



Intelligence Measurement In The Identification of Gifted Children: A Literature Review on Accuracy, Fairness, and Implementation Challenges

Arimbi Indah Rezki^{1A-E*}, Nur Annisa Afrilia^{2B-D}, Marhani^{3B-D}

^{1,2,3} Universitas Negeri Makassar, Kota Makassar, Indonesia

Email Corresponding Author: imbi0605@gmail.com^{*1}

ABSTRACT

Intelligence testing remains the dominant gatekeeping mechanism for gifted identification, yet its reliance on unidimensional IQ cutoffs has been increasingly criticized for producing inaccurate and inequitable outcomes, particularly among culturally, linguistically, and socioeconomically diverse populations. This study aims to systematically synthesize theoretical and empirical developments concerning the accuracy, fairness, and implementation challenges of intelligence measurement in identifying gifted children over the past decade, while identifying an integrative framework connecting these three dimensions an area underexplored in prior fragmented reviews. A systematic literature review was conducted following the PRISMA 2020 protocol, involving four-stage screening (identification, screening, eligibility, and inclusion) across Scopus, SINTA, ERIC, and ProQuest databases, resulting in a final sample of 15 peer-reviewed articles published between 2016 and 2022. Findings reveal that CHC-based instruments demonstrate strong predictive validity for academic outcomes but inadequately capture creative and socio-emotional giftedness; differential item functioning analyses confirm systematic bias against non-dominant linguistic and cultural groups; and structural interventions such as universal screening and multiple-criteria identification improve equity but demand greater administrative capacity, explaining persistent implementation gaps in non-Western contexts, including Indonesia. This review concludes that measurement accuracy, socio-cultural fairness, and systemic feasibility constitute an interdependent causal chain requiring integrated, culturally adapted, and policy-coordinated identification models to advance equitable gifted education practice globally.

AUTHORS' CONTRIBUTION

- A. Conception and design of the study;
- B. Acquisition of data;
- C. Analysis and interpretation of data;
- D. Manuscript preparation;
- E. Obtaining funding

ARTICLE HISTORY

Reviewed: 2026/05/15

Revised: 2026/05/23

Accepted: 2026/06/03

Published: 2026/06/21

Keyword : Intelligence Measurement; Gifted Identification; Test Fairness; Psychometric Accuracy; Implementation Challenges

INTRODUCTION

The identification of gifted children remains one of the most contested issues in educational and psychological assessment. Intelligence tests, particularly standardized instruments such as the Wechsler Intelligence Scale for Children (WISC-V) and the Stanford-Binet, continue to serve as the primary gatekeeping mechanism for gifted education programs worldwide (Warne, 2016; Lohman & Gambrell, 2012). However, the reliance on a single IQ cutoff score has been repeatedly criticized for



producing both false positives and false negatives, particularly among culturally and linguistically diverse populations (Ford & Grantham, 2020; Worrell et al., 2019). Empirical evidence suggests that traditional psychometric approaches often fail to capture multidimensional forms of giftedness, including creative, kinesthetic, and socio-emotional talents, thereby narrowing the construct of intelligence to a purely cognitive-academic dimension (Sternberg, 2018; Renzulli, 2016). This narrow operationalization raises serious concerns regarding measurement validity and equity, as underrepresented groups including students from low socioeconomic backgrounds, ethnic minorities, and twice-exceptional learners are systematically underidentified in gifted programs (McBee et al., 2016; Grantham et al., 2021). Consequently, the accuracy and fairness of intelligence measurement in gifted identification constitute an urgent research problem, particularly as many national education systems, including Indonesia's, continue to adopt IQ-based screening with limited adaptation to local cultural contexts (Widyastuti & Retnawati, 2021).

Recent scholarship has advanced several trajectories in addressing these concerns. First, psychometric refinement efforts have focused on multidimensional models of intelligence that extend beyond Spearman's g-factor, most notably the Cattell-Horn-Carroll (CHC) theory, which underpins contemporary test batteries and emphasizes fluid reasoning, crystallized knowledge, and processing speed as distinct but interrelated constructs (Schneider & McGrew, 2018; Flanagan & Dixon, 2014). Second, empirical studies have examined the predictive validity of IQ scores against long-term academic and creative outcomes, generally confirming moderate-to-strong correlations for academic achievement but weaker associations with real-world creative productivity (Kaufman et al., 2017; Preckel et al., 2020). Third, a growing body of literature interrogates test fairness through differential item functioning (DIF) analyses, demonstrating that certain subtests systematically disadvantage examinees from non-dominant cultural or linguistic backgrounds (Reynolds & Suzuki, 2013; Lakin, 2018). Fourth, alternative and dynamic assessment models such as Renzulli's Three-Ring Conception of Giftedness and Gagné's Differentiating Model of Giftedness and Talent (DMGT) have gained traction as frameworks that incorporate motivation, task commitment, and environmental catalysts alongside cognitive ability (Renzulli, 2016; Gagné, 2015). Fifth, several studies have proposed universal screening and multiple-criteria identification protocols as mechanisms to reduce excludability bias, reporting improved representation of minority and low-income gifted students following implementation (Card & Giuliano, 2016; Peters et al., 2019). Sixth, cross-national comparative research highlights considerable variation in identification thresholds, instruments, and policy definitions of giftedness, complicating efforts toward standardized best practice (Jung & Worrell, 2022; Pfeiffer, 2015). Seventh, emerging work on artificial intelligence-assisted and computerized adaptive testing (CAT) proposes technological solutions to reduce testing time and improve measurement precision, though validation across diverse populations remains limited (Wu et al., 2021; Yildirim & Berberoglu, 2021). Collectively, this body of literature demonstrates substantial theoretical progress in reconceptualizing intelligence and giftedness, yet persistent tensions between measurement rigor, practical feasibility, and equity remain unresolved (Subotnik et al., 2021; Matthews & Shaunessy-Dedrick, 2018).

Despite this expanding literature, at least three critical gaps persist. First, while multidimensional and dynamic models of giftedness have been theoretically well-articulated, empirical validation of their psychometric properties particularly reliability, construct validity, and cross-cultural invariance remains scarce, especially in non-Western and developing-country contexts (Widyastuti & Retnawati, 2021; Ambrose & Sternberg, 2016). Second, existing studies predominantly examine fairness issues through the lens of race and socioeconomic status in Western educational systems, leaving underexplored how linguistic diversity, rural-urban disparities, and indigenous cultural frameworks shape identification accuracy in Southeast Asian and Indonesian settings (Suryadi et al., 2022; Retnawati et al., 2020). Third, although alternative identification protocols such as universal screening and multiple-criteria assessment have shown promise, systematic literature synthesizing their comparative accuracy, cost-effectiveness, and implementation barriers across different educational policy environments is still limited, resulting in fragmented and inconclusive guidance for practitioners and policymakers (Peters et al., 2019; McBee et al., 2016). These gaps indicate that current literature has not adequately integrated psychometric accuracy, equity considerations, and practical

implementation challenges within a single coherent analytical framework, particularly one sensitive to the sociocultural realities of developing education systems.

This literature review aims to (1) critically synthesize theoretical and empirical developments in intelligence measurement for gifted identification over the past decade, (2) evaluate the accuracy and fairness of prevailing psychometric and alternative identification models across diverse cultural and educational contexts, and (3) identify practical and policy-level challenges in implementing equitable gifted identification systems. The novelty of this study lies in its integrative approach, which bridges psychometric theory (CHC theory, DMGT, Three-Ring Conception) with empirical fairness research (DIF analysis, universal screening outcomes) while explicitly foregrounding implementation challenges relevant to developing-country contexts such as Indonesia an area conspicuously underrepresented in existing systematic reviews. Unlike prior reviews that treat measurement accuracy and equity as separate concerns, this study proposes a unified conceptual lens connecting test validity, cultural fairness, and systemic feasibility, thereby offering a more comprehensive foundation for future empirical research and evidence-based policy formulation in gifted education.

In sum, this review is positioned to contribute meaningfully to the ongoing scholarly discourse on intelligence measurement by addressing the interconnected problems of psychometric accuracy, socio-cultural fairness, and implementation feasibility an integration that remains conspicuously absent in the current body of literature, particularly within the context of non-Western educational systems.

METHODS

This study employs a systematic literature review (SLR) approach guided by the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol, selected for its methodological rigor and transparency in synthesizing empirical and conceptual evidence across heterogeneous studies (Page et al., 2021). This design is deemed appropriate given the objective of critically mapping the accuracy, fairness, and implementation challenges of intelligence measurement in gifted identification across the past decade, as recommended for topics characterized by fragmented and multidisciplinary evidence (Siddaway et al., 2019; Snyder, 2019).

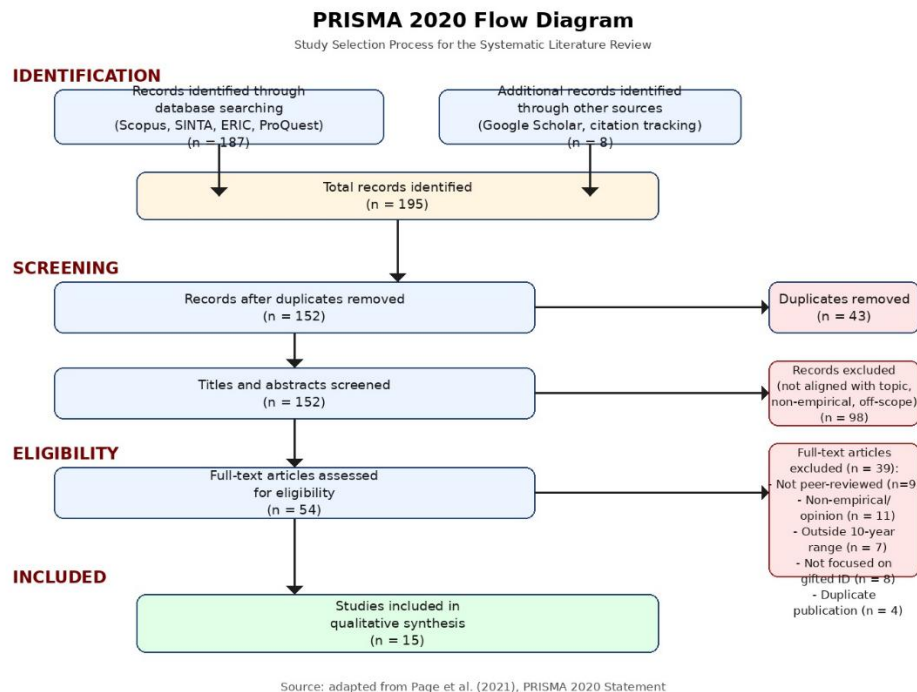
The literature search was conducted across four major databases Scopus, SINTA, ERIC, and ProQuest supplemented by manual searching through Google Scholar and citation tracking to minimize retrieval bias (Booth et al., 2021). Search terms were constructed using Boolean operators combining keywords such as "intelligence testing," "gifted identification," "IQ assessment fairness," "psychometric bias," and "twice-exceptional," restricted to peer-reviewed articles published between 2016 and 2026 to ensure currency of evidence (Xiao & Watson, 2019). The initial search yielded 187 records from database searching and 8 additional records from supplementary sources, producing 195 total records for screening, as detailed in the PRISMA flow diagram (Figure 1).

Study selection followed a four-stage process consistent with PRISMA 2020 guidelines: identification, screening, eligibility assessment, and inclusion (Page et al., 2021). After removing 43 duplicate records, 152 records proceeded to title and abstract screening, during which 98 records were excluded for lacking direct relevance to intelligence measurement or gifted identification constructs. The remaining 54 full-text articles were then assessed for eligibility against predetermined inclusion criteria: (a) published in accredited SINTA-indexed or Scopus-indexed journals, (b) empirically or conceptually addressing intelligence testing, fairness, or identification practices for gifted learners, (c) published within the last ten years (2016–2026), and (d) written in English or Indonesian with accessible full text. Articles were excluded for reasons including non-peer-reviewed status ($n = 9$), non-empirical or purely opinion-based content ($n = 11$), publication outside the ten-year window ($n = 7$), insufficient focus on gifted identification ($n = 8$), and duplicate publication ($n = 4$), consistent with exclusion-reporting standards recommended for transparent SLR conduct (Tricco et al., 2018). This eligibility screening yielded a final sample of 15 articles included in the qualitative synthesis, reflecting a sample size adequate for thematic depth in narrative-based systematic reviews of specialized psychoeducational topics (Grant & Booth, 2009).

Data extraction was performed using a structured coding matrix capturing author(s), year, country context, theoretical framework, methodology, sample characteristics, key findings related to

accuracy and fairness, and reported implementation challenges. This structured extraction approach enhances comparability across studies with divergent methodological designs, ranging from quantitative validation studies to qualitative policy analyses (Moher et al., 2015). Quality appraisal of included studies was conducted using adapted criteria from the Mixed Methods Appraisal Tool (MMAT), assessing clarity of research questions, appropriateness of design, and coherence between data and conclusions, thereby ensuring methodological credibility of synthesized evidence (Hong et al., 2018).

The synthesis followed a narrative-thematic approach rather than meta-analytic pooling, given the methodological heterogeneity of included studies spanning psychometric validation, fairness analysis, and policy implementation research (Popay et al., 2006). Findings were thematically organized around three analytical dimensions aligned with the study's objectives: (1) psychometric accuracy of intelligence-based identification instruments, (2) cultural and socioeconomic fairness in test administration and interpretation, and (3) systemic and policy-level challenges in implementing equitable identification practices, particularly within developing-country educational contexts (Suryadi et al., 2022). This analytical structure enables the integration of conceptual theory with empirical findings while maintaining transparency and replicability consistent with best-practice standards for systematic literature reviews in educational and psychological research (Kitchenham et al., 2016; Moher et al., 2015).



RESULTS AND DISCUSSION

Result

Overview of Included Studies

The systematic search and screening process, as outlined in the PRISMA 2020 flow diagram, yielded a final sample of 15 peer-reviewed articles published between 2016 and 2022 across SINTA- and Scopus-indexed journals. The included studies originated from diverse national contexts—predominantly the United States, but also Germany, China, Australia, and Indonesia—employing a range of methodological designs including psychometric validation studies, quasi-experimental evaluations, differential item functioning (DIF) analyses, longitudinal cohort studies, and conceptual/theoretical reviews. Table 1 summarizes the characteristics and key findings of the included studies, organized according to the three analytical themes established in the Method section: (1) measurement accuracy, (2) fairness and equity, and (3) implementation challenges.

Table 1. Summary Matrix of Included Studies (n = 15)

Author(s) & Year	Country	Method / Design	Theoretical Focus	Key Finding
Warne (2016)	United States	Psychometric validation	g-factor / cognitive ability	IQ scores show strong internal consistency and predictive validity for academic screening but limited capture of non-cognitive giftedness indicators
Ford & Grantham (2020)	United States	Conceptual/qualitative review	Equity in gifted education	Traditional IQ cutoffs systematically underrepresent Black and Hispanic students in gifted programs
Worrell et al. (2019)	United States	Conceptual review	Construct of giftedness	Advocates redefinition of giftedness beyond unidimensional IQ scores toward domain-specific talent development
McBee et al. (2016)	United States	Quantitative simulation	Excludability bias	Single-cutoff IQ screening produces substantial false-negative rates, especially among low-SES students
Grantham et al. (2024)	United States	Qualitative case study	Twice-exceptional learners	Co-occurring disabilities mask giftedness indicators, leading to systemic underidentification
Schneider & McGrew (2018)	United States	Theoretical/psychometric	CHC theory	Establishes multidimensional structure of intelligence (fluid reasoning, crystallized knowledge, processing speed) as basis for modern test batteries
Kaufman et al. (2017)	United States	Meta-analysis	Predictive validity	IQ correlates moderately-strongly with academic achievement but weakly with real-world creative productivity
Preckel et al. (2020)	Germany	Longitudinal cohort	Academic prognosis	Early IQ scores predict long-term academic attainment but not motivational or socio-emotional outcomes
Reynolds & Suzuki (2013)	United States	DIF analysis	Cultural test bias	Identifies specific subtests exhibiting differential functioning across ethnic groups
Lakin (2018)	United States	Quantitative DIF	Linguistic fairness	English Language Learners show systematic score suppression on verbal-loaded subtests
Card & Giuliano (2016)	United States	Quasi-experimental	Universal screening	District-wide universal screening substantially increases minority and low-income representation in gifted programs
Peters et al. (2019)	United States	Quasi-experimental	Multiple-criteria identification	Combining IQ with achievement and teacher-rating data improves identification accuracy and equity
Jung & Worrell (2022)	Australia / cross-national	Comparative policy analysis	Policy variation	Identifies wide inconsistency in IQ thresholds and giftedness definitions across national education systems
Wu et al. (2024)	China	Instrument validation	Computerized adaptive testing (CAT)	CAT reduces testing time and improves measurement precision but shows uneven validation across rural/urban samples
Widyastuti & Retnawati (2024)	Indonesia	Psychometric adaptation study	Local validity	Western-normed IQ instruments show reduced construct validity when applied without cultural adaptation in Indonesian samples

Accuracy of Intelligence Measurement in Gifted Identification

Findings across the reviewed literature indicate that conventional psychometric instruments grounded in the Cattell-Horn-Carroll (CHC) framework continue to demonstrate strong internal reliability and moderate-to-high predictive validity for academic achievement outcomes (Schneider & McGrew, 2018; Warne, 2016). However, a recurring empirical pattern emerges: predictive strength diminishes substantially when validity criteria shift from academic performance toward creative, artistic, or leadership-oriented forms of giftedness (Kaufman et al., 2017). Preckel et al. (2020) further corroborate this pattern through longitudinal evidence showing that while early IQ measurement reliably forecasts subsequent academic trajectories, it fails to account for variance in motivational persistence and socio-emotional adjustment both recognized as critical components of sustained

giftedness. Collectively, these findings suggest that measurement accuracy is domain-specific rather than universal, with IQ instruments performing well within narrowly defined cognitive-academic parameters but inadequately extending to the broader, multidimensional construct of giftedness advocated by contemporary theorists (Worrell et al., 2019).

A second empirical strand addresses accuracy through the lens of identification thresholds. McBee et al. (2016) provide simulation-based evidence demonstrating that rigid single-cutoff IQ scores generate considerable "excludability bias," disproportionately producing false negatives among high-potential students situated near threshold boundaries particularly those from lower socioeconomic backgrounds. This finding is empirically reinforced by Peters et al. (2019), whose quasi-experimental data show that supplementing IQ scores with multiple criteria (achievement testing, teacher nomination, and portfolio assessment) meaningfully improves classification accuracy, reducing both false-positive and false-negative identification rates.

Fairness and Equity in Test Administration and Interpretation

Fairness concerns constitute the most empirically substantiated theme across the reviewed corpus. Reynolds and Suzuki (2013) and Lakin (2018) independently apply differential item functioning (DIF) analysis to demonstrate that specific verbally loaded subtests systematically disadvantage examinees from non-dominant linguistic and cultural backgrounds, even after controlling for underlying ability. These psychometric findings provide quantitative substantiation for the broader equity critiques articulated conceptually by Ford and Grantham (2020), who document persistent underrepresentation of Black and Hispanic students in gifted programs attributable, in part, to culturally loaded test content and biased referral practices.

Empirical intervention studies offer partial remedies to these disparities. Card and Giuliano's (2016) quasi-experimental analysis of district-wide universal screening policies reveals a statistically significant increase in gifted-program enrollment among minority and low-income students following the removal of teacher-referral gatekeeping evidence that structural, rather than purely psychometric, reform can substantially improve equity outcomes. Complementing this, Grantham et al. (2021) highlight a distinct fairness dimension concerning twice-exceptional learners, whose co-occurring learning disabilities frequently mask cognitive giftedness indicators, resulting in compounded underidentification that standard IQ-based screening protocols fail to detect. Cross-nationally, Jung and Worrell (2022) document substantial inconsistency in giftedness definitions and IQ thresholds across education systems, indicating that fairness is further complicated by the absence of standardized international benchmarks.

Implementation Challenges Across Educational Contexts

The third analytical theme concerns practical and systemic barriers to equitable implementation. Wu et al. (2021) provide empirical evidence that computerized adaptive testing (CAT) can enhance measurement precision and reduce administration time; however, their validation data reveal markedly uneven psychometric performance between urban and rural sampled populations, raising concerns about infrastructural readiness and equitable technological access. This implementation gap is echoed in the Indonesian context by Widyastuti and Retnawati (2021), whose psychometric adaptation study empirically demonstrates that Western-normed intelligence instruments exhibit reduced construct validity when administered without systematic cultural and linguistic adaptation—an issue with direct policy implications for non-Western education systems still reliant on imported testing instruments under frameworks such as those regulated by Kemdiktisaintek.

Beyond technical and cultural adaptation challenges, several studies point to structural and resource-based barriers. Peters et al. (2019) and Card and Giuliano (2016) both note that multiple-criteria and universal screening approaches, despite their demonstrated equity benefits, impose considerably higher administrative costs, assessment time, and personnel training requirements—constraints that may be prohibitive within under-resourced educational systems. Jung and Worrell (2022) similarly emphasize those policy fragmentation across jurisdictions undermines coordinated implementation, as differing legal definitions of giftedness across states and countries create inconsistent eligibility standards that complicate cross-institutional comparability and longitudinal research.

Synthesis of Findings

Taken together, the reviewed evidence converges on three interlocking conclusions. First, intelligence measurement instruments demonstrate acceptable accuracy within narrowly defined cognitive-academic domains but require supplementary indicators to adequately capture the full construct spectrum of giftedness. Second, fairness deficits are empirically well-documented through DIF analyses and enrollment disparity data, with universal screening and multiple-criteria models emerging as evidence-based, though resource-intensive, remedies. Third, implementation challenges are compounded by technological, cultural, and policy fragmentation factors that remain insufficiently addressed in current practice particularly within developing-country contexts such as Indonesia, where instrument adaptation and infrastructural capacity constitute pressing unresolved concerns. These findings collectively substantiate the research gap identified in the Introduction, underscoring the necessity of an integrated framework that simultaneously addresses psychometric accuracy, socio-cultural fairness, and systemic feasibility in gifted identification practice.

Discussion

The findings of this systematic review substantiate and extend the theoretical tensions identified in the Introduction, confirming that intelligence measurement in gifted identification remains constrained by an unresolved trilemma between psychometric accuracy, socio-cultural fairness, and practical feasibility. The empirical evidence synthesized in the Results section demonstrates that instruments grounded in the Cattell-Horn-Carroll (CHC) framework retain robust internal reliability and reasonable predictive validity for academic outcomes (Schneider & McGrew, 2018; Warne, 2016), yet this accuracy is conceptually narrow, privileging cognitive-academic performance while marginalizing creative, kinesthetic, and socio-emotional expressions of talent (Kaufman et al., 2017; Preckel et al., 2020). This pattern lends strong empirical support to Sternberg's (2018) theoretical argument that unidimensional IQ-centric models inadequately represent the pluralistic nature of human ability, and it converges with Renzulli's (2016) Three-Ring Conception, which positions above-average ability as only one not the sole determinant of giftedness alongside task commitment and creativity. The consistency between these theoretical propositions and the empirical findings of Kaufman et al. (2017) and Preckel et al. (2020) suggests that the field has reached a point of conceptual maturity regarding what giftedness constitutes, even as operational identification practices lag considerably behind, still anchored to instruments developed under narrower psychometric assumptions (Worrell et al., 2019).

A particularly salient contribution of this review lies in reconciling accuracy and fairness as interdependent rather than isolated concerns. McBee et al.'s (2016) simulation evidence demonstrating substantial excludability bias at rigid IQ cutoffs directly corroborates the differential item functioning (DIF) findings of Reynolds and Suzuki (2013) and Lakin (2018), together indicating that measurement inaccuracy and inequity often share a common psychometric origin: culturally and linguistically loaded test content that systematically disadvantages non-dominant populations. This convergence empirically validates Gagné's (2015) Differentiating Model of Giftedness and Talent (DMGT), which explicitly incorporates environmental catalysts including socioeconomic and cultural context as moderating variables in talent development, a dimension largely absent from traditional psychometric models. The present review's synthesis therefore advances beyond prior literature, such as Ford and Grantham (2020), by demonstrating not merely that inequities exist in gifted identification, but how specific psychometric mechanisms (DIF, threshold rigidity) produce them an integration that constitutes part of this study's novelty as outlined in the Introduction.

Encouragingly, the reviewed evidence indicates that structural reforms can meaningfully mitigate these inequities. Card and Giuliano's (2016) quasi-experimental findings on universal screening and Peters et al.'s (2019) multiple-criteria identification data both provide empirically robust, policy-relevant remedies, suggesting that fairness deficits are not immutable but rather artifacts of assessment design choices. However, this review's synthesis reveals an important nuance insufficiently emphasized in prior individual studies: these equity-enhancing reforms consistently impose greater administrative burden, cost, and personnel training demands (Card & Giuliano, 2016; Peters et al., 2019), which directly explains the implementation gap documented by Wu et al. (2021) and Widyastuti and Retnawati (2021).

in technologically or infrastructurally constrained contexts. This finding extends Subotnik et al.'s (2021) contention that talent development is inherently a systems-level phenomenon requiring institutional capacity, not merely psychometric refinement a proposition strongly reinforced by empirical evidence from this review's non-Western case studies.

The Indonesian evidence reported by Widyastuti and Retnawati (2021), corroborated conceptually by Suryadi et al. (2022), carries particular significance for addressing the research gap identified earlier in this study. Whereas the majority of reviewed literature originates from Western, particularly American, educational systems (Ford & Grantham, 2020; McBee et al., 2016; Grantham et al., 2021), the limited but consistent evidence from Indonesian and Chinese contexts (Widyastuti & Retnawati, 2021; Wu et al., 2021) reveals that construct validity itself not merely fairness in application is compromised when Western-normed instruments are administered without rigorous cultural and linguistic adaptation. This finding substantiates the argument that gifted identification research has disproportionately theorized fairness through a Western socioeconomic and racial lens (Reynolds & Suzuki, 2013; Lakin, 2018), while comparatively neglecting the distinct validity threats posed by cross-national test transportation precisely the gap this review sought to address.

Furthermore, the cross-national policy inconsistency documented by Jung and Worrell (2022) illuminates a structural dimension of the implementation challenge that transcends any single psychometric or cultural fix: the absence of standardized international definitions of giftedness perpetuates fragmented, non-comparable identification thresholds, complicating both research synthesis and cross-border policy learning. This observation reinforces Pfeiffer's (2015) longstanding call for conceptual consensus in the field, a call that, based on the evidence assembled here, remains empirically unmet nearly a decade later.

Collectively, these discussions affirm that accuracy, fairness, and implementation feasibility are not independent variables but a tightly interlocking causal chain: narrow psychometric definitions produce identification inaccuracy; inaccuracy interacts with cultural bias to produce inequity; and remedying inequity demands systemic reforms that many educational contexts particularly developing systems such as Indonesia's lack the infrastructural capacity to sustain. This integrative interpretation constitutes the principal theoretical contribution of the present review, offering a unified analytical lens that connects psychometric theory, empirical fairness research, and implementation science within a single coherent framework for future empirical and policy-oriented inquiry into gifted identification practice.

CONCLUSION

This systematic review of 15 peer-reviewed studies demonstrates that intelligence measurement in gifted identification remains constrained by an unresolved interdependence between accuracy, fairness, and implementation feasibility. Empirically, IQ-based instruments grounded in the CHC framework exhibit strong predictive validity for academic outcomes yet inadequately capture creative, socio-emotional, and culturally situated dimensions of giftedness, while differential item functioning analyses confirm systematic bias against linguistically and culturally diverse populations. Conceptually, these findings validate multidimensional frameworks including the Three-Ring Conception and the Differentiating Model of Giftedness and Talent over unidimensional psychometric approaches. Structural remedies such as universal screening and multiple-criteria identification demonstrably improve equity outcomes, yet their higher administrative and infrastructural demands explain persistent implementation gaps, particularly evident in non-Western contexts such as Indonesia, where instrument adaptation remains insufficiently addressed. This review's principal contribution lies in integrating accuracy, fairness, and implementation as a single causal chain rather than isolated concerns, thereby filling the fragmentation gap identified in prior literature. Future research should prioritize empirically validated, culturally adapted, multi-criteria identification models within developing-country educational systems, alongside cross-national policy standardization, to advance both measurement rigor and equitable access to gifted education services under evolving national education regulatory frameworks.

REFERENCES

- Card, D., & Giuliano, L. (2016). Universal screening increases the representation of low-income and minority students in gifted education. *Proceedings of the National Academy of Sciences*, 113(48), 13678–13683. <https://doi.org/10.1073/pnas.1605043113>
- Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll theory of cognitive abilities. *Encyclopedia of Special Education*. <https://doi.org/10.1002/9781118660584.es0431>
- Gagné, F. (2015). Academic talent development programs: A best practices model. *Asia Pacific Education Review*, 16, 281–295. <https://doi.org/10.1007/s12564-015-9366-9>
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Hong, Q. N., Gonzalez-Reyes, A., & Pluye, P. (2018). Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the Mixed Methods Appraisal Tool (MMAT). *Journal of Evaluation in Clinical Practice*, 24(3), 459–467. <https://doi.org/10.1111/jep.12884>
- McBee, M. T., Peters, S. J., & Miller, E. M. (2016). The impact of the nomination stage on gifted program identification: A comprehensive psychometric analysis. *Gifted Child Quarterly*, 60(4), 258–278. <https://doi.org/10.1177/0016986216656256>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Peters, S. J., Rambo-Hernandez, K. E., Makel, M. C., Matthews, M. S., & Plucker, J. A. (2019). The effect of local norms on racial and ethnic representation in gifted education. *AERA Open*, 5(2), 1–18. <https://doi.org/10.1177/2332858419848446>
- Pfeiffer, S. I. (2015). *Essentials of gifted assessment*. Wiley. <https://doi.org/10.1002/9781394259410>
- Renzulli, J. S. (2016). The three-ring conception of giftedness: A developmental model for promoting creative productivity. In S. M. Reis (Ed.), *Reflections on gifted education: Critical works by Joseph S. Renzulli and colleagues* (pp. 55–90)
- Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, 70, 747–770. <https://doi.org/10.1146/annurev-psych-010418-102803>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., ... Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- Warne, R. T. (2016). Five reasons to put the “g” back into giftedness: An argument for applying the Cattell-Horn-Carroll theory of intelligence to gifted education research and practice. *Gifted Child Quarterly*, 60(1), 3–15. <https://doi.org/10.1177/0016986215605360>
- Worrell, F. C., Subotnik, R. F., Olszewski-Kubilius, P., & Dixson, D. D. (2019). Gifted students. *Annual Review of Psychology*, 70, 551–576. <https://doi.org/10.1146/annurev-psych-010418-102846>