Validation of a Novel Digital Tool in Automatic Scoring of an Online ECG Examination at an International Cardiology Meeting

Kieran L. Quinn, M.D., M.Sc.,^{*,†} Eugene Crystal, M.D., F.R.C.P.C.,[‡]
Ilan Lashevsky, M.D., F.R.C.P.C.,[‡] Banafsheh Arouny, M.D.,[‡]
and Adrian Baranchuk, M.D., F.A.C.C., F.R.C.P.C.[§]

From the ^{*}Keenan Research Centre for Biomedical Science and Department of Medicine, University of Toronto, Toronto, Canada; [†]St. Michael's Hospital, University of Toronto, Toronto, Canada; [‡]Schulich Heart Centre, Sunnybrook Research Institute, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Canada; and [§]Arrhythmia Service, Kingston General Hospital, Queen's University, Kingston, Canada

Background: We have previously developed a novel digital tool capable of automatically recognizing correct electrocardiography (ECG) diagnoses in an online exam and demonstrated a significant improvement in diagnostic accuracy when utilizing an inductive-deductive reasoning strategy over a pattern recognition strategy. In this study, we sought to validate these findings from participants at the International Winter Arrhythmia School meeting, one of the foremost electrophysiology events in Canada.

Methods: Preregistration to the event was sent by e-mail. The exam was administered on day 1 of the conference. Results and analysis were presented the following morning to participants.

Results: Twenty-five attendees completed the exam, providing a total of 500 responses to be marked. The online tool accurately identified 195 of a total of 395 (49%) correct responses (49%). In total, 305 responses required secondary manual review, of which 200 were added to the correct responses pool. The overall accuracy of correct ECG diagnosis for all participants was 69% and 84% when using pattern recognition or inductive-deductive strategies, respectively.

Conclusion: Utilization of a novel digital tool to evaluate ECG competency can be set up as a workshop at international meetings or educational events. Results can be presented during the sessions to ensure immediate feedback.

Ann Noninvasive Electrocardiol 2016;21(4):376–381

ECG; electrocardiogram; evaluation tools

Electrocardiography (ECG) is a critically important clinical tool used to diagnose heart disease, but recent evidence has demonstrated a concerning lack of competency among medical trainees.^{1–6} One possible explanation is that there is considerable controversy surrounding the most effective approach to ECG interpretation. Furthermore, academic institutions and organizers of individual ECG workshops at society meetings have been limited in their ability to conduct a large-scale paper-based ECG exam assessing competency

because of the amount of time needed to manually mark them.

We recently published the results of our study that characterized the performance of a novel digital tool in assessing a large group of medical trainees' competency in ECG interpretation.⁷ In this study, we sought to validate these findings from participants at the International Winter Arrhythmia School (IWAS) meeting, one of the foremost electrophysiology events in Canada.

Address for Correspondence: Adrian Baranchuk, M.D., F.A.C.C., F.R.C.P.C., Arrhythmia Service, Kingston General Hospital, Queen's University, Canada. Fax: +1-613-548-1387; E-mail: barancha@kgh.kari.net

© 2015 Wiley Periodicals, Inc.
DOI: 10.1111/anec.12311

METHODS

Study Participants

The IWAS is a prestigious electrophysiology conference held annually that has steadily grown in participation since its inception with over 175 participants in 2015. Attendees at the 12th annual IWAS held in Collingwood, Ontario, Canada, were encouraged to participate in the ECG workshop held on February 6 and 7, 2015, through invitation via e-mail, advertising on the IWAS website, and in person at the conference. IWAS is designed for professionals within the electrophysiology community, including physicians (cardiologists and electrophysiologists), trainees/fellows, and EP technicians/nurses.

Workshop Design

The first component of the IWAS study consisted of a 1-hour ECG workshop held on the morning of first day of the meeting. A 20-minute didactic lecture was delivered to review the concepts relating to the cognitive strategies of ECG interpretation, including pattern recognition and inductive-deductive reasoning. A 10-minute session followed to provide general instructions for participants and an overview of the components of the examination. Participants who had not preregistered for the session provided their e-mail addresses, and an invitation containing a unique and secure link was sent to all participants via e-mail. Participants then completed the 30-minute exam in the remaining allotted time for the workshop.

After completion of the exam, the data were analyzed onsite within 2 hours. On the following morning, a second 30-minute session was organized for participants to review the correct responses to each exam question and to review the pertinent pathophysiologic descriptions for each ECG and allow for a question and answer period. The analysis of the participants combined results were presented to demonstrate the groups' accuracy for individual ECG diagnoses and to compare diagnostic accuracy for pattern recognition strategies to inductive-deductive reasoning strategies.

Selection of ECGs

We reviewed the existing literature pertaining to electrocardiography to ascertain the most common ECG diagnoses studied. Using a blinded approach,

two individuals (KQ, AB) performed a systematic search strategy to select the 10 most common and clinically relevant ECG diagnoses. Where disagreement existed between the two generated lists, consensus was achieved through involvement of a third expert in both clinical education and electrophysiology, and was based on his clinical opinion.

The 10 most common ECG diagnoses were: anterior STEMI, atrial fibrillation with rapid ventricular response, complete heart block, monomorphic ventricular tachycardia (VT), pseudo VT (artifact), Wolff-Parkinson-White syndrome, left bundle branch block, hyperkalemia, long QT syndrome, and type-1 Brugada pattern (see Figs. S1 through S10 for ECGs).

Clinical vignettes from patients treated at Kingston General Hospital with the same ECG diagnosis were used to provide a clinical history to aid in the inductive-deductive component of the tool. To further test participants knowledge of basic management principles, one multiple choice question addressing the most appropriate management step according to the American College of Cardiology Guidelines⁸⁻¹² accompanied each of the 10 ECG diagnoses.

The complete set of ECGs and their accompanying vignettes were then posted to a secure online website (<https://www.queensecgstudy.ca>).

Administration of the ECG Examination

The exam is separated into two parts, each test different interpretation strategies: part 1 presents an ECG without any accompanying clinical history and therefore tests a participant's accuracy of ECG interpretation as it relates to pattern recognition. In part 2 of the exam, the same ECGs from part 1 are presented but are accompanied by a 1-2 sentence clinical vignette that provides a relevant clinical context for the patient's ECG. In this way, participants utilize inductive-deductive clinical reasoning strategies to look for specific ECG features that may enhance their diagnostic accuracy (e.g., an ECG taken from a 54-year-old man with central crushing chest pain may direct the participant to look for evidence of ST-elevation in vascular territories on the ECG).

In Step 1 of the exam, participants had 90 seconds to provide a single critical diagnosis for each ECG. Once completed, participants entered step 2 of the exam, consisting of the same 10 ECGs, presented in

random order and accompanied by a 1–2 sentence clinical vignette. Participants were not permitted to change their responses at any point during the exam after the 90-second period had passed for an individual question.

Development of the Online ECG Assessment Tool

The ECG Study web application was developed by the Medical Education Technology Unit¹³ at Queen's University.

The application securely saved participants responses to an online repository and allowed study administrators to add and remove participants, as well as score and analyze the response data saved to the database.

This application was developed using the PHP server-side programming language.¹⁴ The web interface itself utilizes Twitter Bootstrap.¹⁵ Finally, the jQuery¹⁶ JavaScript framework is used as a base for the code that automatically submits and advances study participants at 90-second intervals.

The study questions, participants, and responses are stored in a normalized manner throughout a number of tables in a dedicated MySQL¹⁷ database. MySQL is a widely adopted relational database management system (RDBMS) that is based on the Structured Query Language (SQL) ANSI/ISO standard.¹⁸

When a participant accesses the study online using their unique one-time use 32 character hash code, they are asked to provide basic demographic information (i.e., age and gender). After this information has been provided, and they have read a brief study description they can begin the uninterrupted 30-minute study exercise. The "responses" table stores each of the responses provided by study participants and allows data to be exported in .CSV format for processing in Microsoft Excel and final analysis using GraphPad Prism version 6.0.

The scoring mechanism is unique in that it provides study administrators with the ability to automatically score text-based responses. It accomplishes this by using an acceptable variations list that can be dynamically managed as study administrators review "incorrect" responses. When an acceptable variation is encountered as an "incorrect" response (e.g., hypokalaemia vs. hypokalemia), study administrators can correct the variation (i.e., fix a typo) or add the variation to

the acceptable variations list. The tool becomes more powerful and efficient in reducing the need for manual review and correction of acceptable variations as administrators build the acceptable variations response list.

Analysis of Tool Performance

The primary objective of this study was to validate the use of a novel educational tool in the assessment of ECG competency. To do so, we evaluated the following parameters: the percentage of responses the computer automatically identified as correct based on a predefined glossary of terms deemed to be acceptable as a given diagnosis (e.g., for left bundle branch block acceptable answers include "left bundle branch block"; "LBBB"; "left bundle"); and the percentage of responses requiring manual review after they had been flagged by the tool as not having recognized the given response as correct.

Statistical Analysis

Analyses were performed using GraphPad Prism, version 6.0c (GraphPad Software Inc., San Diego, CA, USA). Comparison of overall accuracy of diagnosis between pattern recognition and clinical reasoning for all residents was made using a paired *t*-test. A probability value <0.05 was considered statistically significant.

RESULTS

There were a total of 500 responses provided by the 25 attendees who participated in the study (250 for Step 1 and Step 2 each, total duration of exam is 35 minutes including registration of demographic data). Of the 500 responses, 395 were ultimately correct in the diagnosis. The online tool accurately identified 195 of these correct responses (49% of total correct responses) whereas 305 responses remained that required secondary manual review. Two hundred of these 305 responses (66% of responses requiring secondary review) were accurate in the diagnosis, but were flagged for review and required correction of typographical errors. There were 105 incorrect responses remaining that were appropriately identified by the tool as incorrect (Table 1).

The overall accuracy of correct ECG diagnosis for all participants was 69% and 84% when

Table 1. Performance of the Online Tool in Automatic Marking

Total Responses	Correct Responses Identified n (%) ^a	Responses Requiring Manual Review n (%)	Correct Responses Identified from Manual Review n (%) ^a	Incorrect Responses Identified n (%)
500	195 (49)	305 (61)	200 (51)	105 (21)

^aExpressed as percentage of total correct responses.

Table 2. Baseline Characteristics of Study Participants and their Accuracy of ECG Diagnosis Using Pattern Recognition versus Inductive-Deductive Reasoning Strategies

	Years in Practice (Mean±SEM)	Age (Years±SEM)	Gender ^a		Accuracy of Diagnosis (%±SEM) (Pattern)	Accuracy of Diagnosis (%±SEM) (Clinical)
			Male, n (%)	Female, n (%)		
All participants (25)	9.4 ± 1.9	45.0 ± 1.9	15 (60.0)	9 (36.0)	69.2 ± 4.2	83.8 ± 3.7

^aOne participant did not disclose gender.

Figure 1. Accuracy of diagnosis for individual ECGs using pattern recognition (black bars). The percentage change in accuracy using inductive-deductive reasoning strategies compared to pattern recognition is depicted in the grey bars. *P < 0.05.

using pattern recognition or inductive-deductive strategies, respectively (P < 0.01; Table 2). Fifty-five percent of participants who answered incorrectly in the first part of the study (pattern recognition strategy, 67 responses) changed their answer in the second part to a correct answer (inductive-deductive strategy, 37 responses).

There was considerable variability in the accuracy of individual ECG diagnoses (Figure 1, Table 3). The rates of accuracy in diagnosis for hyperkalemia and long QT were lower than all of the other ECG diagnoses when applying a pattern

recognition strategy. However, when participants applied an inductive-deductive reasoning strategy there was a significant improvement in the rate of accuracy. Accuracy of diagnosis decreased when using inductive-deductive reasoning strategies for LBBB, VT, and atrial fibrillation (Figure 1, Table 3).

DISCUSSION

To the best of our knowledge, our study is the first to employ the use of an automatic online marking tool to carry out an ECG competency evaluation workshop at an international Cardiology CME meeting and present the results within 24 hours to participants for immediate feedback and presentation of the group's combined results. The results were analyzed within 2 hours of completion of the exam, but their presentation was limited to within 24 hours due to the meeting's predetermined schedule. It is feasible to analyze and present the data within the same day, which will be tested at future CME meetings using our tool.

The assessment of ECG competency using nondigital examinations is hindered by the labor-intensive effort in administering the examinations, which require manual review of each individual examinee's accompanying diagnosis for a particular ECG. This appears to have limited this area of education research to one or two-center cohorts,^{2,4} as well as precluding the ability to conduct large-scale workshops on ECG interpretation at any of the major national and international cardiology meetings.

Table 3. Accuracy of Individual ECG Diagnosis Using Pattern Recognition versus Inductive-Deductive Reasoning Strategies

ECG Diagnosis	Accuracy of Diagnosis (%±SEM) (Pattern)	Accuracy of Diagnosis (%±SEM) (Inductive-Deductive)	% Change	P Value
STEMI	76.0±8.7	80.0±8.2	4.0%	0.74
Hyper K	40.0±10.0	80.0±8.2	40.0%	0.003
Complete AVB	72.0±9.2	84.0±7.5	12.0%	0.32
LBBB	80.0±8.2	72.0±9.2	-8.0%	0.52
Long QT	20.0±8.2	76.0±8.7	56.0%	0.00002
Vtach	72.0±9.2	68.0±9.5	-4.0%	0.76
PseudoVT	88.0±6.6	96.0±4.0	8.0%	0.31
Afib	96.0±4.0	87.5±6.9	-8.5%	0.29
Brugada	64.0±9.8	79.2±8.5	15.2%	0.25
WPW	84.0±7.5	95.8±4.2	11.8%	0.18

The average time to manually review a response using our tool interface took 10 seconds. If the same exam were administered at an international meeting to 25 participants without the use of our computer-marking tool, the total marking time would equate to 1.5 hours. By utilizing the automatic scoring tool, our total marking time was reduced by 39%. In our next study, we are planning to conduct an ECG competency evaluation workshop at a large international meeting that attracts over 10,000 attendees. Assuming a 10% participation rate of a workshop conducted at one of these meetings (i.e., 1,000 participants), the total marking time for 20,000 responses would require 65 man hours of manual marking, but would be reduced to 39 hours by the use of our tool. We believe the greatest strength our online tool comes from its ability to automatically recognize free-text responses, which significantly reduces the manual marking time for large-scale examinations.

The results of this validation study are in line with our previous study conducted in 125 Canadian residents containing 2500 responses. The performance of the automated marking function improved from 46% in the cohort of residents to 49% in this study. Marking time was reduced by 46% in that study compared to 39% in our validation cohort.⁷ One possible explanation for the slightly lower rate of efficiency in this study was the significantly higher baseline rate of accuracy of ECG diagnosis, which is not surprising given that IWAS participants had an average of 8 to 10 years of fulltime practice experience. Furthermore, a significantly larger proportion of correct responses were provided as an accurate pathophysiological descrip-

tion of the ECG versus a single correct diagnosis, which the tools is unable to recognize as a correct response and therefore requires manual review.

This study further supports the concept that an inductive-deductive reasoning strategy may be a more accurate approach to ECG interpretation, which may have important implications for medical education in how to teach ECG interpretation. The traditional method involves memorizing the patterns of ECGs and associating them with disease processes, but this is prone to error,¹⁻⁶ whereas a patient's clinical history significantly improves a physician's diagnostic accuracy when reading ECGs.^{1,3,5,6} The concerning lack of trainee proficiency, combined with a lack of educational research examining the most accurate method for ECG interpretation, has left academic institutions without guidance in how to teach ECG interpretation.

LIMITATIONS

One limitation of the tool was its inability to identify responses that were provided as a pathophysiologic interpretation or description of the ECG abnormality, or those responses that contain the correct diagnosis along with other correct descriptions of the ECG (e.g., rate, rhythm, axis, etc.). However, as the common combinations of these are identified through analysis of the types of responses given, the tools performance will improve to capture these and automatically mark them as correct. We are also building in the capacity for the marking function of the tool

to scan an individual free-text response containing multiple answers and mark it correct if it contains the right diagnosis. This will be evaluated in future cohorts for validation and analysis of improvement in the tool's efficiency.

CONCLUSIONS

We present here the validation of a novel tool used for evaluation of ECG competency at an international cardiology meeting. It is widely accessible through its use of a web-based approach to administer the exam, simple in its design of its user interface, flexible for modification in future iterations and is low-cost, making it an attractive tool for use at educational institutions and at society conferences. Individual ECGs can be adjusted to an individual cohort's level of expertise (i.e., nurse practitioners, medical trainees, electrophysiology fellows, and cardiology staff physicians), and used repeatedly to monitor improvement following educational sessions on ECG interpretation.

Acknowledgments: *The authors thank Matt Simpson and Ryan Warner for their valued input and the IWAS organizers to allow us to validate our tool during their prestigious meeting.*

REFERENCES

- Hatala R, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. *J Gen Intern Med* 1999;14:126–129.
- Baranchuk A, Dagnone G, Fowler P, et al. Education at distance: Broadcasting ECG rounds to Southeastern Ontario (BESO Project). An innovative approach for teaching electrocardiography. *Clin Invest Med* 2007;30:51–52.
- Fent G, Gosai J, Purva M. Teaching the interpretation of electrocardiograms: Which method is best? *J Electrocardiol* 2015;48:190–193.
- Hurst JW. Methods used to interpret the 12-lead electrocardiogram: Pattern memorization versus the use of vector concepts. *Clin Cardiol* 2000;23:4–13.
- Sibbald M, Davies EG, Dorian P, Yu EHC. Electrocardiographic interpretation skills of cardiology residents: Are they competent? *Can J Cardiol* 2014;30:1721–1724.
- Lever NA, Larsen PD, Dawes M, Wong A, Harding SA. Are our medical graduates in New Zealand safe and accurate in ECG interpretation? *N Z Med J* 2009;122:9–15.
- Quinn KL, Baranchuk A. Feasibility of a novel digital tool in automatic scoring of an online ECG examination. *Int J Cardiol* 2015;185:88–89.
- January CT, Wann LS, Alpert JS, et al. AHA/ACC/HRS Guideline for the management of patients with atrial fibrillation: A report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the Heart Rhythm Society. *J Am Coll Cardiol* 2014;64(21):e1–76.
- Levine GN, Bates ER, Blankenship JC, et al. ACCF/AHA/SCAI Guideline for percutaneous coronary intervention. *J Am Coll Cardiol* 2011;58:e44–e122.
- O'Gara PT, Kushner FG, Ascheim DD, et al. ACCF/AHA guideline for the management of ST-elevation myocardial infarction: A report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. *J Am Coll Cardiol* 2013;61:e78–e140.
- Epstein AE, DiMarco JP, Ellenbogen KA, et al. American College of Cardiology/American Heart Association Task Force on practice guidelines (writing committee to revise the ACC/AHA/NASPE 2002 guideline update for implantation of cardiac pacemakers and antiarrhythmia devices), American Association for Thoracic Surgery, Society of Thoracic Surgeons. ACC/AHA/HRS 2008 guidelines for device-based therapy of cardiac rhythm abnormalities: A report of the American College of Cardiology/American Heart Association Task Force on practice guidelines (writing committee to revise the ACC/AHA/NASPE 2002 guideline update for implantation of cardiac pacemakers and antiarrhythmia devices): Developed in collaboration with the American Association for Thoracic Surgery and Society of Thoracic Surgeons. *Circulation* 2008;117:e350–e408.
- Zipes DP, Camm AJ, Borggrefe M, et al. ACC/AHA/ESC 2006 Guidelines for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death. *J Am Coll Cardiol* 2006;48:e247–e346.
- Queen's Medical Education Technology Unit [Internet]. Available at: <https://healthsci.queensu.ca/technology>. Accessed February 1, 2014.
- PHP [Internet]. Available at: <http://php.net>. Accessed February 1, 2014.
- Twitter Bootstrap [Internet]. Available at: <http://getbootstrap.com>. Accessed February 1, 2014.
- jQuery [Internet]. Available at: <http://jquery.com>. Accessed February 1, 2014.
- MySQL [Internet]. Available at: <http://www.mysql.com>. Accessed February 1, 2014.
- Structured Query Language ANSI/ISO [Internet]. Available at: http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=53681. Accessed February 1, 2014.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

- Figure S1.** Anterior STEMI.
- Figure S2.** Hyperkalemia.
- Figure S3.** Complete heart block.
- Figure S4.** Left bundle branch block.
- Figure S5.** Long QT.
- Figure S6.** Monomorphic ventricular tachycardia.
- Figure S7.** Pseudo-ventricular tachycardia.
- Figure S8.** Rapid atrial fibrillation.
- Figure S9.** Type-1 Brugada ECG pattern.
- Figure S10.** Wolf-Parkinson-White.