be modelled (Section 5.3) with the ability being estimated. Results will be reported as a five-domain profile rather than a single composite, with a separate item bank for each domain. The essential unidimensionality of each bank will be tested before adaptive calibration [28]. A general reasoning score will also be derived from the three reasoning scales. The giftedness threshold will be applied to this score, preventing working-memory weaknesses from lowering it [7]. How this rule handles uneven reasoning profiles will be evaluated empirically. Items will target aptitude rather than achievement, although the verbal domain will draw partly on grade-appropriate vocabulary (Gc).

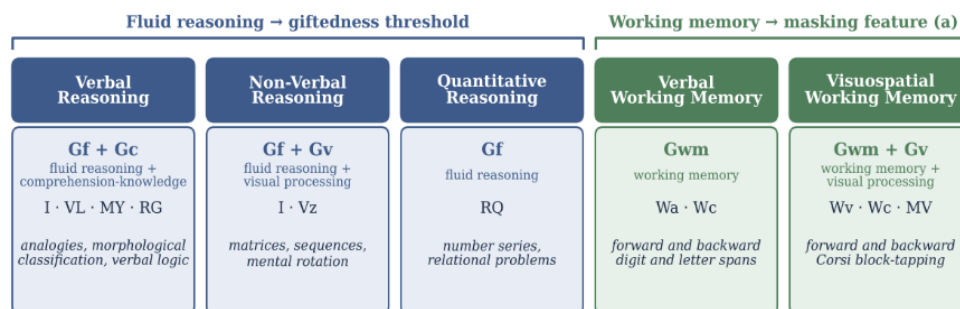


**Fig. 2.** Proposed METIS domains with their CHC broad and narrow abilities and task types.

Fig. 3 illustrates the rationale for profile-based reporting. Two students obtain the same composite score of 105, within the average range. Student A scores consistently near this value. Student B, however, scores 122–132 on the three reasoning scales but only 76 and 67 on working memory.

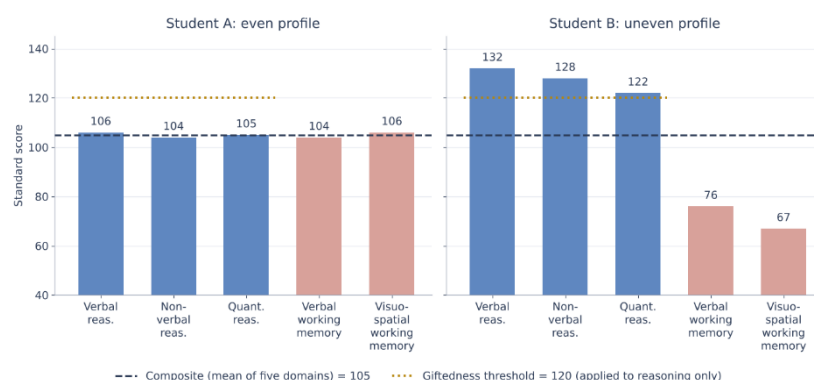


**Fig. 3.** Two profiles with the same five-domain composite (standard scores,  $M = 100$ ,  $SD = 15$ ). Dashed line: composite (mean of five domains); dotted line: giftedness threshold of 120, applied to the reasoning scales only.

The composite obscures these contrasting strengths and weaknesses. A reasoning mean of 127 would exceed the illustrative giftedness threshold of 120, while the working-memory scores would indicate a possible need for support. This example illustrates how a composite may conceal high reasoning ability alongside substantial weaknesses; such a profile may warrant further assessment for 2e but does not by itself establish twice-exceptionality.

## 5.2 Adaptive Administration Beyond Ceiling Effects

METIS will be administered as five computerised adaptive tests, one per domain. After each response, the ability estimate will be updated and the next item selected accordingly. A simulation study of an adaptive version of a cognitive-ability test showed that this approach maintains measurement precision with substantially fewer items [29]. Items across Grades 2–9 will be calibrated and linked to a common scale per domain, subject to evidence of cross-grade comparability, so that grade sets only the starting item and a younger student who reasons at the level of much older peers continues to receive items of corresponding difficulty instead of reaching the ceiling of a grade-level form. Adaptation will operate separately within each domain, so that a weakness in one domain does not restrict access to difficult items in another. Fixed test lengths in the reasoning tests and item-level scoring of memory span trials will prevent an isolated lapse from automatically ending testing prematurely.

## 5.3 Response Behaviour as Information

Digital administration makes it possible to record not only whether an item was answered correctly but also how it was answered, for instance how long the student took to respond. In most instruments such process information is not used for interpretation. METIS will record item-level response times and model them jointly with response accuracy [30], to separate reasoning ability from response speed, so that a slow but able student is not mistaken for a less able one. These latencies will not be interpreted as a measure of a CHC speed factor such as  $G_s$  or  $G_t$ , since response times on complex reasoning items also reflect strategy, persistence and attention. Aberrant combinations of accuracy and response latency are a documented signal in adaptive testing [31]. We hypothesise that, in 2e students, *masking* appears as a structured divergence between reasoning, working memory and response timing, operationalised as three proposed features: (a) working memory markedly below reasoning performance, (b) slow but accurate responding with persistence, and (c) erratic response times. Whether this divergence is stable enough to be detected reliably is the central empirical question METIS is designed to answer. Validation will separate identification, whether METIS detects high ability in students with ADHD, from *masking* detection, whether it detects high ability in students whose earlier assessment fell below the giftedness threshold. An ADHD diagnosis alone will not count as evidence of *masking*; the instrument will support, not replace, the judgement of an experienced psychologist.

## 5.4 National Calibration and Fairness

Finally, METIS will be calibrated on a nationally representative sample of Greek students, instead of being adapted from an instrument developed for another linguistic and cultural context [25]. The same



sample will yield national expected profiles for response timing alongside ability. Item-level fairness across gender, socio-economic background, language background and grade will be examined through differential item functioning (DIF) analyses. Identification decisions are intended to be embedded in a multiple-criteria framework in which the test score is one source of evidence among several [32].

## 6. Discussion And Implications

In answer to RQ1, the comparison identified the three limitations described in Section 4. In answer to RQ2, these are not incidental weaknesses of individual instruments. They follow from a shared design logic that combines limited precision at the upper end with global scores that may obscure domain-specific strengths. A student whose strengths are constrained by ceiling effects and whose weaknesses are absorbed into the composite is the 2e student who remains unidentified. The design principles in Section 5 (RQ3) are therefore intended to operate jointly, not as separate improvements. The validation will test whether response timing adds value beyond ability scores; a null result would be equally informative. For Greece, and for other systems lacking dedicated infrastructure for gifted education [1], a nationally calibrated instrument is not only a psychometric but also an infrastructural improvement. It creates the conditions for a public identification pathway comparable to the one that already exists for SEN. For language teachers the implications are twofold. First, teachers are often the first to notice a discrepancy between a student's reasoning and their performance, for example a learner who grasps grammatical regularities quickly but struggles with spelling or sustained attention. Such observations are valuable as a trigger for referral, not as a substitute for assessment. Second, teachers cannot use profile information they never receive [2]; a profile-based report shared with teachers would allow them to differentiate instruction on the basis of documented strengths as well as difficulties. This study has several limitations. It is a critical review, not an empirical comparison, and the evidence summarised in Table 1 comes from independent studies with different samples and methods. The design principles in Section 5 (RQ3) are theoretically grounded and will be evaluated in forthcoming pilot, calibration, and clinical validation studies, including whether ability plus timing improves detection beyond ability alone; generalisation beyond ADHD remains to be tested.

## 7. Conclusion

The identification of gifted and 2e students in Greece currently relies on instruments not specifically designed for national giftedness and 2e identification and that, to different degrees, obscure uneven profiles through composite scores, may lack precision at the upper end and leave 2e recognition to post-hoc judgement. This paper has argued that these limitations share a common root, and that a nationally calibrated, CHC-based, digital adaptive battery, which would report domain profiles, adapt item difficulty within each domain, apply the giftedness threshold to reasoning alone and treat response behaviour as information, offers a theoretically grounded alternative whose added value requires empirical validation. Its development is an opportunity to build a public, evidence-based identification framework in which high ability is recognised as systematically as other educational needs.

## REFERENCES

- [1] Baccassino F., Pinnelli S., "Giftedness and gifted education: A systematic literature review", *Frontiers in Education*, 7, 2023, 1073007. <https://doi.org/10.3389/feduc.2022.1073007>
- [2] Tsakalidou S., "The gifted paradox in the Greek foreign language classroom: Invisible talent and the motivation trap", Manuscript submitted for publication, 2026.
- [3] Card D., Giuliano L., "Universal screening increases the representation of low-income and minority students in gifted education", *Proceedings of the National Academy of Sciences*, 113(48), 2016, 13678-13683. <https://doi.org/10.1073/pnas.1605043113>
- [4] Subotnik R. F., Olszewski-Kubilius P., Worrell F. C., "Rethinking giftedness and gifted education: A proposed direction forward based on psychological science", *Psychological Science in the Public Interest*, 12(1), 2011, 3-54. <https://doi.org/10.1177/1529100611418056>
- [5] Foley Nicpon M., Allmon A., Sieck B., Stinson R. D., "Empirical investigation of twice-exceptionality: Where have we been and where are we going?", *Gifted Child Quarterly*, 55(1), 2011, 3-17. <https://doi.org/10.1177/0016986210382575>



- [6] van Viersen S., Kroesbergen E. H., Slot E. M., de Bree E. H., "High reading skills mask dyslexia in gifted children", *Journal of Learning Disabilities*, 49(2), 2016, 189-199. <https://doi.org/10.1177/0022219414538517>
- [7] Gilman B. J., Peters D. B., Silverman L. K., Lovecky D. V., Kearney K., Rogers K. B., Amend E. R., Rimm S. B., "Use of the WISC-V for gifted and twice-exceptional identification: Strength-based indexes address discrepant scoring and uninterpretable full scale IQs", *SAGE Open*, 16(1), 2026, 21582440261426072. <https://doi.org/10.1177/21582440261426072>
- [8] Cornoldi C., Giofrè D., Toffalini E., "Cognitive characteristics of intellectually gifted children with a diagnosis of ADHD", *Intelligence*, 97, 2023, 101736. <https://doi.org/10.1016/j.intell.2023.101736>
- [9] Milioni A. L. V., Chaim T. M., Cavallet M., Moleda de Oliveira N., Annes M., dos Santos B., Louzã M., da Silva M. A., Miguel C. S., Serpa M. H., Zanetti M. V., Busatto G., Cunha P. J., "High IQ may "mask" the diagnosis of ADHD by compensating for deficits in executive functions in treatment-naïve adults with ADHD", *Journal of Attention Disorders*, 21(6), 2017, 455-464. <https://doi.org/10.1177/1087054714554933>
- [10] Kofler M. J., Rapport M. D., Sarver D. E., Raiker J. S., Orban S. A., Friedman L. M., Kolomeyer E. G., "Reaction time variability in ADHD: A meta-analytic review of 319 studies", *Clinical Psychology Review*, 33(6), 2013, 795-811. <https://doi.org/10.1016/j.cpr.2013.06.001>
- [11] Schneider W. J., McGrew K. S., "The Cattell-Horn-Carroll theory of cognitive abilities", in Flanagan D. P., McDonough E. M. (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (4th ed.), New York, Guilford Press, 2018, 73-163.
- [12] Flanagan D. P., Dixon S. G., "The Cattell-Horn-Carroll theory of cognitive abilities", in Reynolds C. R., Vannest K. J., Fletcher-Janzen E. (Eds.), *Encyclopedia of Special Education* (4th ed.), Hoboken, NJ, John Wiley & Sons, 2014. <https://doi.org/10.1002/9781118660584.ese0431>
- [13] Jabůrek M., Portešová Š., Straka O., Palíšek P., "Jak identifikovat mimořádně nadané děti objektivně a zábavně? Pomocí diagnostického systému Invenio [How to identify exceptionally gifted children objectively and enjoyably: The Invenio diagnostic system]", *E-psychologie*, 17(2), 2023, 88-93. <https://doi.org/10.29364/epsy.472>
- [14] de Jong P. F., "The validity of WISC-V profiles of strengths and weaknesses", *Journal of Psychoeducational Assessment*, 41(4), 2023, 363-379. <https://doi.org/10.1177/07342829221150868>
- [15] Farmer R. L., Kim S. Y., "Difference score reliabilities within the RIAS-2 and WISC-V", *Psychology in the Schools*, 57(8), 2020, 1273-1288. <https://doi.org/10.1002/pits.22369>
- [16] Canivez G. L., Watkins M. W., Dombrowski S. C., "Structural validity of the Wechsler Intelligence Scale for Children-Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests", *Psychological Assessment*, 29(4), 2017, 458-472. <https://doi.org/10.1037/pas0000358>
- [17] Dombrowski S. C., McGill R. J., Canivez G. L., "Exploratory and hierarchical factor analysis of the WJ-IV Cognitive at school age", *Psychological Assessment*, 29(4), 2017, 394-407. <https://doi.org/10.1037/pas0000350>
- [18] Lohman D. F., Gambrell J., Lakin J., "The commonality of extreme discrepancies in the ability profiles of academically gifted students", *Psychology Science Quarterly*, 50(2), 2008, 269-282.
- [19] Weiss D. J., "Better data from better measurements using computerized adaptive testing", *Journal of Methods and Measurement in the Social Sciences*, 2(1), 2011, 1-27. <https://doi.org/10.2458/v2i1.12351>
- [20] Minton B. A., Pratt S., "Gifted and highly gifted students: How do they score on the SB5?", *Roeper Review*, 28(4), 2006, 232-236. <https://doi.org/10.1080/02783190609554369>
- [21] Lohman D. F., Korb K. A., Lakin J. M., "Identifying academically gifted English-language learners using nonverbal tests: A comparison of the Raven, NNAT, and CogAT", *Gifted Child Quarterly*, 52(4), 2008, 275-296. <https://doi.org/10.1177/0016986208321808>
- [22] Kubinger K. D., Holocher-Benetka S., "Adaptives Intelligenz Diagnostikum 3 (AID 3), Version 3.2", Göttingen, Hogrefe, 2023.
- [23] Atmaca F., Baloglu M., "Using machine learning for identification: Development of the Cognitive Assessment Battery for Twice Exceptionality", *Gifted Child Quarterly*, 2025, advance online publication. <https://doi.org/10.1177/00169862251394277>
- [24] Wechsler D., "WISC-V GR: Wechsler Intelligence Scale for Children, Fifth Edition, Greek Edition" (A. Stogiannidou, Scientific Ed.), Athens, Motivo, 2017.
- [25] McGill R. J., Ward T. J., Canivez G. L., "Use of translated and adapted versions of the WISC-V: Caveat emptor", *School Psychology International*, 41(3), 2020, 276-294. <https://doi.org/10.1177/0143034320903790>



- [26] Martinussen R., Hayden J., Hogg-Johnson S., Tannock R., "A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder", *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(4), 2005, 377-384. <https://doi.org/10.1097/01.chi.0000153228.72591.73>
- [27] Almulhim N., Aljughaiman A. M., Alfaiz F. S., "Beyond the total score: Revealing cognitive heterogeneity in high ability through latent profile analysis", *Journal of Intelligence*, 14(8), 2026, 166. <https://doi.org/10.3390/jintelligence14080166>
- [28] Rodriguez A., Reise S. P., Haviland M. G., "Evaluating bifactor models: Calculating and interpreting statistical indices", *Psychological Methods*, 21(2), 2016, 137-150. <https://doi.org/10.1037/met0000045>
- [29] Tsaousis I., Sideridis G. D., AlGhamdi H. M., "Evaluating a computerized adaptive testing version of a cognitive ability test using a simulation study", *Journal of Psychoeducational Assessment*, 39(8), 2021, 954-968. <https://doi.org/10.1177/07342829211027753>
- [30] van der Linden W. J., "A hierarchical framework for modeling speed and accuracy on test items", *Psychometrika*, 72(3), 2007, 287-308. <https://doi.org/10.1007/s11336-006-1478-z>
- [31] van der Linden W. J., van Krimpen-Stoop E. M. L. A., "Using response times to detect aberrant responses in computerized adaptive testing", *Psychometrika*, 68(2), 2003, 251-265. <https://doi.org/10.1007/BF02294800>
- [32] Worrell F. C., Erwin J. O., "Best practices in identifying students for gifted and talented education programs", *Journal of Applied School Psychology*, 27(4), 2011, 319-340. <https://doi.org/10.1080/15377903.2011.615817>