TEACHING INNOVATIONS

Comparison of two formats of journal club for postgraduate students at two centers in developing critical appraisal skills

Manpreet Kaur,1 Hanjabam B. Sharma,2 Simran Kaur,2 Ratna Sharma,2 Renuka Sharma,1 Raj Kapoor,1 and Kishore K. Deepak2
1Department of Physiology, Vardhman Mahavir Medical College and Safdarjung Hospital, New Delhi, India; and
2Department of Physiology, All India Institute of Medical Sciences, New Delhi, India

Submitted 8 August 2019; accepted in final form 13 August 2020

Kaur M, Sharma HB, Kaur S, Sharma R, Sharma R, Kapoor R, Deepak KK. Comparison of two formats of journal club for postgraduate students at two centers in developing critical appraisal skills. Adv Physiol Educ 44: 592–601, 2020; doi:10.1152/advan.00111.2019.—Journal club (JC) is an integral part of postgraduate medical education. Although several innovations have been attempted to improve its effectiveness, the influence of the number of sessions remains unexplored. The current study investigated the effect of the number of sessions (one vs. four) of conduction of journal club on critical appraisal skills of postgraduate students at two institutions. A quasi-experimental study was conducted comparing the effectiveness of JC in a one- versus four-session JC format using two objective tests: Critical Appraisal Skills Test (CAST) and Population, Intervention, Comparison, Outcome (PICO) score. The degree of changes (pre-JC vs. post-JC session) after each format exposure, association of prescores to postscores, and correlation between the CAST and PICO scores were discerned. CAST score was higher in the four-session format (P = 0.03), whereas PICO score increased in the single-session format of JC (P = 0.03). Also, the change in CAST score was higher in the four-session versus single-session JC format (P = 0.04). We conclude that a four-session format of JC is more effective in augmenting critical appraisal skills of postgraduate students, whereas a single session is sufficient for improving factual knowledge.

Critical Appraisal Skills Test; medical education; objective tests; one versus four session; Population Intervention Comparison Outcome

INTRODUCTION

In the present era, evidence-based medicine (EBM) has emerged as an important component of medical practice (13, 16, 23). An essential part of EBM is the ability to critically appraise the available literature to assess its content reliability and relevance (1). Substantial efforts are being made in many countries to provide critical appraisal skills training as part of healthcare professionals’ undergraduate and postgraduate activities. Journal clubs (JC) have been an essential part of the teaching-learning process for postgraduate and undergraduate training (10, 39). The major goals of JC include acquisition of critical appraisal skills, keeping abreast of the present medical literature, translating the learned knowledge and skills into evidence-based practice, and providing training in research and education (4, 9, 14). Additionally, it has been used as a tool to gain in-depth knowledge of the concerned topic and provide an opportunity to increase the social interaction between postgraduate students and faculty members (9, 14). Although there exist multiple methodologies and inter/intrdisciplinary diversity in the format of JC conduction, imparting critical appraisal skills still remains one of the most important learning objective of JC sessions (41).

Although JC is an established medical education tool, it has been moderately successful in maintaining the students’ interest and achieving the targeted educational objectives (40). For the presenting students, JC is often seen as a daunting process, as they have not been adequately prepared for this in their undergraduate medical education (7). Some studies have also reported discontinuation of JC, mainly due to lack of time, inadequate preparation, and lack of goals, interest, or participation (14). Thus the need of the hour is to design an effective JC format, which is stimulating, interesting, sustainable, and educational for the students (9).

There is a vast diversity in the format of JC being conducted at the departmental as well as institutional level (33). To date, there is no consensus on the procedure for conduction of JC or methods of evaluation of its effectiveness in achieving the desired objectives (9). Many authors have tried to change the traditional structure of JC by furnishing it with novel objectives and construction and subsequently found an increase in attendance and competency of the postgraduates (3, 25, 27, 33, 36). Hence, continued steps for initiating modifications may lead to improvement in the current format of JC, such that it may successfully achieve its intended objectives.

A monthly discussion of a single research article is the most accepted format of JC in most of the universities (21, 22, 27, 35). Literature regarding the ability of the traditional single monthly session of JC in enhancing the critical appraisal skills of the postgraduates has conflicting findings, with few studies suggesting improvement (17, 27, 35, 38), whereas some report no significant change (5, 8, 12, 21, 24). The traditional JC is less effective in facilitating the deep learning process (which essentially means learning something with full understanding so that one remembers it for a longer time). It is hypothesized that increasing the number of sessions of JC discussion would enhance this process by allowing the students to progress sequentially by initially using their existing knowledge on the subject, building on new knowledge over and above it, and finally consolidating the newly acquired knowledge.

To inculcate in-depth learning (detailed and thorough understanding) of the research paper for JC and induce active
involvement of the participants, innovations in terms of the number of sessions have been done. Kitching and Ross (19) structured JC into four distinct steps, wherein in the first session the case is described, followed by search strategy to obtain relevant literature. Subsequently, the article is discussed with respect to the review and its applicability in clinical practice. The Department of Physiology in All India Institute of Medical Sciences (AIIMS), New Delhi, India, has been practicing a four-session JC format in which the chosen scientific paper is discussed in four sessions over 1 mo (1 session per week). Thus, perhaps by increasing the number of sessions of JC, one could achieve more enhancement in the overall critical appraisal skills of postgraduates, which is the most significant intended objective of a JC.

To the best of our knowledge, there is paucity of literature investigating the difference between the above two formats of JC conduction, in terms of the critical appraisal skills acquired by the postgraduate students. Hence, the present study was designed with an aim to evaluate the effectiveness of one-session versus four-session JC in terms of enhancing the critical appraisal skills of the students at the two centers.

METHODS

We conducted a quasi-experimental study at the Department of Physiology, AIIMS, New Delhi, India (center 1), and Vardhman Mahavir Medical College and Safdarjung Hospital, New Delhi, India (center 2). The participants of the study were the postgraduate students [MD (Medical Doctor), MSc (Master of Science), and PhD (Doctor of Philosophy) scholars] from both the centers with physiology as their major subject. The study was approved by the institutional ethics committee at both the centers (IEC/263/5/2018; IEC/VMMC/ SJH/Project/July/2018/1089). Informed, written consent was obtained from all of the participants. The nature and purpose of the study was explained in a language comprehensible to them, and they were informed that they were free to withdraw at any time from the study, without giving any reason. Although all postgraduate students attended the JC and took part in the discussion, data for analysis was generated using suitable questionnaires only from those who gave informed consent for participation in the study.

The study design was composed of introducing the participants to both formats of journal club (JC) and collecting the data of their critical appraisal skills before and after exposure to each of the formats of JC. The step-by-step sequence of events in the process of preparation, conduction, and analysis is given in Fig. 1 and is elaborated henceforth.

The JC paper was chosen by or on the recommendation of the faculty in charge of postgraduate teaching, who then sent it to all of the faculty, including the moderator of the JC, and to all of the postgraduate students, at least 1 mo before the final JC presentation. Scientific papers of equivalent academic quality were selected for both types of JC sessions in the two centers.

We conducted the two formats of JC, single-session and four-session, at both of the centers. In the single-session JC, a single student was allotted the JC paper 1 mo in advance, with an assigned faculty moderator. The faculty moderator guided and prepared the presenting student for the final deliberation via frequent informal meetings before the actual presentation. During the final session, the student presented the paper in front of all of the departmental faculty and postgraduate students in a 2-h long session, wherein questions were asked of the presenter, who made attempts to answer and clarify doubts being raised and principally led the discussion. The other attendees further interacted and clarified the doubts, if required.

However, in the four-session JC, three sessions of 2 h each, once a week for 3 wk, were conducted before the final presentation. These sessions involved the moderator, presenting group (comprising 3 or 4 students), and all other postgraduates. In the four-session JC, the following protocol was being followed:

Session 1: The JC paper is read by the postgraduates, who come up with a list of learning objectives that they want to achieve as a group based on the JC paper. These objectives are divided among the postgraduates who are supposed to read in detail and discuss the same with their peers in the subsequent session.

Session 2: The postgraduates discuss their respective topics and address queries of the group. The group as a whole tries to attain the learning objectives decided by them in the first session.

Session 3: The presenting group gives a mock presentation of the JC paper in detail in the format of IMRAD (Introduction, Methods, Results, and Discussion) to the other postgraduates and clarifies their doubts.

Session 4: This is the final presentation with the entire faculty and the postgraduates together. Here the presenting group formally presents the JC paper in the IMRAD format. There is an active interaction between the faculty members and all of postgraduates to ensure the educational objectives of JC are met.

For a given semester, a single cycle, i.e., one single-session and one four-session JC was done at both of the centers.

The aim of this study was to compare the critical appraisal skills in a group-based multisession versus individual-based single-session JC format at two centers. The primary objective was to assess the improvement in the critical appraisal skills after exposure to a single-versus four-session format of JC, and the secondary objective was to assess the differences between the two centers involved in the present study.

On the pattern of Alavi-Moghaddam et al. (3), we used a fictitious paper to test the critical appraisal skills of the postgraduate students. It is provided in Supplemental Material S1 (available at https://doi.org/10.6084/m9.figshare.12363566). To assess the critical appraisal skills, two objective measures, Critical Appraisal Skills Test (CAST) score (3) and Population Intervention Comparison Outcome (PICO) score (2), were used as a part of questionnaire. It can be accessed in Supplemental Material S2 (available at https://doi.org/10.6084/m9.figshare.12971060). The questionnaire involved critical analysis of a fictitious manuscript, in which major (e.g., improper randomization, data not supporting the conclusion, major ethical concerns, statistical errors, etc.) and minor flaws (e.g., methodological errors that were not emphasized in EBM approach to critical appraisal, errors in referencing, spelling mistakes, grammatical and language errors, etc.) were intentionally incorporated in an already published paper (37) to test the critical appraisal skills of the postgraduates.

Five faculty members (subject experts) at both the centers analyzed the fictitious manuscript and achieved consensus score regarding the major and minor flaws of the manuscript and validated the critical appraisal skills test (based on Delphi method) (29, 31). The process of validation is provided in Supplemental Material S3 (available at https://doi.org/10.6084/m9.figshare.12971081).

Two fictitious manuscripts, with a similar difficulty level, as decided by the subject experts, were used at the two centers, one for single-session and another for four-session JC. Both of the centers were exposed to the same manuscript for the corresponding format of JC session. The questionnaire to assess the critical appraisal skills was given, along with a copy of the fictitious manuscript, under the supervision of a faculty to each of the participating students. A 1-h duration was given to the postgraduate students to fill out the questionnaire and submit it to the concerned faculty. They were expected to critically analyze the manuscript given and detect those flaws that were intentionally incorporated. Based on the number of flaws successfully detected by the postgraduates, they were respectively scored as follows, i.e., for each major flaw detected, two points were given, and for each minor flaw, one point was given (3). The total score was converted into a percentage and reported as the CAST score of the
student. We quantified the competency of the critical appraisal skills of the postgraduate students by using this objective scoring.

Additionally, knowledge regarding the research question of the study was assessed by asking the postgraduates about the PICO analysis of the given manuscript (2). A well-built research question has essentially four basic components: population: who was studied; intervention or exposure: what was the therapy, risk factor, tests, etc.; comparison or control: what was the alternative to intervention or exposure; and outcome: clinical, functional, economic, etc. (2). Outlining the PICO of a research paper is essentially the first step in critical appraisal of literature in EBM. The students were expected to write the corresponding response to each component of the PICO, which was assessed and scored, with one point for each of the component of the PICO given (maximum score of 4). All of the subjects from both of the centers were made to fill out the pre-JC and post-JC questionnaire form (33 for single-session and 29 for four-session format). The above-mentioned procedure was followed for both the single-session and four-session JC in the two centers.

Thus the main hypothesis of the study was that the four-session JC format will be a better tool for teaching critical appraisal skills.

The rationale behind the analysis were based on the following questions:
1. Whether exposure to the JC format changes the critical appraisal skills scores (pre-JC vs. post-JC comparison).
2. Whether the critical appraisal skills scores differed in the two formats of JC (single- versus four-session comparison).
3. Whether exposure to the JC format changes the critical appraisal skills scores at each of the centers (pre-JC vs. post-JC comparison in center 1 and center 2).
4. Whether the critical appraisal skills scores differed in the two formats of JC at the two centers (single- vs. four-session comparison in center 1 and center 2).
5. Whether the critical appraisal scores differed between the two centers (center 1 vs. center 2 comparison of single- and four-session JC format).
6. Whether there is any correlation between the baseline parameters with post-JC data scores or between the two scores used in this study.

Statistical Analysis

We analyzed data using IBM SPSS (Statistical Package for the Social Sciences) Statistics for Windows, version 25.0 (IBM Corp., Armonk, NY). Normality tests were done using the Shapiro–Wilk test and visual methods (histogram and Q-Q or quantile-quantile plot). Pre-JC versus post-JC study parameter comparison for normally distributed data was done using paired t test (overall single-session JC and at center 1 for both single- and four-session JC format), while all other pre-JC versus post-JC comparisons were performed using the Wilcoxon signed ranks test (as data were nonparametric). Independent group comparisons (single- vs. four-session or center 1 vs. center 2) were done using unpaired t test for parametric data (at center 1: single- vs. four-session CAST percentage comparison for both pre-JC and post-JC scores), whereas the Mann–Whitney test was used for all others (nonparametric data). Additionally, correlation of the pre-JC score of CAST percentage and PICO was done with their corresponding post-JC score and the change in the score (post-JC minus pre-JC), respectively. Furthermore, correlation was also performed between change in CAST percentage and change in PICO score. We evaluated association between these selected variables using Pearson correlation for parametric variables (pre-JC CAST percentage combined and at center 1 with post-JC CAST percentage), while all other nonparametric variables used Spearman correlation. Statistical significance was taken at P value (two-tailed) of <0.05.

RESULTS

We analyzed 33 and 29 pre-post (pre-JC vs. post-JC session) responses from both of the centers for single-session and four-session JC, respectively. From center 1, the responses analyzed for single- and four-session JC were 26 and 12, respectively, whereas 7 and 17 responses were analyzed, respectively, for center 2 (refer to Fig. 1). All of these participants were present for both the pre-JC and post-JC evaluation of critical appraisal skills test (CAST) and population, intervention, comparison, outcome (PICO) scores, and hence there was 100% compliance, and no drop out.

Result 1: Pre-JC Versus Post-JC Comparison in Single- and Four-Session JC Format

The changes in CAST and PICO scores in the postgraduate students after exposure to the JC session (single-session and four-session, respectively) were compared (Table 1, Fig. 2, A and B).

The two scores showed a dichotomy in the results with respect to the pre-post change as shown below: a. CAST had a statistically significant pre-post improvement after four-session JC (P = 0.04). b. PICO had a statistically significant increase after a single-session JC (P = 0.03).

This finding reveals that PICO and CAST scores probably assess two different areas of critical assessment skills and thus show a differential improvement, wherein a single-session JC suffices to improve the PICO score, whereas four-session JC improves the CAST score of postgraduate students.

Table 1. Comparison of CAST and PICO parameters during conduction of single- and four-session format of JC at the two centers

| JC Session/Parameters | Pre-JC Session | Post-JC Session | Change (Post-Pre) | P Valueb |
|-----------------------|---------------|----------------|-----------------|----------|
| CAST, %               |               |                |                 |          |
| Single-session (n = 33) | 19.70 (9.76)  | 18.64 (9.62)  | 0.00 (5.00)     | 0.52c    |
| Median (QD)           | 0, 35         | 15.22–22.05   | –25, 15         | 0–5      |
| Minimum, maximum      | 16.24–23.16   |                |                 |          |
| 95% CI                |               |                |                 |          |
| Four-session (n = 29) | 12.12 (10.46) | 18.18 (10.19) | 3.23 (5.00)     | 0.04c,d  |
| Median (QD)           | 0, 60         | 0, 70          | –24.24, 20      |          |
| Minimum, maximum      | 6.45–24.24    | 9.09–24.24     | 3.03–10         |          |
| 95% CI                |               |                |                 |          |
| P valueb              | 0.14c         | 0.77c          | 0.03c,d         |          |
| PICO score            |               |                |                 |          |
| Single-session (n = 33) | 3.00 (0.50)   | 3.00 (0.50)    | 0 (0)           | 0.03c,d  |
| Median (QD)           | 0, 4          | 1, 4           | 0, 3            |          |
| Minimum, maximum      | 2–3           | 2–3            | 0–0             |          |
| 95% CI                |               |                |                 |          |
| Four-session (n = 29) | 3.00 (1.00)   | 4.00 (1.00)    | 0 (0)           | 0.07d    |
| Median (QD)           | 0, 4          | 0, 4           | 0, 4            |          |
| Minimum, maximum      | 2–4           | 3–4            | 0–0             |          |
| 95% CI                |               |                |                 |          |
| P valueb              | 0.13f         | 0.044f         | 0.94f           |          |

Values are median with quartile deviation (QD) in parentheses, minimum and maximum, and 95% confidence interval (CI). n, No. of students. CAST, Critical Appraisal Skills Test; JC, journal club; PICO, Population Intervention Comparison Outcome. bP value for comparison of pre-JC and post-JC CAST and PICO parameters in single- and four-session JC, columnwise. bP value for comparison of pre-JC and post-JC CAST and PICO parameters between single- and four-session JC done at the two centers, rowwise. cP value < 0.05. dWilcoxon signed ranks test. ePaired t test. f Mann–Whitney Test.

Advances in Physiology Education • doi:10.1152/advan.00111.2019 • http://advan.physiology.org
The two centers were dissimilar in the pre-post improvement profile of CAST and PICO scores.

The only significant finding observed in center 1 was the pre-JC versus post-JC increase in PICO after a single-session JC (in line with the findings of result 1b). However, there was no significant change in CAST after both single- and four-session JC, and no change in PICO after four-session JC.

At center 2, CAST had an outcome similar to result 1a, with a significant pre-post improvement after a four-session JC, while no change in PICO score (pre-JC vs. post-JC) was observed in both single- and four-session JC. Also, similar CAST scores (pre-JC and post-JC) were reported in single-session JC.

A noteworthy difference between the two centers was the format of JC being followed previously: at center 1 were adherents of four-session JC, and center 2 had a single-session JC format. We propose that this difference in their previous exposure could account for the distinctive results obtained at the two centers.

Result 4: Single- Versus Four-Session Comparison in Center 1 and Center 2

We wanted to test whether the pre-JC, post-JC, and change (pre-post) scores in single- versus four-session JC format were distinct at the two centers (Table 2).

The only significant finding of this analysis was that, at center 1, the value of the PICO score (post-JC) was higher in four-session than single-session JC.

Result 5: Center 1 Versus Center 2 Comparison of Single- and Four-Session JC Format

As a part of the secondary objective, we wanted to observe whether the two centers were similar in the CAST and PICO (pre-, post-, and change) scores after single- and four-session JC (Table 3).

The delta change in the CAST score in the four-session JC format was greater in center 2 compared with center 1. This observation further emphasizes the difference in exposure between the two centers and is in accordance with result 3 (where we observed that center 2 had a significant pre-post change in CAST after a four-session JC). Thus, as expected, the delta change in CAST in four-session JC is higher in center 2 than center 1.

Result 6: Correlational Analysis

To analyze further the complex relationship between these parameters, correlational analysis was done between pre-JC, post-JC, and change of CAST and PICO together at the two centers and at center 1 and center 2 separately (Table 4).

Pre-CAST to post-CAST and change in CAST score in single- and four-session JC.

1. Combined results: In the single-session, the pre-CAST score correlated to the post-CAST score and change, whereas, in the four-session, the pre-CAST score correlated to the post-CAST score.

2. At center 1: In both single- and four-session JC, the pre-CAST score correlated to both post-CAST score and change in score.
In both single- and four-session JC, pre-CAST score correlated to the post-CAST score.

The uniformly positive association of pre-CAST score with post-CAST score, as observed above, is suggestive that a higher prescore would obviously culminate in a higher postscore and vice versa, whereas the negative association between change and pre-CAST score seen in combined (single) and center 1 (single- and four-session) suggests that those with lower prescores had higher improvement in CAST after JC exposure. This establishes the fact that, in single-session, clearly prescore determines the degree of improvement in CAST.

Pre-PICO to post-PICO and change in PICO score in single- and four-session JC.

At center 1: In both single- and four-session pre-PICO score correlated to both post-PICO score and change in score.

At center 2: In both single- and four-session pre-PICO score correlated to both post-PICO score and change in score.

Table 2. Centerwise comparison of CAST and PICO parameters in single- and four-session format of JC

| JC Session/Parameters | Pre-JC Session | Post-JC Session | Change (Post-Pre) | P Valuea |
|-----------------------|----------------|----------------|------------------|---------|
| **CAST, %**           |                |                |                  |         |
| Single-session (n = 26)|                |                |                  |         |
| Median (QD)           | 20.00 (9.27)   | 17.88 (8.62)   | 0.00 (5.63)      | 0.28e   |
| Minimum, maximum      | 0, 35          | 0, 35          | −25, 15          |         |
| 95% CI                | 16.25–23.75    | 14.40–21.37    | 0–5              |         |
| Four-session (n = 12) |                |                |                  |         |
| Median (QD)           | 15.66 (13.29)  | 14.65 (9.57)   | 3.03 (3.79)      | 0.73e   |
| Minimum, maximum      | 0, 42.42       | 3.03, 30.3     | −24.24, 9.09     |         |
| 95% CI                | 7.21–24.10     | 8.57–20.73     | 0–6.06           |         |

| P valueb              | 0.25f          | 0.31g          | 0.36f            |         |
| **PICO score**        |                |                |                  |         |
| Single-session (n = 26)|                |                |                  |         |
| Median (QD)           | 2.50 (0.63)    | 3.00 (0.50)    | 0.00 (0.00)      | 0.03c,d |
| Minimum, maximum      | 0, 4           | 1, 4           | 0, 3             |         |
| 95% CI                | 2–3            | 2–3            | 0–0              |         |
| Four-session (n = 12) |                |                |                  |         |
| Median (QD)           | 4.00 (0.88)    | 4.00 (0.50)    | 0.00 (0.00)      | 0.18d   |
| Minimum, maximum      | 0, 4           | 1, 4           | 0, 4             |         |
| 95% CI                | 2–4            | 3–4            | 0–0              |         |

| P valueb              | 0.05f          | 0.01c,f        | 0.93f            |         |

| **CAST, %**           |                |                |                  |         |
| Single-session (n = 7)|                |                |                  |         |
| Median (QD)           | 25.00 (12.50)  | 25.00 (15.00)  | 5.00 (5.00)      | 0.33d   |
| Minimum, maximum      | 0, 30          | 5, 35          | −10, 10          |         |
| 95% CI                | 0–30           | 5–35           | 0–10             |         |
| Four-session (n = 17) |                |                |                  |         |
| Median (QD)           | 10.00 (14.28)  | 22.58 (24.27)  | 9.68 (7.26)      | 0.02c,d |
| Minimum, maximum      | 0, 60          | 0, 70          | −10, 20          |         |
| 95% CI                | 6.45–30        | 9.68–50        | 3.23–16.13       |         |

| P valueb              | 0.77f          | 0.92f          | 0.24f            |         |

| **PICO score**        |                |                |                  |         |
| Single-session (n = 7)|                |                |                  |         |
| Median (QD)           | 3.00 (1.00)    | 3.00 (1.00)    | 0.00 (0.00)      | 1.0d    |
| Minimum, maximum      | 2, 4           | 2, 4           | 0, 0             |         |
| 95% CI                | 2–4            | 2–4            | 0–0              |         |
| Four-session (n = 17) |                |                |                  |         |
| Median (QD)           | 3.00 (1.00)    | 3.00 (1.00)    | 0.00 (0.00)      | 0.18d   |
| Minimum, maximum      | 0, 4           | 0, 4           | 0, 2             |         |
| 95% CI                | 2–4            | 2–4            | 0–0              |         |

| P valueb              | 0.87f          | 0.79f          | 0.35f            |         |

Values are median with quartile deviation (QD) in parentheses, minimum and maximum, and 95% confidence interval (CI). n, No. of students. CAST, Critical Appraisal Skills Test; JC, journal club; PICO, Population Intervention Comparison Outcome. *P value for centerwise comparison of pre-JC and post-JC CAST and PICO parameters in single- and four-session JC, columnwise. bP value for centerwise comparison of pre-JC and post-JC CAST and PICO parameters between single- and four-session JC done at the two centers, rowwise. cP value < 0.05. dWilcoxon signed ranks test. ePaired t test. fMann–Whitney test. gUnpaired t test.

3. At center 2: In both single- and four-session JC, pre-CAST score correlated to the post-CAST score.

The uniformly positive association of pre-CAST score with post-CAST score, as observed above, is suggestive that a higher prescore would obviously culminate in a higher postscore and vice versa, whereas the negative association between change and pre-CAST score seen in combined (single) and center 1 (single- and four-session) suggests that those with lower prescores had higher improvement in CAST after JC exposure. This establishes the fact that, in single-session, clearly prescore determines the degree of improvement in CAST.

Pre-PICO to post-PICO and change in PICO score in single- and four-session JC.

1. Combined results: In both single- and four-session JC, pre-PICO score correlated to both post-PICO score and change in score.

2. At center 1: Similar to above, in both single- and four-session, pre-PICO score correlated to both post-PICO score and change in score.

3. At center 2: In both single- and four-session JC, pre-PICO score correlated to post-PICO score.

In accordance with CAST, the positive association of pre-PICO score with post-PICO score unambiguously seen is suggestive of the fact that a higher prescore would unquestionably lead to a higher postscore and vice versa. Whereas the negative association between change and pre-PICO score seen in combined (single and four) and center 1 (single- and four-session) suggests that those with lower prescores had higher improvement in PICO after JC exposure.

Change in PICO with change in CAST in single- and four-session JC (combined and at center 1 and center 2). No association was found suggesting that the two scores are independent of each other and are likely to be assessing different aspects of critical appraisal.
The salient findings of the present study are:

- Four-session format of journal club (JC) is more effective in enhancing the in-depth critical appraisal skills of postgraduate students.
- Single-session format of JC is more beneficial in improving the factual knowledge of postgraduate students.
- Centerwise differences in the results lie in the context of the novelty of the introduced format in the respective centers.

**DISCUSSION**

The salient findings of the present study are:

| Parameters | Center 1 (n = 26) | Center 2 (n = 7) | P Value |
|------------|-------------------|-----------------|---------|
| CAST, %    |                   |                 |         |
| Pre-JC     |                   |                 |         |
| Median (QD)| 20.00 (9.27)      | 25.00 (12.50)   | 1.0     |
| Minimum, maximum | 0, 35 | 0, 30 | |
| 95% CI     | 16.25–23.75       | 0–30            |         |
| Post-JC    |                   |                 |         |
| Median (QD)| 17.88 (8.62)      | 25.00 (15.00)   | 0.46    |
| Minimum, maximum | 0, 35 | 5, 35 | |
| 95% CI     | 14.40–21.37       | 5–35            |         |
| Change (post-pre) | 0.00 (5.63) | 5.00 (5.00) | 0.16 |
| Minimum, maximum | −25, 15 | −10, 10 | |
| 95% CI     | 0–5               | 0–10            |         |
| PICO Score |                   |                 |         |
| Pre-JC     |                   |                 |         |
| Median (QD)| 2.50 (0.63)       | 3.00 (1.00)     | 0.35    |
| Minimum, maximum | 0, 4 | 2, 4 | |
| 95% CI     | 2–3               | 2–4             |         |
| Post-JC    |                   |                 |         |
| Median (QD)| 3.00 (0.50)       | 3.00 (1.00)     | 0.73    |
| Minimum, maximum | 1, 4 | 2, 4 | |
| 95% CI     | 2–3               | 2–4             |         |
| Change (post-pre) | 0.00 (0.00) | 0.00 (0.00) | 0.22 |
| Minimum, maximum | 0, 3 | 0, 0 | |
| 95% CI     | 0–0               | 0–0             |         |

Values are median with quartile deviation (QD) in parentheses, minimum and maximum, and 95% confidence interval (CI). n, No. of students. CAST, Critical Appraisal Skills Test; JC, journal club; PICO, Population Intervention Comparison Outcome. *P value < 0.05 (Mann–Whitney test).

**Table 3. Journal club sessionwise comparison of pre-JC, post-JC, and change in CAST and PICO parameters between the two centers**

CAST, %

| Parameters | Center 1 (n = 29) | Center 2 (n = 7) | P Value |
|------------|-------------------|-----------------|---------|
| CAST, %    |                   |                 |         |
| Pre-JC     |                   |                 |         |
| Median (QD)| 15.66 (13.29)     | 10.00 (14.28)   | 0.93    |
| Minimum, maximum | 0, 42.42 | 0, 60 | |
| 95% CI     | 7.21–24.10        | 6.45–30         |         |
| Post-JC    |                   |                 |         |
| Median (QD)| 14.65 (9.57)      | 22.58 (24.27)   | 0.35    |
| Minimum, maximum | 3.03, 30.3 | 0, 70 | |
| 95% CI     | 8.57–20.73        | 9.68–50         |         |
| Change (post-pre) | 3.03 (3.79) | 9.68 (7.26) | 0.02* |
| Minimum, maximum | −24.24, 9.09 | −10, 20 | |
| 95% CI     | 0–6.06            | 3.23–16.13      |         |
| PICO Score |                   |                 |         |
| Pre-JC     |                   |                 |         |
| Median (QD)| 4.00 (0.88)       | 3.00 (1.00)     | 0.30    |
| Minimum, maximum | 0, 4 | 0, 4 | |
| 95% CI     | 2–4               | 2–4             |         |
| Post-JC    |                   |                 |         |
| Median (QD)| 4.00 (0.50)       | 3.00 (1.00)     | 0.18    |
| Minimum, maximum | 1, 4 | 0, 4 | |
| 95% CI     | 3–4               | 2–4             |         |
| Change (post-pre) | 0.00 (0.00) | 0.00 (0.00) | 0.69 |
| Minimum, maximum | 0, 4 | 0, 2 | |
| 95% CI     | 0–0               | 0–0             |         |
critical appraisal skills (6, 9, 12, 15, 29, 30). JC, besides being a proven effective postgraduate teaching tool, has also been successfully implemented in undergraduate medical students, to inculcate critical thinking and real-world application skills (10, 11, 26, 34). Many variations exist in terms of frequency of conduction (weekly or monthly), facilitators (subject expert, faculty or resident themselves), attendance (mandatory or not), and outcome assessment instruments (Likert scale, Delphi method, written feedback). Several innovations have been tried to improvise on the effectiveness and productivity of JC, like using a more structured format (5, 8), varying the facilitator (22), use of internet for its conduction (27, 28), and use of more objective criteria to determine its efficacy in meeting the desired learning goals (3, 5, 25). One such modification could be the number of sessions dedicated to discussing a single research article to successfully teach and develop critical appraisal skills in the participating postgraduates. Thus the focus of the present paper was to assess the critical appraisal skills (using two objective scores: CAST and PICO) in two different formats of JC by varying the number of sessions allotted to a single research paper (one-session vs. four-session).

One-Session Versus Four-Session JC Format

Interestingly, a differential pattern of results in the augmentation of critical appraisal skills was found in the two formats of JC. While the CAST score significantly improved in four-session JC on post-post comparison, the PICO score significantly increased in a single-session JC. Furthermore, on comparing single- versus four-session JC, we found that there was a significant improvement in the change in CAST score during four-session versus single-session, whereas the PICO score, which deals with factual information, was significantly higher after a single session. These findings suggest that more sessions of discussion of a JC would culminate in better skills in terms of critical evaluation of flaws and usefulness of a research article, as deciphered from the CAST score, whereas a single session of discussion is sufficient to appraise about the PICO analysis (which is also a part of critical appraisal skills). To the best of our knowledge, ours is the first study assessing the impact of one- versus four-session JC format in imparting critical appraisal skills. Several studies have assessed the effect of the one-session JC format (most commonly followed pattern) on critical appraisal skills using objective tests with confounding results: some suggest an improvement (17, 18, 35), whereas others documented no change (25). Moreover, only two of these studies have used validated tests (18, 25), whereas others have not reported on validation-related data (17, 35). Additionally, Linzer et al. (25) found that the greater number of (single-session format type) JCs a subject attended, the more knowledge was gained by them, which they depicted in a “dose response” relationship (r = 0.40, P < 0.07). Interestingly, they reported that the knowledge scores increased more in the JC group compared with the control group, and a trend was found toward more knowledge gained as more sessions were attended, but the two groups did not differ in critical appraisal skills (25). This study is in line with our hypothesis that more sessions of discussion increase deep learning and, due to more sessions focused on discussion of a single research paper, will also result in better critical appraisal skills, which they failed to achieve in their single-session study design. A study by Phillips et al. (32) discusses the different variations in JC format, one-session to four-session cycle, and mention that they follow a two-session cycle versus the earlier three-session cycle (due to more boredom observed in this pattern in their study group). However, they have not compared the two formats in terms of their ability to augment critical appraisal skills.

Centerwise Differences

We further performed a detailed analysis of centerwise differences, as the previous format of JC at the two places were different. We found that, at center 1 (AIIMS), where previously also four-session JC format was being followed, there was no significant improvement in CAST after four-session JC, whereas PICO had a significant accentuation after a single-session JC. However, at center 2 (Vardhman Mahavir Medical Institute of Medical Sciences, New Delhi) the change was significant in the CAST score (P = 0.020). The change was also significant in the PICO score (P = 0.001).

Table 4. Correlational analysis of PICO and CAST scores (combined and centerwise)

| Variable                        | Combined | Center 1 | Center 2 |
|---------------------------------|----------|----------|----------|
| Post-JC CAST, %                 | 0.528f, 0.718 | 0.400f, 0.659d | 0.815, 0.716 |
| CC value                        |          |          |          |
| Post-JC CAST, %                 | 0.002d, <0.001f | 0.043d, <0.001d-e, 0.020d-e | 0.020d, <0.001f |
| Change in CAST, %               | -0.428, 0.066 | -0.558, -0.667 | -0.206, 0.287 |
| CC value                        | 0.013f, 0.733 | 0.003f, 0.018f | 0.658, 0.264 |
| Change in CAST, %               | -0.428, -0.373 | -0.453, -0.661 | —, -0.108 |
| CC value                        | 0.013f, 0.046f | 0.020f, 0.019f | —, 0.680 |
| Change in CAST, %               | 0.143, -0.022 | 0.235, -0.082 | —, -0.097 |
| CC value                        | 0.426, 0.910 | 0.248, 0.799 | -0.711 |
| Change in PICO Score            |          |          |          |
| CC value                        | <0.001f, <0.001f | <0.001f, 0.011f | <0.001f, <0.001f |
| Change in PICO Score            | -0.428, -0.373 | -0.453, -0.661 | —, -0.108 |
| CC value                        | 0.013f, 0.046f | 0.020f, 0.019f | —, 0.680 |
| Change in CAST, %               | 0.143, -0.022 | 0.235, -0.082 | —, -0.097 |
| CC value                        | 0.426, 0.910 | 0.248, 0.799 | -0.711 |

CAST, Critical Appraisal Skills Test; CC, correlation coefficient, JC, journal club. PICO, Population Intervention Comparison Outcome. aBetween pre-JC CAST with post-JC CAST and change in CAST. bBetween pre-JC PICO with post-JC PICO and change in PICO. cBetween change in CAST and change in PICO. dCC with Pearson correlation coefficient. All other CC Spearman correlation coefficients: *P value < 0.05 and †P value < 0.01.
College and Safdarjung Hospital), the opposite pattern was observed: CAST had a significant increase after four-session JC, whereas no significant change was reported in PICO score after a single-session JC. When we compared the individual values of CAST, PICO, and change at the two centers, the only significant difference was the change in CAST in four-session JC, which was higher at center 2 than center 1. More learning is possible for novel stimuli, and learning about them is consequently more reinforcing than learning about other relatively familiar stimuli (39). Thus this could probably explain why, at center 2, where the JC four-session was conducted as a novel format, postgraduate students had higher augmentation in their detailed critical evaluative skills, whereas, at center 1, where single-session JC was a new format, postgraduates experienced more improvement in PICO score after a single session of JC. This observation emerged as the most consistent finding, emphasizing that a greater number of sessions would aid in better critical appraisal skills but also taking into account the previous experience.

We also compared absolute values of CAST and PICO in single-session and four-session JC at both centers; the only significant finding was that the post-PICO score in single-versus four-session JC at center 1 was higher in four-session JC, probably signifying that it might also improve in four-session JC, although the pre-post PICO change was not statistically different.

**Correlational Analysis**

Although, in both the sessions of JC, those who scored higher pre-JC CAST and PICO values also tended to score more after the JC sessions, but higher improvement was seen among those who scored less at the pre-JC for CAST value in single-session JC, and in both the sessions for PICO value, indicating that the JC seemed to benefit the relatively weaker students more.

In both the centers, for both types of JC session, those who scored higher pre-JC CAST and PICO values also tended to score higher after the JC sessions. The improvement in both CAST and PICO scores seemed to be more for those who scored less in the pre-JC test in center 1, indicating that JC sessions tended to be more beneficial for postgraduate students, who had lesser critical appraisal ability compared with others. Also, delta change in CAST and delta change in PICO were not associated with each other, suggestive of the premise that, although both are evaluating critical appraisal skills, the two mirror different aspects of the same.

The salient finding of the present study is that, first and foremost, critical appraisal skills can be improved using a JC as an educational intervention. Second, innovating with more sessions of JC discussion may be adopted to develop the critical-thinking abilities. Once the student has been adequately trained, a single session maybe used in conjunction to reduce boredom and also achieve the ability to grasp relevant information in a smaller timespan and also expose the student to a greater number of case scenarios.

**Strengths, Limitations, and Future Prospects**

The strength of this study lies in the conception of a novel idea for improvisation of the utility of JC in augmenting critical appraisal skills and assessing the effectiveness of the same using objective tests. Although the target participants were postgraduate students, the findings of the study can be extrapolated to undergraduate students as well.

This study has some inherent limitations, as the design is restricted to single- versus four-session JC format (and not tried to optimize the number of sessions) at two centers only.

**Conclusion**

Our results indicate that a four-session format of JC is more effective in augmenting critical appraisal skills of postgraduate students, whereas a single session is sufficient for improving factual knowledge. We recommend that a blended approach in terms of the number of sessions of JC discussion may be adopted. Wherein a multisession format may be more beneficial to develop critical-thinking abilities in the initial years of training and thereafter, after adequate grounding, a single session may be used in conjunction to provide a wider exposure to scientific literature.

**ACKNOWLEDGMENTS**

We acknowledge all of the faculty members and the postgraduate students for active participation and support during the conduct of the study.

**DISCLOSURES**

No conflicts of interest, financial or otherwise, are declared by the authors.

**AUTHOR CONTRIBUTIONS**

M.K., S.K., Ratna Sharma, Renuka Sharma, R.K., and K.K.D. conceived and designed research; M.K., H.B.S., and S.K. performed experiments; M.K., H.B.S., and S.K. analyzed data; M.K. and S.K. interpreted results of experiments; M.K., H.B.S., and S.K. prepared figures; M.K., H.B.S., and S.K. drafted manuscript; M.K., S.K., and K.K.D. edited and revised manuscript; M.K., S.K., Ratna Sharma, Renuka Sharma, R.K., and K.K.D. approved final version of manuscript.

**REFERENCES**

1. Akobeng AK. Principles of evidence based medicine. *Arch Dis Child* 90: 837–840, 2005. doi:10.1136/adc.2005.071761.
