

ABCD Framework-Based Stratified Analysis of ECG Interpretation Among Chinese Internal Medicine Residents: A Cross-Sectional Study

Gaigai Ma, Rui Zhang, Haiquan Li, Xinhong Wang, Dan Su, Zhenhua Han

Department of Cardiology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, 710004, People's Republic of China

Correspondence: Zhenhua Han, Department of Cardiology, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, 710004, People's Republic of China, Email 15829254977@163.com

Introduction: Deficiencies in electrocardiogram (ECG) interpretation remain a challenge in residency training. The ABCD framework enables competency-based ECG assessment by stratifying findings by clinical urgency and prevalence. However, empirical data on its application in non-Western training settings are scarce. This study applied the framework to evaluate ECG interpretation skills among Chinese internal medicine residents.

Methods: In this cross-sectional study, 158 internal medicine residents (PGY-1/2/3) at a Chinese tertiary hospital completed a 40-item ECG examination aligned with the ABCD framework in December 2025. The primary outcome was the proportion meeting a pre-specified Category A (common emergencies) benchmark ($\geq 90\%$ correct). Secondary outcomes included mean accuracy, proportion at or below a screening threshold ($\leq 70\%$ correct), and flawless performance (100% correct). Factors associated with benchmark attainment were explored using multiple logistic regression.

Results: Mean overall accuracy was 68.6% (SD 15.2%). Category A accuracy increased with training year (PGY-1: 78.1%; PGY-3: 92.1%). Only 30.2% of PGY-1 residents met the benchmark ($\geq 90\%$ correct), compared with 94.7% of PGY-3 residents. Performance at or below the screening threshold was observed in 34.0% of PGY-1 residents versus 0% of PGY-3 residents. Flawless performance was achieved by 5.7% of PGY-1 and 34.2% of PGY-3 residents. In adjusted analysis, seniority was independently associated with benchmark attainment (PGY-3 vs. PGY-1: adjusted OR=21.46, 95% CI 4.18–110.23; Firth-adjusted OR=13.07, 95% CI 3.52–79.49). The wide confidence interval reflects near-ceiling performance and warrants conservative interpretation.

Conclusion: Under timed examination conditions, over one-third of first-year residents scored below a pre-specified screening threshold for common emergency patterns, despite moderate group-mean accuracy, highlighting a gap between aggregate knowledge and consistent individual performance. These findings support competency-based ECG curricula that emphasize early mastery of emergency patterns through structured, repetitive practice.

Keywords: residency training, electrocardiogram interpretation, competency assessment, ABCD framework, graduate medical education

Introduction

ECG interpretation is a core competency for internal medicine residents,¹ yet substantial evidence documents persistently high misinterpretation rates among trainees worldwide.^{1–3} Recent studies report suboptimal performance among medical interns in Saudi Arabia,² pediatric residents in Ethiopia,³ and healthcare professionals across disciplines in Iran⁴ and North America.⁵ A recent review has revisited the persistent challenge of ECG interpretation competence across different educational contexts.⁶ Missed recognition of acute myocardial infarction or malignant arrhythmias can delay reperfusion therapy or increase sudden cardiac death risk, directly threatening patient safety.¹

China's standardized residency training system, implemented nationwide in 2014, follows a "5+3" model, five years of undergraduate education plus three years of standardized residency.^{7,8} During the undergraduate phase, ECG instruction varies considerably across institutions, often limited to didactic lectures with minimal structured interpretation practice. The subsequent

three-year residency progressively increases clinical responsibility, creating a critical transition window in which basic ECG knowledge must be converted into reliable, clinically applied skills. Many Western programs have adopted framework-driven, competency-based ECG curricula. In contrast, ECG training in Chinese residency programs is typically integrated into clinical rotations without a standardized competency-based assessment framework.⁹ This lack of systematic, framework-guided assessment makes it difficult to identify specific competency gaps or to target remediation effectively. Within this system, ECG interpretation is mandatory for internal medicine residents, yet empirical performance data using validated competency frameworks are scarce. This gap mirrors a broader challenge: an American College of Physicians systematic review concluded that no evidence-based minimum number of ECG interpretations exists for attaining or maintaining competency.¹⁰

An international task force addressed this issue by developing the ABCD framework, which stratifies ECG findings by clinical urgency and prevalence: A (common emergencies), B (common non-emergencies), C (rare emergencies), and D (rare non-emergencies).¹¹ From an educational measurement perspective, the framework supports competency-based assessment by classifying examinees according to mastery of specific attributes.¹² By identifying knowledge gaps precisely, it enables targeted instruction rather than inefficient broad remediation. Since its initial proposal by Antipervitch et al,¹¹ the ABCD framework has been endorsed as a foundation for designing competency-based ECG curricula.⁹ However, its application has largely been limited to defining learning objectives and curricular scope in Western training contexts. Although the ABCD framework was designed to link assessment with curriculum design, empirical data on its item-level application for stratified performance assessment in non-Western settings are limited.

We addressed this gap by applying the ABCD framework to perform a stratified analysis in a cohort of Chinese internal medicine residents. Specifically, this study aimed to characterize stage-specific differences in competency mastery, defined as differences in the proportion of residents achieving predefined educational benchmarks ($\geq 90\%$ correct, flawless performance, and $\leq 70\%$ screening threshold) across PGY-1, PGY-2, and PGY-3, for common ECG patterns, and to identify specific educational needs that can inform targeted curriculum development.

Materials and Methods

Study Participants and Sample Size

This cross-sectional study was conducted at the Second Affiliated Hospital of Xi'an Jiaotong University in December 2025. All internal medicine residents were invited via cluster sampling. An a priori power analysis ($\alpha=0.05$, power=0.80, anticipated effect size $f=0.25$) required 132 participants. Of 165 invitees, 158 submitted complete responses (response rate 95.8%): 53 PGY-1, 67 PGY-2, and 38 PGY-3 residents; seven were excluded due to clinical scheduling conflicts. All participants provided informed consent; the hospital's Institutional Review Board approved the study (Approval No: 2025–260).

Assessment Tool and Content

Three cardiology educators and two internal medicine program directors independently classified 60 candidate ECG findings into ABCD categories. After two rounds of moderated discussion, full consensus was reached on 40 findings, from which a 40-item multiple-choice examination was developed (10 items per category).¹³ Standardized clinical vignettes were constructed for each item, as clinical contextualization has been shown to influence ECG diagnostic accuracy,¹⁴ and two independent cardiologists verified the matching between ECG images and vignettes as well as the diagnostic accuracy.

Pilot testing with 20 residents from the same program was conducted prior to formal administration. The overall Cronbach's α was 0.81, indicating acceptable internal consistency. Category-specific α coefficients calculated from the same pilot data were as follows: Category A = 0.72, Category B = 0.68, Category C = 0.60, and Category D = 0.55. Given the small pilot sample ($n=20$), these category-level estimates should be interpreted as preliminary.

To provide a concise overview of the ABCD classification, representative diagnoses for each category are summarized in [Table 1](#). A complete example item with ECG tracing and clinical vignette is provided in [Supplementary File 1](#).

The examination included all four ABCD categories to ensure comprehensive content coverage. We adopted a two-tiered analytical approach. Categories A and B, representing common, educationally prioritized patterns, constituted the primary analysis, with educational benchmarks pre-specified based on clinical urgency. Categories C and D were included for content

Table 1 Representative ECG Diagnoses Within Each ABCD Category

Category	Clinical Classification	Representative Examples (Partial)
A	Common Emergencies	Acute ST-elevation myocardial infarction, ventricular tachycardia, high-grade atrioventricular block, ventricular fibrillation
B	Common Non-emergencies	Atrial fibrillation, left bundle branch block, right bundle branch block, first-degree atrioventricular block, premature atrial contractions
C	Rare Emergencies	Brugada syndrome, long QT syndrome, arrhythmogenic right ventricular cardiomyopathy
D	Rare Non-emergencies	Right atrial enlargement, left posterior fascicular block, dextrocardia

Notes: Only partial representative diagnoses are listed as examples. For a complete example item with ECG tracing and clinical vignette, please refer to [Supplementary File 1](#).

validity but were not psychometrically calibrated (eg, via Item Response Theory). They serve solely to document content coverage and are reported in supplementary material without performance-based inference. The ABCD classification reflects clinical judgments about urgency and prevalence, not psychometric item difficulty. Items between and within Categories A and B were not calibrated for difficulty equivalence; observed accuracy differences may therefore partly reflect intrinsic item characteristics (eg, subtlety of ECG changes, vignette complexity) rather than true competency differences. All between- and within-category comparisons should be interpreted with this caveat.

Assessment Implementation and Bias Control

The examination was administered in a computer laboratory via an institutional online platform, supervised by a research assistant, with a 90-second-per-question time limit to simulate clinical pressure (set by faculty consensus). The primary outcome was the proportion of residents meeting a pre-specified Category A educational benchmark ($\geq 90\%$ correct). This aspirational target aligns with standards in advanced ECG education.⁹ Secondary outcomes included mean accuracy, flawless performance (100% correct), and performance at or below an educational screening threshold ($\leq 70\%$ correct), set to identify trainees who may particularly benefit from additional practice. This threshold has not been validated against clinical competence measures or patient outcomes.

Statistical Analysis

Analyses used SPSS v26.0 (IBM Corp). Continuous variables are presented as mean $\pm$ SD; group comparisons used one-way ANOVA with eta-squared (η^2) effect sizes and Bonferroni-corrected post-hoc tests. Categorical variables are presented as frequencies with 95% CIs; between-group comparisons used χ^2 -tests. As an exploratory analysis, multiple logistic regression examined factors independently associated with meeting the Category A educational benchmark ($\geq 9/10$ correct), with training year, gender, entrance exam score, and pre-training ECG hours as independent variables. Model fit was assessed via the Hosmer–Lemeshow test. Since the PGY-3 subgroup exhibited near-ceiling performance (94.7% benchmark achievement), raising concern for quasi-complete separation, Firth’s penalized maximum likelihood logistic regression (R package logistf, version 1.26) was performed as a sensitivity analysis to provide more conservative coefficient estimates and better-calibrated confidence intervals. A two-tailed $P < 0.05$ defined statistical significance.

Results

Participant Characteristics

A total of 158 residents completed the examination (response rate 95.8%): 53 PGY-1, 67 PGY-2, and 38 PGY-3. [Table 2](#) presents their baseline characteristics. No significant differences were observed across groups in age, gender, or entrance examination scores (scored on a 0–100 scale; all $P > 0.05$). Pre-training ECG hours differed significantly across training years ($P < 0.001$), with PGY-3 residents reporting substantially more prior ECG learning than PGY-2 and PGY-1 residents (post-hoc Bonferroni correction: both $P < 0.001$).

Overall and Stratified ECG Interpretation Accuracy

Mean overall accuracy was 68.6% (SD 15.2%; [Table 3](#) and [Figure 1](#)). Category A mean accuracy was 85.6% (SD 16.0%) and Category B 72.5% (SD 15.8%; [Table 3](#)); the 13-percentage-point difference (95% CI 8.5–17.7) warrants cautious

Table 2 Baseline Characteristics of 158 Internal Medicine Residents, Overall and Stratified by Post-Graduate Year (PGY)

Characteristic	All Residents (n=158)	PGY-1 (n=53)	PGY-2 (n=67)	PGY-3 (n=38)	P-value
Age, years	25.1 ± 1.8	24.6 ± 1.6	25.2 ± 1.7	25.8 ± 1.9	0.052
Female, n (%)	95 (60.1%)	34 (64.2%)	38 (56.7%)	23 (60.5%)	0.724
Entrance Exam Score	80.2 ± 7.8	78.5 ± 7.9	80.8 ± 7.6	81.8 ± 7.7	0.102
Pre-training ECG Hours	11.2 ± 6.8	8.2 ± 5.9	11.9 ± 6.5	15.1 ± 6.3	<0.001

Notes: Post-hoc analysis with Bonferroni correction showed significantly more pre-training ECG hours in PGY-3 compared to PGY-2 and PGY-1 (both $P < 0.001$).

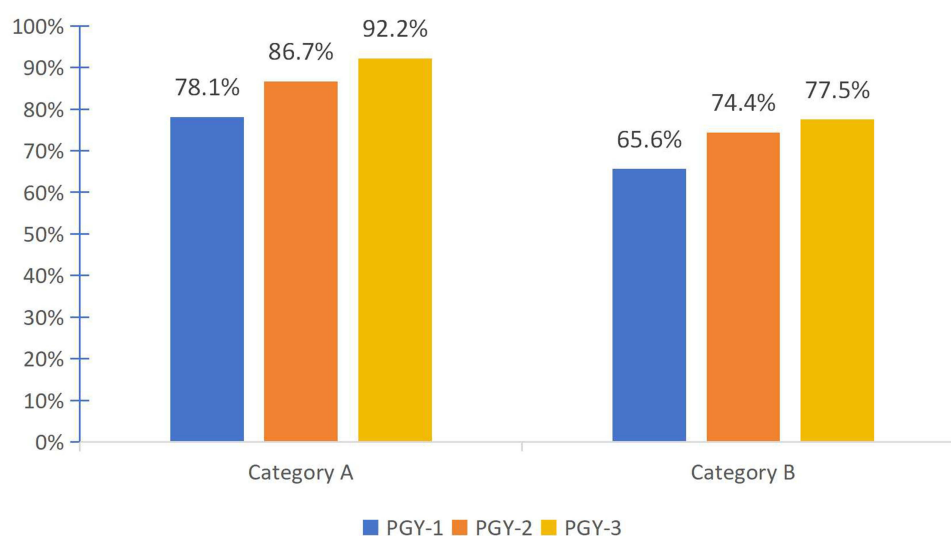
Table 3 Comparison of ECG Interpretation Accuracy (%) for Categories A and B by Training Year

Group	n	Overall Accuracy, % (SD)	Category A, % (SD)	Category B, % (SD)
PGY-1	53	59.7 (13.2)	78.1 (18.2)	65.5 (16.8)
PGY-2	67	69.4 (11.9)	86.7 (14.5)	74.3 (15.3)
PGY-3	38	76.9 (10.1)	92.1 (11.3)	77.6 (13.4)
Total	158	68.6 (15.2)	85.6 (16.0)	72.5 (15.8)
P-value		<0.001	<0.001	0.003

Notes: Post-hoc analysis with Bonferroni correction for overall accuracy showed that PGY-3 outperformed PGY-2, and PGY-2 outperformed PGY-1, with all P values below 0.05. Comparable improvements were observed across training years for Categories A and B individually.

interpretation, as item difficulty was not equated across categories, and differences may partly reflect intrinsic item characteristics rather than true competency gaps. Accuracy improved with training year across categories (both $P < 0.05$; $\eta^2=0.216$). PGY-3 residents outperformed PGY-1 residents (Cohen's $d = 1.45$).

Categories C (rare emergencies) and D (rare non-emergencies) were included for content validity but not psychometrically calibrated; their raw accuracy values confound trainee ability with item difficulty and cannot be interpreted as competency indicators. Detailed results and limitations appear in [Supplementary File 2](#).

**Figure 1** Comparison of ECG Interpretation Accuracy Across Competency Categories by Training Year. Grouped bar chart showing mean accuracy (%; Y-axis) for Category A (common emergencies) and Category B (common non-emergencies) among PGY-1, PGY-2, and PGY-3 resident groups (X-axis).

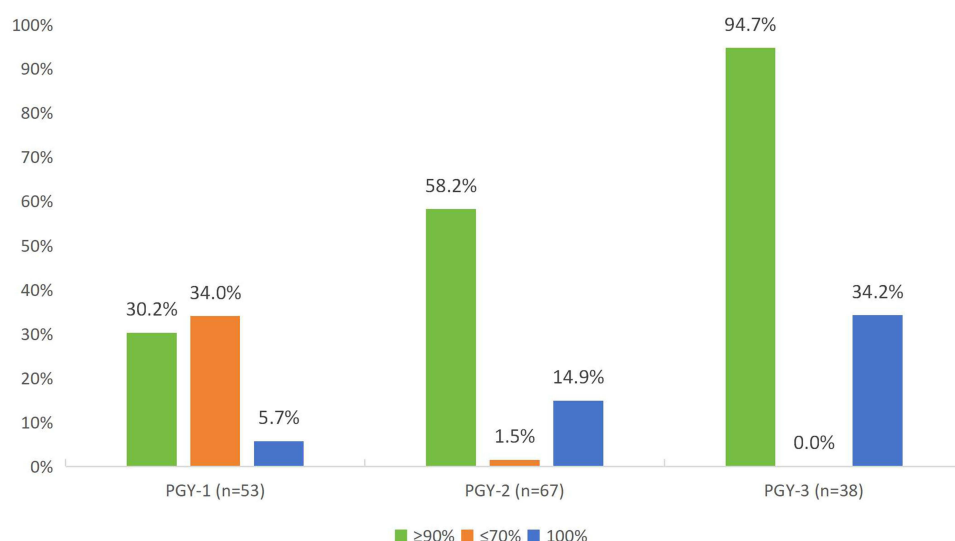


Figure 2 Category A Proficiency Outcomes by Training Year. Grouped bar chart illustrating the proportion of PGY-1, PGY-2, and PGY-3 residents achieving high proficiency ($\geq 9/10$ correct, green), flawless performance ($10/10$ correct, blue), and performance below remediation threshold ($\leq 7/10$ correct, red) on the 10-item Category A ECG examination.

Category A Proficiency Analysis

Category A performance was further examined using pre-specified educational thresholds (Figure 2 and [Supplementary Figure 1](#)). These thresholds serve as educational reference points for curriculum planning and resource allocation, not validated clinical competency standards. The proportion meeting the educational benchmark ($\geq 90\%$ correct) was 30.2% (95% CI 18.3–44.8%) in PGY-1, 58.2% (95% CI 45.5–70.1%) in PGY-2, and 94.7% (95% CI 82.3–99.4%) in PGY-3 ($\chi^2 = 38.52$, $P < 0.001$). Proportions at or below the screening threshold ($\leq 70\%$ correct) were 34.0% (95% CI 21.7–48.3%), 1.5% (95% CI 0.04–8.0%), and 0% (95% CI 0–9.3%) for PGY-1, PGY-2, and PGY-3, respectively ($\chi^2 = 45.21$, $P < 0.001$). Flawless performance (100% correct) was achieved by 5.7%, 14.9%, and 34.2% of PGY-1, PGY-2, and PGY-3 residents, respectively ($\chi^2 = 13.78$, $P = 0.001$); 95% CIs for each are 1.2%–15.7%, 7.4%–25.7%, and 19.6%–51.4%. [Supplementary Figure 2](#) shows score distributions by training year.

Table 4 presents the adjusted logistic regression model. Seniority was independently associated with meeting the benchmark (PGY-2 vs. PGY-1: adjusted OR = 2.54, 95% CI 1.16–5.57, $P = 0.020$; PGY-3 vs. PGY-1: adjusted OR = 21.46, 95% CI 4.18–110.23, $P < 0.001$). Pre-training ECG hours (adjusted OR = 1.11 per hour, 95% CI 1.04–1.18, $P = 0.002$), and entrance exam scores (adjusted OR = 1.06 per point, 95% CI 1.01–1.11, $P = 0.024$) were also independently associated. Gender was not

Table 4 Multiple Logistic Regression Analysis of Factors Associated with Meeting the Category A Educational Benchmark ($\geq 90\%$ Correct)

Variable	Standard Logistic Regression			Firth Penalized Likelihood		
	Adjusted OR	95% CI	P-value	Adjusted OR	95% CI	P-value
PGY-1 (reference)	1.00	—	—	1.00	—	—
PGY-2	2.54	1.16–5.57	0.020	2.47	1.14–5.43	0.021
PGY-3	21.46	4.18–110.23	<0.001	13.07	3.52–79.49	<0.001
Pre-training ECG hours (per hour)	1.11	1.04–1.18	0.002	1.10	1.03–1.17	0.003
Entrance exam score (per point)	1.06	1.01–1.11	0.024	1.06	1.01–1.11	0.018
Female gender	0.92	0.41–2.03	0.83	0.93	0.42–2.02	0.84

Notes: Standard logistic regression model: Hosmer-Lemeshow goodness-of-fit $\chi^2 = 7.89$, $df = 8$, $P = 0.44$. Firth penalized likelihood model: likelihood ratio $\chi^2 = 53.25$, $df = 5$, $P < 0.001$. Firth regression was performed as a sensitivity analysis to address quasi-complete separation due to the near-perfect benchmark achievement rate (94.7%) in the PGY-3 subgroup. The Firth-adjusted estimates are more conservative and provide better-calibrated confidence intervals under sparse event conditions.

significant (adjusted OR = 0.92, 95% CI 0.41–2.03, $P = 0.83$). Model fit was acceptable (Hosmer–Lemeshow $P = 0.44$). Inclusion of pre-training ECG hours attenuated the PGY-3 odds ratio from 33.75 (95% CI 7.49–152.14, unadjusted) to 21.46 (95% CI 4.18–110.23, adjusted), indicating that some of the seniority effects were attributable to self-directed learning.

The wide confidence interval for the PGY-3 comparison in the standard model (95% CI 4.18–110.23) reflects the near-perfect benchmark achievement rate (94.7%; 36/38) in this subgroup, which results in quasi-complete separation and limited information for precise odds ratio estimation. The Firth-adjusted analysis yielded a more conservative point estimate for PGY-3 (OR = 13.07, 95% CI 3.52–79.49, $P < 0.001$), confirming the robustness of the association's direction and statistical significance. These findings should be interpreted as evidence of a strong positive association rather than a precisely estimated effect size.

Discussion

Main Findings

In this cohort of 158 residents, overall mean ECG interpretation accuracy (68.6%) aligns with prior studies,^{5,15,16} including a recent multi-center Chinese report by Liang et al (65.2%).¹⁶ Applying the ABCD framework revealed a pattern where moderate mean performance coexists with wide individual variability. PGY-1 residents achieved a Category A mean accuracy of 78.1%, suggesting foundational knowledge, yet only 30.2% (95% CI 18.3–44.8%) met the educational benchmark ($\geq 90\%$ correct), and 34.0% (95% CI 21.7–48.3%) scored at or below the screening threshold. This pattern illustrates that aggregate scores may obscure meaningful variation among early-career trainees. These thresholds are educational reference points, not validated clinical competency standards; their association with clinical decision-making requires further study before any patient safety inferences can be drawn. It should be noted that the wide confidence interval for PGY-3 benchmark attainment (standard model 95% CI 4.18–110.23; Firth-adjusted 95% CI 3.52–79.49) warrants conservative interpretation of the effect magnitude for this subgroup.

Incremental Value of the ABCD Framework

A question is whether similar findings could emerge from analyzing individual high-risk items without the ABCD framework. Recognizing that observed accuracy differences may partly reflect uncalibrated item difficulty, we base the framework's value on structural contributions, not unique data patterns. First, the ABCD framework provides a transparent, reproducible design: expert-consensus categorization ensures systematic representation of all clinically relevant domains, enables cross-institutional comparison, and eliminates post-hoc categorization bias. Second, it maps assessment results directly onto differentiated educational actions. Category A defines patterns for which near-perfect accuracy is the aspirational goal; lower performance warrants attention. Category B informs knowledge breadth in non-urgent contexts. Categories C and D define the boundaries of exposure-based learning. This mapping, from category to educational response, is the framework's principal pragmatic advantage: it directs educators to where instructional effort is needed. In our study, this structure enabled systematic examination of the spread between mean accuracy and individual-level consistency.

Interpretation of the Reliability Gap

One possible interpretation of this dissociation between mean accuracy and consistent mastery is that it is consistent with cognitive load theory. ECG interpretation under time constraints imposes high intrinsic cognitive load, which may overwhelm novice working memory and lead to errors even when relevant knowledge is present.^{17,18} In early training, emergency pattern recognition may not yet be sufficiently automated to support consistent performance under conditions that simulate clinical pressure. This cognitive framework provides a possible explanation for why many PGY-1 residents in our study could recognize emergency patterns on average, yet only a minority demonstrated the consistent mastery captured by the $\geq 90\%$ threshold. However, we emphasize that this interpretation is based on the congruence between observed performance patterns and the cognitive load theory framework, not on direct measurement of individual response times or cognitive processes. Future studies incorporating response time data, eye-tracking, or think-aloud protocols would be needed to directly test this mechanistic hypothesis.

If this interpretation is correct, the educational implication is not that residents below the threshold lack knowledge, but that their recognition skills may not yet be sufficiently practiced for consistent performance under time pressure. This frames the gap as a remediable learning need, transitioning from effortful recall to fluent pattern recognition, rather than an inability to recognize emergencies. Repetitive retrieval practice, spaced exposure, and structured algorithms are proven strategies for building such fluency, and may be particularly relevant for PGY-1 residents before they assume greater clinical autonomy.^{19,20}

Comparison with Existing Literature

The stepwise accuracy improvement with advancing training year mirrors prior reports.^{5,16} While overall accuracy rates are similar across studies, our use of pre-specified educational benchmarks enables more detailed examination of variability in early training. Over one-third (34.0%) of PGY-1 residents scored at or below the screening threshold, set as an educational reference point, not a definition of clinical incompetence. While direct comparison is limited by methodological differences, this proportion is broadly comparable to the 36% overall miss rate for critical ECG findings reported by Sibbald et al in a cardiology resident cohort.²¹ Consistent with prior literature, baseline knowledge (entrance exam scores) correlated with ECG performance, supporting readiness-based training approaches.⁵ A recent systematic review by Oh et al confirmed that while instructional interventions can improve ECG learning, many evidence-based strategies remain insufficiently investigated.²²

Explanation of Pre-Training ECG Hours Differences

The significant difference in self-reported pre-training ECG hours across PGY groups (Table 2) was unexpected and warrants discussion. Three complementary explanations may account for this finding. First, recall bias may have played a role: senior residents, having accumulated richer clinical ECG experience, may have adopted a broader definition of what constitutes formal “ECG training,” leading to higher retrospective estimates, whereas PGY-1 residents may have provided more conservative estimates limited to clearly recalled classroom hours. Second, a cohort effect is plausible: Chinese undergraduate medical curricula have undergone gradual reforms over the past decade, with more recent cohorts potentially receiving fewer standalone ECG teaching hours as curricula shifted from discipline-based to integrated system-based approaches. Third, a selection effect cannot be excluded in this single-center design: residents with higher intrinsic motivation and more pre-training self-study may have been more likely to enter and remain in a university-affiliated residency program, thereby inflating the mean in senior groups. Regardless of the underlying mechanism, the independent association of pre-training hours with benchmark attainment (adjusted OR = 1.11 in the standard model and 1.10 in the Firth model, per hour) indicates that part of the observed seniority effect may stem from baseline differences. Nevertheless, the fact that training year remained strongly significant after adjustment suggests that both baseline preparation and residency training contribute independently to competency development.

Implications for Training and Future Research

The patterns observed in our data, moderate group-mean accuracy coexisting with wide individual variability, particularly among PGY-1 residents, directly inform several educational priorities. First, the 34.0% of PGY-1 residents scoring below the screening threshold for Category A emergencies suggests a need for structured, repetitive pattern-recognition drills early in residency, such as daily microlearning modules (5–8 ECG items per session with immediate feedback).²⁰ Second, the progressive but incomplete Category B mastery supports a staged curriculum in which Category A competency is verified before introducing Category B content. Third, the high benchmark achievement among PGY-3 residents (94.7%) suggests that senior residents could serve as peer-teaching resources while their own learning shifts toward advanced interpretation and rare emergency patterns. Beyond these specific recommendations, the ABCD framework serves as a formative assessment tool throughout residency, enabling progress tracking, deficiency identification, and tailored remediation, an advantage over traditional aggregate scoring approaches.

Future research should employ formal psychometric models (eg, Rasch analysis, cognitive diagnostic models) to calibrate items across all ABCD categories, enabling meaningful interpretation of C and D performance. Standard-setting exercises linking benchmark performance to clinical outcomes would validate the educational thresholds used here.

Prospective intervention studies comparing framework-based curricula with conventional teaching are warranted to determine whether this diagnostic precision translates into improved educational outcomes.

Limitations

Several limitations warrant consideration. First, the single-center, cross-sectional design restricts generalizability, although our program mirrors China's national "5+3" model. The cross-sectional design precludes causal inference; associations with training year may partly reflect between cohort differences in undergraduate preparation, selection, or clinical exposure rather than residency training itself. Unmeasured confounders, including undergraduate ECG instruction quality and clinical exposure volume, could not be accounted for, and residual confounding may persist despite adjustment.

Second, several psychometric limitations should be noted. Items across all ABCD categories were not calibrated for difficulty equivalence. The ABCD classification reflects clinical urgency and prevalence, not psychometric item difficulty. Observed accuracy differences between and within categories may therefore partly reflect intrinsic item characteristics (eg, ECG subtlety, distractor similarity, vignette complexity) rather than true competency differences. We addressed this for Categories C and D by restricting their results to supplementary material; for Categories A and B, we reported 95% CIs and avoided strong between-category claims. Additionally, the category-specific Cronbach's α coefficients (A = 0.72, B = 0.68, C = 0.60, D = 0.55) were derived from a small pilot sample (n = 20) and are subject to considerable sampling error; confirmatory psychometric evaluation in a larger sample is warranted. Furthermore, the pre-training ECG hours variable relied on self-report and may be affected by recall bias.

Third, the educational thresholds ($\geq 90\%$ benchmark; $\leq 70\%$ screening threshold) were set based on face validity and educational discrimination needs. They lack validation against clinical decision-making, patient outcomes, or consensus-based standard-setting. The $\leq 70\%$ threshold should be regarded solely as a criterion for identifying residents who may benefit from intensified educational support; it does not define clinical incompetence.

Fourth, the logistic regression estimates for the PGY-3 subgroup should be interpreted with caution, as near-ceiling benchmark achievement (94.7%; 36/38) resulted in quasi-complete separation. The Firth-adjusted sensitivity analysis confirmed the robustness of the association's direction (OR = 13.07, 95% CI 3.52–79.49), but the persistently wide confidence interval reflects the inherent limitation of estimating effect sizes under these conditions.

Conclusion

Applying the ABCD framework for competency-based stratified assessment identified stage-specific patterns in ECG interpretation among Chinese internal medicine residents. Under timed examination conditions, over one-third of PGY-1 residents did not achieve the pre-specified educational benchmark for common emergency patterns, despite moderate group-mean accuracy. This gap between aggregate knowledge and consistent individual performance suggests that early residency training should prioritize the transition from foundational knowledge to fluent, reliable pattern recognition, a hypothesis that requires prospective validation. Although the single-center, cross-sectional design and the variable internal consistency across ABCD categories limit generalizability, this study provides a structured assessment model and evidence base for designing competency-based ECG curricula that emphasize early mastery of emergency patterns through systematic, framework-guided training.

Abbreviations

ECG, electrocardiogram; PGY, Post-Graduate Year; SD, standard deviation; OR, odds ratio; CI, confidence interval; ANOVA, analysis of variance; IRT, item response theory; CDM, Cognitive Diagnostic Model.

Data Sharing Statement

The datasets used and/or analyzed during the current study are available from the corresponding author (Zhenhua Han, Email: 15829254977@163.com) upon reasonable request.

Ethics Approval and Informed Consent

This study was approved by the Institutional Review Board of the Second Affiliated Hospital of Xi'an Jiaotong University (Approval No: 2025-260). All participants provided written informed consent prior to participation. The study was conducted in accordance with the Declaration of Helsinki.

Acknowledgments

The authors thank the residents who participated in this study and the medical education department for their administrative support.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agreed to be accountable for all aspects of the work.

Funding

There is no funding to report.

Disclosure

The authors report no conflicts of interest in this work.

References

- Breen CJ, Kelly GP, Kernohan WG. ECG interpretation skill acquisition: a review of learning, teaching and assessment. *J Electrocardiol.* 2022;73:125–128. doi:10.1016/j.jelectrocard.2019.03.010
- Al Mousa AM, Alhubail FM, Almulhim M, et al. Electrocardiogram interpretation competency of medical interns in Saudi Arabia: a cross-sectional study. *Cureus.* 2023;15(4):e37557. doi:10.7759/cureus.37557
- Mintesnot H, Tadele H, Ahmed H, et al. Electrocardiography interpretation competency among pediatric and child health residents at Addis Ababa University, Ethiopia. *BMC Med Educ.* 2024;24(1):1548. doi:10.1186/s12909-024-06614-5
- Amini K, Mirzaei A, Hosseini M, et al. Assessment of electrocardiogram interpretation competency among healthcare professionals and students of Ardabil University of Medical Sciences: a multidisciplinary study. *BMC Med Educ.* 2022;22(1):448. doi:10.1186/s12909-022-03518-0
- Kashou AH, Noseworthy PA, Beckman TJ, et al. Predictors of ECG Interpretation Proficiency in Healthcare Professionals. *Curr Prob Cardiol.* 2023;48(12):102011. doi:10.1016/j.cpcardiol.2023.102011
- Carmona-Puerta R, Lorenzo-Martínez E. The issue of electrocardiography interpretation competence revisited. *Educación Médica Superior.* 2025;39:e4563.
- Zhu J, Li W, Chen L. Doctors in China: improving quality through modernisation of residency education. *Lancet.* 2016;388(10054):1922–1929. doi:10.1016/S0140-6736(16)00582-1
- Xia X, Qin C, Li M, et al. Standardized nuclear medicine residency training program in China. *Eur J Nucl Med Mol Imaging.* 2024;51(11):3153–3156. doi:10.1007/s00259-024-06798-4
- Kaye MG, Kwiatkowski AV, Khan HA, et al. Designing an ECG curriculum for residents: evidence-based approaches to improving resident ECG interpretation skills. *J Electrocardiol.* 2024;82:64–68. doi:10.1016/j.jelectrocard.2023.10.012
- Kadish AH, Buxton AE, Kennedy HL, et al. ACC/AHA clinical competence statement on electrocardiography and ambulatory electrocardiography. A report of the ACC/AHA/ACP-ASIM Task Force on Clinical Competence (ACC/AHA Committee to Develop a Clinical Competence Statement on Electrocardiography and Ambulatory Electrocardiography). *J Am Coll Cardiol.* 2001;38(7):2091–2100. doi:10.1016/s0735-1097(01)01680-1
- Antiperovitch P, Zareba W, Steinberg JS, et al. Proposed in-training electrocardiogram interpretation competencies for undergraduate and postgraduate trainees. *J Hosp Med.* 2018;13(3):185–193. doi:10.12788/jhm.2876
- Holmboe ES, Sherbino J, Englander R, et al. A call to action: the controversy of and rationale for competency-based medical education. *Med Teach.* 2017;39(6):574–581. doi:10.1080/0142159X.2017.1315067
- Parekh P, Bahadoor V. The utility of multiple-choice assessment in current medical education: a critical review. *Cureus.* 2024;16(5):e59778. doi:10.7759/cureus.59778
- Viljoen CA, Millar RS, Manning K, et al. Clinically contextualised ECG interpretation: the impact of prior clinical exposure and case vignettes on ECG diagnostic accuracy. *BMC Med Educ.* 2021;21(1):417. doi:10.1186/s12909-021-02854-x
- Cook DA, Oh S, Pusic MV. Assessments of physicians' electrocardiogram interpretation skill: a systematic review. *Acad Med.* 2022;97(4):603–615. doi:10.1097/ACM.0000000000004140
- Liang S, Lai R, Xu T. Evaluating the competency and confidence in ECG interpretation among residents: a multi-center pilot study. *Adv Med Educ Pract.* 2025;16:2481–2489. doi:10.2147/AMEP.S550536
- Venkat MV, O'Sullivan PS, Young JQ, et al. Using cognitive load theory to improve teaching in the clinical workplace. *MedEdPORTAL.* 2020;16:10983. doi:10.15766/mep_2374-8265.10983

18. Thomas MG. The use of cognitive load theory to assist in the teaching of electrocardiogram interpretation within paramedical science education. *Clin Teach*. 2024;21(4):e13759. doi:10.1111/tct.13759
19. Cunningham JM, Johnson M, Kincaid T, et al. The spacing effect: improving electrocardiogram interpretation. *Clin Teach*. 2024;21(1):e13626. doi:10.1111/tct.13626
20. Moulin T, Lellouche N, Gandjbakhch E, et al. Daily Quiz-based microlearning program to support electrocardiogram interpretation training for medical students: a feasibility study. *CJC Open*. 2025;8(1):43–50. doi:10.1016/j.cjco.2025.09.009
21. Sibbald M, Davies EG, Dorian P, et al. Electrocardiographic interpretation skills of cardiology residents: are they competent? *Can J Cardiol*. 2014;30(12):1721–1724. doi:10.1016/j.cjca.2014.08.026
22. Oh S, Cook DA, Van Gerven PWM, et al. Physician training for electrocardiogram interpretation: a systematic review and meta-analysis. *Acad Med*. 2022;97(4):593–602. doi:10.1097/ACM.0000000000004607

Advances in Medical Education and Practice

Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>

Dovepress
Taylor & Francis Group