Original Article

JHSMR Journal of Health Science and Medical Research

Validity of the Gesell Figure Drawing Test for Assessing Intelligence in Children Aged 6–10 Years

Supamas Supabanpot, M.D.[1], Tippawan Duangchoo, M.sc.[2,3], Juthamas Worachotekamjorn, M.D.[2,3]

[1]Department of Pediatrics, Faculty of Medicine, Burapha University, Mueang, Chonburi 20131, Thailand.

[2]Division of Developmental and Behavioral Pediatrics, Department of Pediatrics, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand.

[3]Child Development Center, Songklanagarind Hospital, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand.

Received 17 October 2025 ● Revised 24 November 2025 ● Accepted 5 January 2026 ● Published online 8 September 2026.

Abstract:

Objective: Standardized intelligence quotient (IQ) tests are useful measures of cognitive functioning; nonetheless, their applicability in general pediatric practice remains limited. Accordingly, the Gesell Figure Drawing Test was used for preliminary IQ assessment. This study evaluated the validity of the Gesell Figure Drawing Test in measuring IQ compared with standard tests among children aged 6–10 years.

Material and Methods: This cross-sectional study recruited children aged 6–10 years who underwent IQ testing at at the Pediatric Outpatient Department of Songklanagarind Hospital. Each child completed the Gesell Figure Drawing Test with a psychologist, followed by a standard IQ assessment (either the Wechsler Intelligence Scale for Children or Stanford-Binet LM) administered by another psychologist blinded to the Gesell Test results. The IQ scores obtained from the Gesell Figure Drawing Test and standard tests were compared to determine test validity. Inter-rater reliability was assessed by comparing the Gesell Test IQ scores obtained from three independent raters.

Results: Of the 59 participants, 43 were male (72.9%), and the mean age was 8.6 years (standard deviation, 1.1 years). Inter-rater reliability exhibited intra-class correlation values ranging from 0.70 to 0.84 across the three raters. The sensitivity, specificity, positive predictive value, and negative predictive value of the Gesell Figure Drawing Test to detect participants with IQ scores <70 were 87.5%, 80.4%, 41.2%, and 97.6%, respectively.

Conclusion: The findings of this study support the favorable psychometric properties of the Gesell Figure Drawing Test for assessing IQ in children 6–10 years of age.

Contact: Juthamas Worachotekamjorn, M.D.
Division of Developmental and Behavioral Pediatrics, Department of Pediatrics,
Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand.
E-mail: wjuthamas@yahoo.com

J Health Sci Med Res
doi: 10.31584/jhsmr.20261413
www.jhsmr.org

Keywords: gesell figure drawing test, IQ test, reliability, school-aged children, validity

Introduction

Determining intelligence quotient (IQ) scores is essential for diagnosing, assessing severity, and planning treatments or interventions in pediatric patients presenting with developmental, learning, and behavioral challenges. These conditions are particularly common among school-aged children; notably, learning disability, attention-deficit/hyperactivity disorder (ADHD), and intellectual disability (ID) are observed in approximately 15% [1], 8%[2], and 2–3% of such a population, respectively[3].

Standard intelligence tests, including the Wechsler Intelligence Scale for Children (WISC)[4] (for ages 6–16) and the Stanford Binet Intelligence Scale[5] (for ages 2–85), exhibit limited widespread application, primarily because of their administrative demands: they require clinical psychologists, involve a prolonged evaluation period (50–70 min), and incur substantial costs. Therefore, simpler alternative methods, notably drawing-based evaluations, are instrumental for pediatric IQ initial assessment. While the human figure drawing test is widely used, available data on its validity compared to standard IQ assessments are inconsistent[6,7]. Thus, the Gesell Figure Drawing Test[8] was employed as an alternative method for intellectual assessment that utilizes visual-motor perception, requiring the child to reproduce nine geometric figures (e.g., circles, squares, and triangles) with increasing age-appropriate difficulty. The resulting score is then converted into mental age, which, when compared to the chronological age of the child, allows for the calculation of the IQ score.

The Gesell Figure Drawing Test offers advantages as an intelligence-screening tool. Specifically, it is applicable to children aged 3–11 years and can be administered by a diverse range of professionals, not exclusively clinical psychologists. Importantly, the materials required are minimal and accessible (paper and pencils). Moreover, the test is time-efficient, requiring only a few minutes, and exhibits no linguistic limitations owing to its reliance on reproducing geometric shapes. Consequently, the test is well-suited for implementation in resource-limited settings.

The reliability of the Gesell Figure Drawing Test in evaluating pediatric intellectual and developmental levels has been reported[9]. A study employing both inter-rater reliability (comparing two physicians) and test-retest reliability (two assessments by the same physician at different times) yielded relatively good results for both measures. Nevertheless, to our knowledge, research concerning the validity of the Gesell Figure Drawing Test when compared to a standardized IQ assessment tool remains limited in the existing literature, with no relevant published reports available among the Thai pediatric population. Therefore, we measured the validity of the Gesell Figure Drawing Test by comparing it with the established standard IQ tests, the Wechsler Intelligence Scale for Children, third edition (WISC-III), and the Stanford-Binet Intelligence Scale, in assessing the intellectual ability of Thai children aged 6–10 years.

Material and Methods

Study design and setting

This cross-sectional study was conducted at the Pediatric Outpatient Department of Songklanagarind Hospital between March 2014 and February 2015. The study population included children aged 6–10 years referred for intelligence assessment. Children unable to hold a pencil or diagnosed with bilateral blindness or deafness were excluded. The sample size was determined using parameters derived from a preliminary survey (January–June 2013), which demonstrated a 61.9% incidence of low average IQ (score less than 80) among the tested children. This calculation assumed a probability of an expected sensitivity (P) of 0.9 and an acceptable probability of error (d) of 0.1, which were subsequently input into the sample size formula. Therefore, the minimum sample size required for this study was 57.

Informed consent was obtained from the parents or guardians of the participants after explaining the research details to them. Baseline demographic data, including age, sex, and educational level of the participants, along with the education of the caregivers and the diagnosis assigned by developmental pediatricians based on the DSM-5 criteria, were recorded. The intellectual assessment involved two independent sequential steps on the same day. First, a clinical psychologist administered the Gesell Figure Drawing Test, asking participants to reproduce three instances of each geometric shape until they could no longer draw the figures accurately. Second, another psychologist, who was blinded to the Gesell results, administered a standardized IQ assessment using either the WISC-III (for school-aged) or the Stanford-Binet Intelligence Scale (for preschoolers). Finally, the Gesell Figure Drawing Test results were independently interpreted by three raters—a psychologist, a general practitioner, and a developmental pediatrician—all blinded to the standardized test scores. The final mental age was determined by selecting the score agreed upon by at least two of the three raters; disagreements were resolved through consensus to calculate the definitive IQ score. Additionally, interrater reliability was assessed by comparing the IQ scores from the Gesell Test assigned by three independent raters.

Statistical analysis

Descriptive statistics, including percentages and means, were used for baseline demographic data. The validity of the Gesell Figure Drawing Test compared to standardized instruments (WISC-III or Stanford-Binet intelligence scale) was assessed by calculating the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) using varying IQ levels (IQ<70, <80, and <90) as the cutoff points. Additionally, receiver operating characteristic (ROC) curve analysis was conducted to estimate the area under the curve (AUC) and to identify the optimal cutoff point that maximized both sensitivity and specificity for the screening test. The clinical accuracy based on the AUC was interpreted as follows: 0.5–0.7, low accuracy; 0.7–0.9, useful accuracy; and values greater than 0.9, high accuracy.

Ethical consideration

This research was approved by the Ethics Committee for Research of the Faculty of Medicine, Prince of Songkla University (project code: REC 57-006-01-1). Informed consent was obtained from the parents or guardians of the participants.

Results

Fifty-nine children were enrolled. The mean age of the participants was 8.6 years (standard deviation, 1.1), with a notable male predominance (n=43, 72.9%). The most common clinical diagnoses among the groups were ADHD (n=24, 40.7%), learning disabilities (n =10, 16.9%), and both borderline IQ and intellectual disabilities (n=9 each, 15.3%) (Table 1). Approximately half of the participants exhibited intellectual abilities within an average range (IQ $\geq$ 90) (Table 2). The inter-rater reliability for interpreting the Gesell Figure Drawing Test was good, with ICC values ranging from 0.70 to 0.84 across the three raters.

The validity analysis (Table 3) demonstrated that the Gesell Figure Drawing Test exhibited satisfactory screening performance across multiple cutoffs compared with the standard assessments. Test validity analysis revealed that when screening for ID (IQ <70), the test exhibited good sensitivity (87.5%), specificity (80.4%), and NPV (97.6%), although the PPV was low (41.2%). Using the cutoff for borderline IQ (IQ <80), all four metrics (sensitivity, 82.4%;

specificity, 78.6%; PPV, 60.9%; and NPV, 91.7%) were classified as good. Similarly, for screening IQ less than 90, all four metrics were favorable (ranging 67.7–78.6%). The

ROC curve analysis (Figure 1.) further confirmed the utility of the test, yielding an AUC of 0.84, 0.80, and 0.73 for IQ cutoffs of <70, <80, and <90, respectively. Based on the established criteria, all three AUC values fell within the range of "useful accuracy" (0.7–0.9).

Discussion

The Gesell Figure Drawing Test is advantageous as a pediatric intelligence screening tool because of its feasibility: it is easy to administer and conserves resources, including personnel and time, given its short duration. Additionally, the test minimizes the impact of linguistic or

Table 1 Demographic data

Participant characteristics	N=59
Male (n, %)	43 (72.9)
Age (mean±S.D. (yr)	8.6±1.1
Participant education level (n, %)	
Kindergarten	2 (3.4)
Early primary school (grades 1–3)	48 (81.4)
Late primary school (grades 4–6)	9 (15.2)
Education level of the caregiver (n, %)	
Primary school	13 (22.0)
Secondary school	6 (10.2)
Vocational certificate or diploma	8 (13.6)
Bachelor's degree or higher	32 (54.2)
Principal diagnosis (n, %)	
Attention deficit hyperactivity disorder	24 (40.7)
Learning disabilities	10 (16.9)
Borderline IQ	9 (15.3)
Intellectual disabilities	9 (15.3)
Others (e.g., autism spectrum disorder, mood disorder)	8 (13.5)

IQ=intelligence quotient, S.D.=standard deviation

Table 2 IQ results from the gesell figure drawing test and standardized tests (N=59)

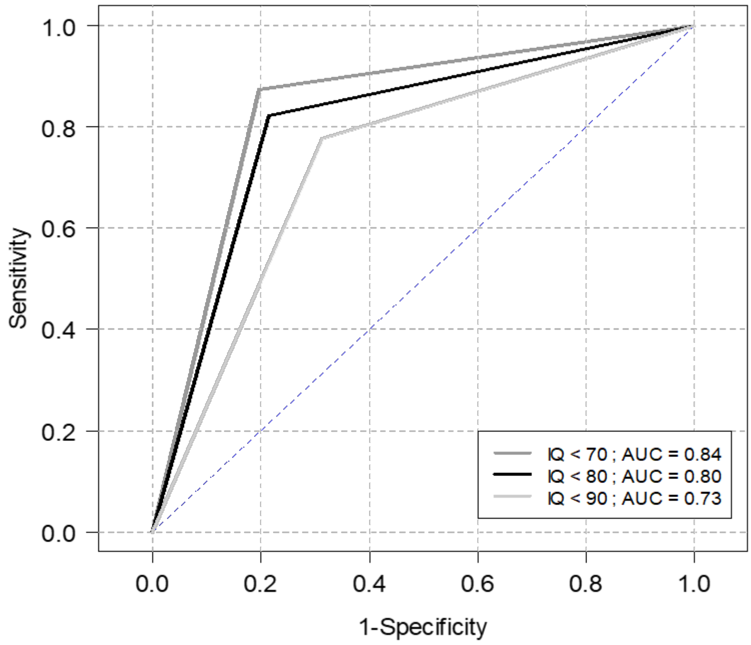
IQ results from the gesell figure drawing	IQ results from a standardized test			
	<70	70–79	80–89	≥90
<70	7	6	1	3
70–79	0	1	1	4
80–89	1	1	2	3
≥90	0	1	5	23
Total, n, (%)	8 (13.6%)	9 (15.3%)	9 (15.3%)	33 (55.9%)

IQ=intelligence quotient

Table 3 sensitivity, specificity, positive predictive value, and negative predictive value of the gesell figure drawing Test compared to those of standardized tests at various intelligence quotient (IQ) cutoff points: intellectual disability (IQ <70), borderline (IQ <80), and low average (IQ <90)

IQ cutoff point	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
IQ <70	87.5	80.4	41.2	97.6
IQ <80	82.4	78.6	60.9	91.7
IQ <90	77.8	68.8	67.7	78.6

IQ=intelligence quotient, PPV=positive predictive value, NPV=negative predictive value



AUC=area under the curve, IQ=intelligence quotient

Figure 1 receiver operating characteristic curve of the gesell figure drawing test compared to that of standardized Tests at various intelligence quotient (IQ) cutoff points: low average (IQ <90), borderline (IQ <80), and intellectual disabilities (IQ<70)

cultural differences by relying on the drawing of engaging geometric shapes, thereby helping children remain focused and cooperative throughout the assessment.

Besides its ease of administration, the Gesell Test result interpretation is reliable. Here, the results were interpreted independently by three medical professionals: a psychologist, a general practitioner, and a developmental pediatrician. A comparison of the results of the three raters using the ICC showed good agreement, corroborating the findings of Chueasuwanchai et al.[9], who reported good reliability for the interpretation of the Gesell Figure Drawing Test by two physicians (ICC=0.75; 95% confidence interval 0.61–0.84). Furthermore, both geometric figures used in the test are generally accessible. These inherent benefits make the Gesell Figure Drawing Test particularly suitable for widespread application, especially in communities or regions facing a shortage of specialists who are proficient in administering standard intelligence assessments.

In the assessment of intellectual ability, the WISC–III dictates interpretive criteria, defining an average IQ range of 90–109. An IQ score in the 80–89 range is categorized as low average, whereas scores of 70–79 are classified as borderline intellectual functioning. An IQ score of less than 70 indicates an individual in the ID group. Therefore, this study utilized these varying IQ levels (IQ<70, <80, and <90) as cutoff points for calculating validity. The resulting validity metrics differed across these thresholds, allowing the test to be used for specific screening purposes.

The analysis revealed that the IQ <70 cutoff provided the highest AUC in the ROC analysis, coupled with superior sensitivity (87.5%), specificity (80.4%), and NPV (97.6%) compared to those of other thresholds. Consequently, this cutoff is highly valuable for screening ID, as its high sensitivity and specificity are expected to remain stable even when applied to populations with a lower prevalence, such as the general population. While the IQ <80 cutoff exhibited slightly lower sensitivity and specificity, it offered a better PPV (60.9%) than did the IQ <70 threshold, making it a viable alternative for minimizing false positives. Additionally, research suggests that a Gesell score of ≥90 may be used to exclude ID. Participants scoring ≥ 90 did not score below 70 on the standard tests, with only one scoring in the borderline range (70–79). Therefore, patients scoring ≥90 on the Gesell Test may not require a standard IQ test, potentially allowing resource-constrained settings to safely reduce unnecessary referrals for full IQ assessment.

Potential limitations affecting the Gesell Figure Drawing Test and other drawing-based assessments include the requirement for adequate fine motor function and participant cooperation. Those with poor fine motor development or muscular weakness may require alternative intelligence assessments because their motor difficulties could lead to an inaccurate and falsely low intellectual ability estimation. Similarly, a lack of cooperation during test administration could yield a measured IQ score lower than the actual level. Although the Gesell Figure Drawing Test is language-independent, social and environmental factors can influence performance. Specifically, test takers from environments that strongly encourage drawing and fine-motor practice may yield different results compared to those of individuals from contrasting social contexts.

This study has limitations, as it was conducted exclusively in a pediatric clinical setting. Because more than half of the caregivers in this study held bachelor-level degrees or higher, this sample may not adequately represent populations with different sociodemographic characteristics. Moreover, study outcomes may differ when the test is applied to specific clinical groups, such as children with autism or learning disabilities. Further studies in more diverse populations are warranted.

Conclusion

Assessing the intellectual ability of children is a requisite component of clinical diagnosis and therapeutic intervention planning, in addition to the rights afforded under disability legislation. Therefore, any screening tool must possess sufficient validity to correctly identify this patient group for subsequent definitive assessments using standardized IQ tests. The Gesell Figure Drawing Test demonstrated good convenience, reliability, and validity relative to standard instruments. Specifically, IQ <70 and IQ <80 cutoffs are beneficial for screening intellectual disabilities owing to high sensitivity and specificity. Conversely, an IQ >90 cutoff is valuable for guiding clinical decisions, allowing clinicians to reduce referrals for standard IQ testing, particularly in settings with limited resources. Nonetheless, clinical application of the Gesell Test must account for other influential factors such as fine-motor function and current behavioral or emotional states in the child. Although limitations regarding generalization must be acknowledged, this research supports the conclusion that the Gesell Figure Drawing Test is a beneficial and reliable instrument for screening intellectual ability in children aged 6–10 years.

Acknowledgement

We are grateful to the Department of Pediatrics, Faculty of Medicine, Prince of Songkla University for their generous support throughout this project. Moreover, we thank all our colleagues and staff members for their administrative and technical assistance. Finally, we sincerely thank the participating children and their parents or guardians for their essential contributions to this research.

Funding sources

This study received no specific grants from any funding agency.

Conflict of interest

There are no conflicts of interest.

References

1. Mogasale VV, Patil VD, Patil NM, Mogasale V. Prevalence of specific learning disabilities among primary school children in a south Indian city. Indian J Pediatr 2012;79:342–7.
2. Ayano G, Demelash S, Gizachew Y, Tsegay L, Alati R. The global prevalence of attention deficit hyperactivity disorder in children and adolescents: an umbrella review of meta-analyses. J Affect Disord 2023;339:860–6.
3. Totsika V, Liew A, Absoud M, Adnams C, Emerson E. Mental health problems in children with intellectual disability. Lancet Child Adolesc Health 2022;6:432–44.
4. Wechsler D. Wechsler Intelligence Scale for Children: 3rd ed. manual. San Antonio, TX: The Psychological Corporation; 1991.
5. Roid GH. Stanford Binet intelligence scales: 5th ed. Itasca, IL: RiversidePublishing; 2003.
6. Aikman KG, Belter RW, Finch AJ. Human figure drawings: validity in assessing intellectual level and academic achievement. J Clin Psychol 1992;48:114–20.
7. Plubrukarn R, Theeramanoparp S. Human figure drawing test: validity in assessing intelligence in children aged 3–10 years. J Med Assoc Thai 2003;86(Suppl3):S610–7.
8. Verma SK, Pershad D, Kaushal P. Gesell drawing test as a measure of intelligence in the mentally retarded children. Indian J Ment Retard 1972;5:64–8.
9. Chueasuwanchai S, Suphakunpinyo C, Areemit RS. Reliability of Gesell figure drawing test for screening intelligence level in Thai children. Thai J Pediatr 2011;50:290–4.