


Evaluation of a Preclerkship Learning Community Model for Delivering Clinical Skills Curriculum

Danielle Roussel¹ , Katherine Anderson², Tiffany Glasgow³ and Jorie M Colbert-Getz⁴

¹Department of Anesthesiology, University of Utah School of Medicine, Salt Lake City, UT, USA.

²Department of Internal Medicine, Division of Geriatrics, University of Utah School of Medicine, Salt Lake City, UT, USA. ³Department of Pediatric, University of Utah School of Medicine, Salt Lake City, UT, USA. ⁴Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA.

Journal of Medical Education and Curricular Development
Volume 6: 1–7
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2382120519855061



ABSTRACT

BACKGROUND: Medical schools are increasingly using learning communities (LCs) for clinical skills curriculum delivery despite little research on LCs employed for this purpose. We evaluated an LC model compared with a non-LC model for preclerkship clinical skills curriculum using Kirkpatrick's hierarchy as an evaluation framework.

METHODS: The first LC cohort's (N=101; matriculating Fall 2013) reaction to the LC model was assessed with self-reported surveys. Change in skills and learning transfer to clerkships was measured with objective structured clinical examinations (OSCEs) at the end of years 2 and 3 and first and last clerkship preceptor evaluations; the LC cohort and the prior cohort (N=86; matriculating Fall 2012) that received clinical skills instruction in a non-LC format were compared with Mann-Whitney *U* tests.

RESULTS: The LC model for preclerkship clinical skills curriculum was rated as excellent or good by 96% of respondents in Semesters 1 to 3 (N=95). Across multiple performance domains, 96% to 99% of students were satisfied to very satisfied with their LC faculty preceptors (N varied by item). For the end of preclerkship OSCE, the LC cohort scored higher than the non-LC cohort in history gathering ($P=.003$, $d=0.50$), physical examination ($P=.019$, $d=0.32$), and encounter documentation ($P\leq .001$, $d=0.47$); the non-LC cohort scored higher than the LC cohort in communication ($P=.001$, $d=0.43$). For the end of year 3 OSCE, the LC cohort scored higher than the non-LC cohort in history gathering ($P=.006$, $d=0.50$) and encounter note documentation ($P=.027$, $d=0.24$); there was no difference in physical examination or communication scores between cohorts. There was no detectable difference between LC and non-LC student performance on the preceptor evaluation forms at either the beginning or end of the clerkship curriculum.

CONCLUSIONS: We observed limited performance improvements for LC compared with non-LC students on the end of the preclerkship OSCE but not on the clerkship preceptor evaluations. Additional studies of LC models for clinical skills curriculum delivery are needed to further elucidate their impact on the professional development of medical students.

KEYWORDS: Learning community, clinical skills, curriculum development, medical education

RECEIVED: May 3, 2019. **ACCEPTED:** May 14, 2019.

TYPE: Original Research

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Danielle Roussel, University of Utah School of Medicine, 30 North 1900 East, Salt Lake City, UT 84132, USA. Email: Danielle.roussel@hsc.utah.edu

Introduction

Higher education learning communities (LCs) can be broadly defined as groups of people assembled to improve learning.¹ The past 2 decades have seen dramatic growth in the number of US medical schools developing LCs.^{2,3} Though the emphases of LCs vary considerably across institutions, the broad categories of LC functions in medical schools include mentoring/advising, curriculum delivery, social, and community service.^{3,4}

By design, curricular LCs embrace a hands-on model of instruction; the underlying assumption is that instruction and feedback occurring within a committed community of learners will improve learning.⁵ Based on this assumption of improved learning, LC models for delivering preclerkship clinical skills curriculum have become increasingly common.^{3,4} Early studies examining the benefits of preclerkship clinical skills training in a LC model have found improved student comfort when starting third-year clerkships and improved proficiency with clinical skills based on clerkship evaluations.^{6,7}

In evaluating LC curricula, it is important to include comparisons to alternative or predecessor curricula to determine comparative effectiveness as opposed to student capacity to learn.⁸ In evaluating our LC model for preclerkship clinical skills curriculum, we therefore included comparison with a lecture and small group model (non-LC) with clinic encounters for delivering clinical skills instruction. We hypothesized that LC students would perform better than non-LC students at the end of the preclerkship curriculum and on their first clerkship rotation but we were unsure whether there would be any differences in performance at the end of the clerkship year.

Methods

Participants

Participants were University of Utah School of Medicine (UUSOM) medical students who matriculated in fall 2012



(N=86) and Fall 2013 (N=101). Fall 2012 entering medical students were in the last cohort before LCs were implemented for years 1 to 2 clinical skills instruction (non-LC cohort) and fall 2013 entering medical students represented the first cohort of the LC model for years 1 to 2 clinical skills instruction (LC cohort).

Curriculum

Non-LC curriculum. Prior to fall 2013, preclerkship clinical skill instruction was primarily delivered in the first 4 months of year 1. During this time, students received approximately 4 hours of clinical skills instruction weekly consisting of lecture and some skills practice in student-selected small groups with 3 to 4 faculty instructors circulating between the groups. The curriculum emphasized history taking, physical examination, communication, and documenting the clinical encounter. In addition, between month 5 of year 1 and the end of year 2, each student participated in 30 half-day clinic encounters. These encounters included outpatient and inpatient primary care and subspecialty experiences; the purpose of these clinical experiences was for students to practice and expand their clinical skills before they started the formal year 3 clerkship curriculum.

LC curriculum. Starting in fall 2013, a revised preclerkship clinical skills curriculum was implemented. LCs composed of a clinical core faculty instructor and 10 to 11 students met for four hours during most weeks in years 1 and 2. The year 1 content was similar to that for non-LC students, but instruction primarily occurred in the LCs with fewer large group lectures. The additional curricular time in the LC model allowed instruction in early clinical reasoning, hypothesis driven data acquisition, oral patient presentations, and advanced physical examination techniques. Simultaneously, the number of half-day clinical experiences was decreased to 18. The LC curriculum included more individualized instruction, structured performance feedback for students, and formalized mentorship. Weekly faculty development sessions were held to ensure consistency across LCs.

Program evaluation

To measure the effectiveness of the LC model, we used Kirkpatrick's adapted hierarchy for the health professions as a program evaluation framework.⁹⁻¹¹ Specifically, data for level 1—reaction, level 2—change in skills, and level 3—transfer of learning to the workplace were collected.

Data analysis

Level 1—reaction. Student satisfaction with the LC model was measured with self-reported data from required end-of-course surveys after each semester. LC students rated the overall quality of each semester of the LC curriculum in years 1 to 2

as excellent, good, fair, or poor. Students were surveyed at the midpoint of year 2 about their satisfaction with their LC faculty members' ability to perform various tasks on a very satisfied, satisfied, dissatisfied, very dissatisfied scale. Descriptive statistics (frequencies and percentages) were computed for reaction data.

Level 2—change in skills. Change in skills was measured with objective structured clinical examinations (OSCEs) at the end-of-year 2 (EOY2) and at end-of-year 3 (EOY3). Students were required to pass each OSCE to progress to the next phase in the curriculum.

The EOY2 OSCE included 3 stations for non-LC students and 2 stations for LC students. All non-LC students completed stations 1, 2, and 3. All LC students completed station 1 (same station 1 as non-LC students), half of the LC students completed station 2 (same station 2 as non-LC students), and the other half of the LC students completed station 3 (same station 3 as non-LC students). Four domain means were computed for each student: history gathering, physical examination, encounter note documentation, and communication. The same rubric was used to grade encounter notes for LC and non-LC cohorts; however, 2 trained nonphysicians graded the non-LC cohort encounter notes, whereas LC faculty members graded the LC cohort encounter notes. LC faculty did not grade encounter notes for their own LC students. All other domain means were based on the number of checklist items performed correctly as rated by standardized patients.

The EOY3 OSCE had 7 stations and non-LC and LC students completed the same 7 stations. Four domain means were computed for each student: history gathering, physical examination, communication, and encounter note documentation. Two trained nonphysicians used the same rubric to grade encounter notes for LC and non-LC cohorts. All other domain means were based on the number of checklist items performed correctly as rated by standardized patients.

To ensure content validity evidence for the EOY2 and EOY3 OSCEs, a blueprint was used to determine case diagnoses, and checklists and encounter note rubrics were created by a clinician educator and reviewed by a medical education expert. Standardized patients received annual training for their respective cases and nonphysician encounter note graders participated in a calibration session to ensure response process validity evidence. Internal consistency evidence was not strong given the low number of stations for both OSCEs; however, it is difficult to reach a high reliability coefficient due to case specificity.¹² Relationship to other variables and consequence validity evidence were not collected for the purpose of this study.

Performance was computed in each domain for each OSCE and compared between the non-LC LC cohorts with Mann-Whitney *U* tests. Cohen's *d* was used to measure effect sizes for any significant differences with 0.20 to 0.49 indicating a small

Table 1. Sample 19-item rating form rating scale and behavioral anchors for Data Gathering: Initial History/Interviewing Skills and Data Recording/Reporting: Written Histories and Physicals.

DATA GATHERING: INITIAL HISTORY/INTERVIEWING SKILLS									
4		3		2		1		0	
Resourceful, efficient, appreciates subtleties, insightful, obtains all relevant data including psychosocial components		Precise, detailed, broad-based, obtains almost all relevant data including psychosocial components		Obtains basic history, accurate, obtains most of the relevant data and most of the psychosocial components		Incomplete or unfocused, relevant data missing, psychosocial components absent or sketchy		Inaccurate, major omissions, inappropriate, psychosocial component entirely absent	Not observed
○	○	○	○	○	○	○	○	○	○
DATA RECORDING/REPORTING: WRITTEN HISTORIES AND PHYSICALS									
4		3		2		1		0	
Concise, reflects thorough understanding of disease process and patient situation		Documents key information, focused, comprehensive		Accurate, complete		Poor flow in History of Present Illness, lacks supporting detail or incomplete problem lists		Inaccurate data or major omissions	Not observed
○	○	○	○	○	○	○	○	○	○

effect size, 0.50 to 0.79 representing a moderate effect size, and ≥ 0.80 representing a large effect size.

Level 3—transfer of learning to the workplace. Transfer of clinical skills instruction in years 1 to 2 to the clerkship environment in year 3 was measured with a 19-item rating form completed by attending faculty and resident preceptors during each of the 7 clerkships in year 3; this instrument was used for both cohorts. During the third year, students complete clerkships in Family Medicine, Internal Medicine, Neurology, Obstetrics and Gynecology, Pediatrics, Psychiatry, and Surgery. In each clerkship, with the exception of Family Medicine, multiple faculty and resident preceptors in outpatient and inpatient settings rated each student. In Family Medicine students often worked with only 1 or 2 community-based faculty members. For each of the 19 rating items, preceptors rated students on a 5-item Likert-type scale where the lowest level of performance was scored as 0 and the highest level of performance was score as 4; preceptors could rate students between levels such that a total of 9 distinct ratings were possible for each item. Each of the Likert-type scale items for each of the questions had a distinct behavioral anchor. Seven of the rating items related to a patient care / clinical skill competency, 5 items related to a medical knowledge competency, and 7 items related to a professionalism competency. By example, the Likert-type scale and behavioral anchors for Data Gathering: Initial History/Interviewing Skills and Data Recording/Reporting: Written Histories & Physicals are displayed in Table 1.

The preceptor rating form was developed more than 10 years ago and there is not a record of a blueprint or review process so it is difficult to determine content validity evidence. The form

had a 5-point scale for each item, but usually preceptors gave 1 of 3 ratings thereby limiting the response process evidence. There was good internal consistency across items (Cronbach $\alpha > 0.80$), but assessing interrater reliability was difficult due to different preceptors rating each student. Relationship to other variables and consequences validity evidence was not collected for the purpose of this study.

Average scores were computed for the first clerkship in each competency and for the last clerkship in each competency and compared between LC and non-LC cohorts with Mann-Whitney *U* tests. Each student completed a different clerkship in rotation 1 and rotation 7 (eg, Internal Medicine Clerkship in rotation 1 and Surgery clerkship in rotation 7). Cohen's *d* was used to measure effect sizes for any significant differences. Any rating form completed by a LC faculty member was omitted from the analyses. In addition, the analyses for Level 3 were limited to students who completed 1 of the 7 clerkships in the first block and/or last block of year 3.

This study was deemed exempt by the University of Utah Institutional Review Board.

Results

Reaction data were collected from the 101 students in the LC cohort. Two of the LC cohort students did not advance to complete the EOY2 exam. Change in skills data were based on the 86 non-LC cohort students and 99 LC cohort students who completed the EOY2 OSCE and the 73 non-LC cohort students and 89 LC cohort students who completed the EOY3 OSCE. Transfer of Learning to the Workplace data were based on 64 non-LC students for clerkship rotation 1 and 82 for rotation 7 of year 3 compared with 83 LC students for rotation

Table 2. Average percentages of student ratings with frequencies in parenthesis for the first learning community faculty cohort at the University of Utah School of Medicine academic year 2014-2015.

	VERY SATISFIED	SATISFIED	DISSATISFIED	VERY DISSATISFIED	NOT ABLE TO RATE
Ability to teach history taking	78% (71)	19% (17)	0%	1% (1)	2% (2)
Ability to teach community skills	76% (69)	18% (16)	3% (3)	1% (1)	2% (2)
Ability to teach physical exam techniques	74% (67)	23% (21)	1% (1)	1% (1)	1% (1)
Ability to provide meaningful feedback	74% (67)	23% (21)	0%	2% (2)	1% (1)
Ability to act in a professional manner	78% (71)	18% (16)	2% (2)	1% (1)	1% (1)
Ability to guide clinical reasoning	80% (73)	18% (16)	0%	1% (1)	1% (1)
Ability to teach presentation skills	76% (69)	20% (18)	1% (1)	1% (1)	2% (2)
Adequacy of feedback about my documentation skills	76% (69)	21% (19)	1% (1)	1% (1)	1% (1)

The total Ns vary by task due to omitting “not able to rate” ratings or if a student did not provide any rating for a task.

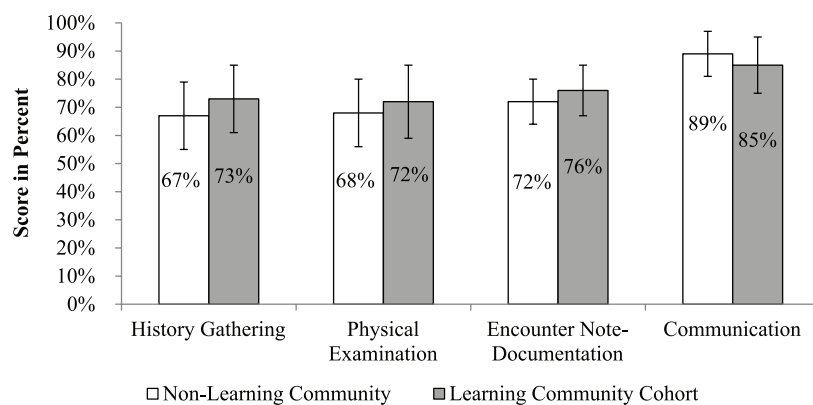


Figure 1. End of year 2 objective structured clinical examination mean performance with error bars indicating 1 standard deviation above and below the mean for the nonlearning community cohort in academic year 2013-2014 and the first learning community cohort in academic year 2014-2015 at the University of Utah School of Medicine.

1 and 83 for rotation 7 (not entirely the same 83 students). Differences within cohorts in the number of students completing clerkships in rotation 1 and rotation 7 were due to student leaves of absence after year 2 or during year 3 (eg, for another degree program, academic or personal reasons) or to completing an elective in either the first or last block. The EOY3 OSCE analysis was limited to 73 students in the non-LC cohort and 89 students in the LC cohort.

Level 1—reaction

The percent of LC students rating the LC curriculum as excellent or good by semester were as follows: Semester 1: 96% (N=97), Semester 2: 96% (N=97), Semester 3: 96% (N=97), Semester 4: 26% (N=26).

Table 2 displays student satisfaction ratings for all LC faculty. Ten students did not complete the satisfaction survey for their faculty member. The majority of students were very satisfied or satisfied with their LC faculty members' ability to

perform all tasks. One LC faculty received dissatisfied or very dissatisfied ratings by 1 to 2 students depending on the task.

Level 2—change in skills

Figure 1 illustrates EOY2 OSCE performance for non-LC and LC cohorts. The LC cohort scored significantly higher than the non-LC cohort in history gathering ($P=.003$, $d=0.50$), physical examination ($P=.019$, $d=0.32$), and the encounter note documentation ($P\leq.001$, $d=0.47$). The non-LC cohort scored significantly higher than the LC cohort in communication ($P=.001$, $d=0.43$).

Figure 2 shows EOY3 OSCE performance for non-LC and LC cohorts. The LC cohort scored significantly higher than the non-LC cohort in history gathering ($P=.006$, $d=0.50$), and encounter note documentation ($P=.027$, $d=0.24$). There were no significant differences between cohorts in physical examination ($P=.860$) or communication ($P=.121$). Internal consistency evidence was not strong given the low number of stations

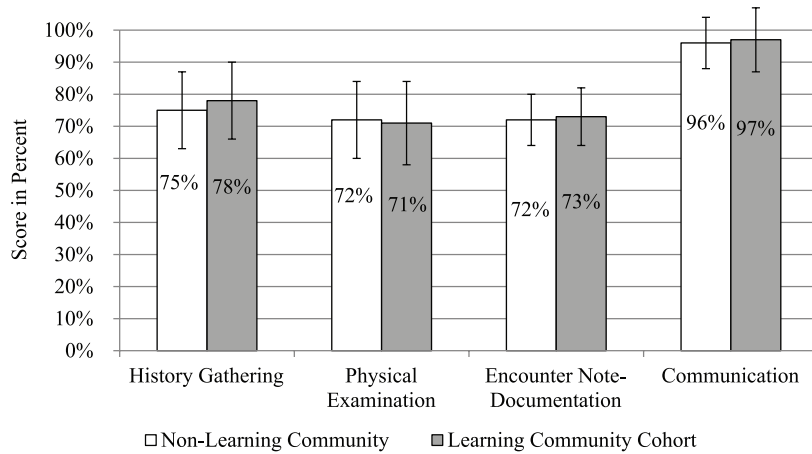


Figure 2. End of year 3 objective structured clinical examination mean performance with error bars indicating 1 standard deviation above and below the mean for the nonlearning community cohort in academic year 2013-2014 and the first learning community cohort in academic year 2014-2015 at the University of Utah School of Medicine.

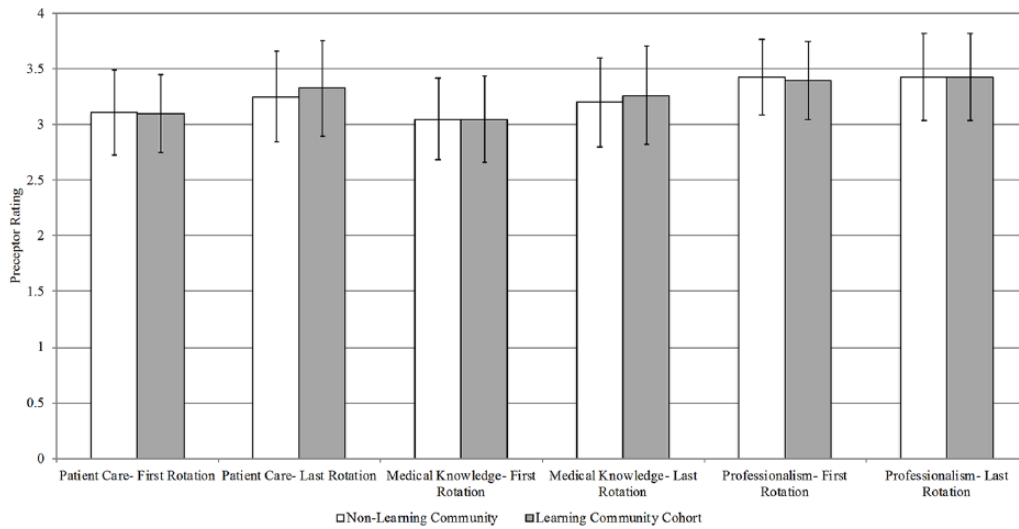


Figure 3. Average preceptor ratings in first and last clerkship with error bars indicating 1 standard deviation above and below the mean for the nonlearning community cohort in academic year 2014-2015 and the first learning community cohort in academic year 2015-2016 at the University of Utah School of Medicine.

for both OSCEs and the fact that it is difficult to reach high reliability coefficient due to case specificity.¹²

Level 3—transfer of learning to the workplace

Figure 3 illustrates average clerkship preceptor ratings by competency for non-LC and LC cohorts. There were no significant differences between cohorts for first clerkship ratings in patient care ($P = .528$), medical knowledge ($P = .793$) or professionalism ($P = .458$). There were also no significant differences between cohorts for last clerkship ratings in patient care ($P = .273$), medical knowledge ($P = .328$) or professionalism ($P = .831$).

Discussion

While the majority of medical education studies only reach level 2 of the Kirkpatrick hierarchy, our evaluation is

strengthened by measuring the impact of the LC model with levels 1 to 3 of the hierarchy and may therefore provide a more complete picture of postintervention LC effects.¹³ The findings of our study, in which favorable reaction level data (ie, end of course student surveys) was accompanied by only limited change in skill (ie, EOY2 OSCE) and no measured differences in workplace performance (ie, first clerkship preceptor ratings), highlight the risks of major curricular decisions made on only lower hierarchy reaction level data when higher level program evaluation data may not similarly support the benefits of the program. For LCs in particular, the implementation costs especially related to securing clinical faculty mentors can be significant, and higher level outcome data may be needed to justify program continuation.

Our study is also unique in that we gathered evaluation data at both early (ie, end of preclerkship, first clerkship) and late timepoints after intervention (ie, last clerkship, end of clerkship

curriculum) to evaluate both immediate and potentially sustained benefits of our educational intervention; most educational intervention studies have short intervals between intervention and last assessment.¹³ Our detection of only limited late performance differences in skills (ie, EOY3 OSCE) and not difference in workplace transfer (ie, last clerkship preceptor ratings) illustrates the importance of measuring both immediate and sustained benefits for educational interventions. Though limited difference in performance at the end of clerkships likely primarily reflects the limited differences in measured performance at the end of the preclerkship curriculum, it may also reflect a dominant effect of the immersive clerkship experience on clinical skills development that overshadows benefits of a particular preclerkship clinical skills curriculum delivery method like LCs.

Also importantly, as opposed to measuring learning outcomes of a new curriculum to no curriculum, our evaluation is strengthened by comparing clinical skills content delivered in a LC model compared with clinical skills content delivered in a non-LC model; we were able to perform these comparisons between the 2 curricula for both change in skills and transfer of learning to the workplace.⁸ Again, the lack of major differences between cohorts in measures of clinical skills reinforces the importance of comparing existing curriculum with new curriculum as opposed to new curriculum compared with no curriculum. Our results are consistent with literature suggesting the challenges of achieving incremental performance enhancements between different pedagogical approaches.⁸ Our results also highlight the limitations of traditional experimental approaches for large-scale curricular interventions and may support suggestions that qualitative methods and theory-guided experimentation should complement comparative effectiveness research methods.^{8,14}

In relating our findings to existing literature on LC models for preclerkship clinical skills curriculum delivery, it is notable that except for low fourth-semester satisfaction, the positive student reaction to the first 3 semesters of the curriculum is consistent with prior research reporting more favorable student ratings of a LC clinical skills curriculum compared with a non-LC curriculum.¹⁵ In contrast, our finding on skills transfer to the workplace is to some degree inconsistent with prior research by Jackson et al who observed higher ratings on Internal Medicine clerkship rating forms; our finding is however consistent with that same study's finding of very few significant differences in individual item ratings in other clerkships.⁷ Some degree of inconsistency in findings for our 2 institutions may be related to factors including differences in the LC-based curricula between the institutions, analysis of clinical clerkships jointly in our study as opposed to analyzing differences at the individual clerkship level in the Jackson study, and/or differences in the rating forms used to gather preceptor perceptions about student performance. In addition, Jackson et al⁷ assessed 3 cohorts of pre- and post-LC curricular implementation and excluded the first cohort of the new LC curriculum due to substantial changes that occurred

after the first iteration; we analyzed only our first LC cohort compared with our last non-LC cohort.

A number of limitations in our study are worth noting. First, and highlighting limitations of reaction data, the profoundly lower student satisfaction ratings for the fourth semester may have been overly influenced by student reaction to professionalism issues that course directors needed to address with the entire class as well as general student resistance to attempted integration of an evidence-based medicine course into the LC curriculum. In addition, our analysis of student reactions to their preclerkship clinical skills training did not include direct comparison of LC and non-LC cohort; this was due to the lack of a comparable reaction question being asked of non-LC students. Though the AAMC Graduate Questionnaire includes a question on student satisfaction with preclerkship clinical skills training, we were unable to use these data due to some mixing of LC and non-LC students that resulted from off-track graduation caused by pursuit of additional degrees, academic performance, or personal leaves of absence. Our study is also limited generally by untested measurements/limited validity evidence for the OSCEs used to measure student skills acquisition and the preceptor rating form used to measure transfer of skills to the workplace. Specifically regarding the clerkship preceptor rating form, in an attempt to remove bias from this study we excluded clerkship preceptor rating forms completed by LC core faculty members. However, including these faculty ratings in our analysis may have changed the clerkship performance rating results as these faculty may have more thoroughly assessed student capabilities. In addition, inability to detect first or last clerkship performance rating differences between cohorts may have been related to preceptors most commonly rating students within a narrow band of 3 ratings. As mentioned related to the existing LC literature, our study is limited by comparison of only 1 cohort of students in a LC model to 1 cohort of students in a non-LC model at a single institution.

In sum, our experimental approach of comparing one curriculum with another using Kirkpatrick hierarchy level 1 to 3 measures at both near and distant timepoints did not demonstrate substantially altered trajectory of student clinical skills acquisition in a preclerkship LC model for clinical skills curriculum delivery. These results are important to other institutions considering the time, effort, and cost required for implementation and evaluation of LC models for this purpose. However, the limitations of our study make us cautious about rejecting LCs for this purpose; it is possible that preclerkship LCs support student development in important but unmeasured areas such as student perceived self-efficacy or professional identity formation.¹⁶ And, for institutions with limited capability to provide authentic in situ preclerkship clinic experiences, our findings of limited differences between non-LC and LC models may actually suggest that LCs may be a feasible option to supplant some of the preclerkship clinic

experiences. It will be important as LCs continue to experience increasing presence in medical schools to clarify their strengths and limitations so that they can be optimally deployed.

Conclusions

Implementation of a LC model for preclerkship clinical skills curriculum was associated with high student satisfaction, limited improvements in performance on the EOY2 and EOY3 OSCEs compared with a non-LC cohort, and no differences in performance on first or last clerkship preceptor ratings compared with non-LC students. As the number of medical schools with clinical skills curricular LCs is increasing, further studies evaluating the short- and long-term impact of this educational modality via both qualitative and quantitative methods should be undertaken to continue to elucidate the impact of LCs on the professional development of medical students.

Author Contributions

DR, KA, TG, JMCG contributed to the design and implementation of the evaluation, analysis of data, and writing of the manuscript.

ORCID iD

Danielle Roussel  <https://orcid.org/0000-0002-3993-6987>

REFERENCES

1. Lenning O, Ebbers L. *The Powerful Potential of Learning Communities: Improving Education for the Future* (ASHE-ERIC Higher Education Report. 1999;26(6)). Washington, DC: Graduate School of Education and Human Development, The George Washington University.
2. Ferguson KJ, Wolter EM, Yarbrough DB, Carline JD, Krupat E. Defining and describing medical learning communities: results of a national survey. *Acad Med.* 2009;84:1549–1556. doi:10.1097/ACM.0b013e3181bf5183.
3. Smith S, Shochet R, Keeley M, Fleming A, Moynahan K. The growth of learning communities in undergraduate medical education. *Acad Med.* 2014;89:928–933. doi:10.1097/ACM.0000000000000239.
4. Osterberg LG, Goldstein E, Hatem DS, Moynahan K, Shochet R. Back to the future: what learning communities offer to medical education. *J Med Educ Curric Dev.* 2016;3:67–70.
5. Goldstein EA, MacLaren CF, Smith S, et al. Promoting fundamental clinical skills: a competency-based college approach at the University of Washington. *Acad Med.* 2005;80:423–433.
6. Whipple ME, Barlow CB, Smith S, Goldstein EA. Early introduction of clinical skills improves medical student comfort at the start of third-year clerkships. *Acad Med.* 2006;81:S40–S43.
7. Jackson MB, Keen M, Wenrich MD, Schaad DC, Robins L, Goldstein EA. Impact of a pre-clinical clinical skills curriculum on student performance in third-year clerkships. *J Gen Intern Med.* 2009;24:929–933. doi:10.1007/s11606-009-1032-7.
8. Cook DA. If you teach them, they will learn: why medical education needs comparative effectiveness research. *Adv Health Sci Educ Theory Pract.* 2012;17:305–310. doi:10.1007/s10459-012-9381-0.
9. Kirkpatrick DL, Kirkpatrick KJ. *Evaluating Training Programs: The Four Levels*. 3rd ed. San Francisco, CA: Berrett-Koehler; 2006.
10. Praslova L. Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in higher education. *Educ Asse Eval Acc.* 2010;22:215–225.
11. Dudas RA, Colbert-Getz JM, Balighian E, et al. Evaluation of a simulation-based pediatric clinical skills curriculum for medical students. *Simul Healthc.* 2014;9:21–32. doi:10.1097/SIH.0b013e3182a89154.
12. Regehr G, Freeman R, Hodges B, Russell L. Assessing the generalizability of OSCE measures across content domains. *Acad Med.* 1999;74:1320–1322.
13. Baernstein A, Liss HK, Carney PA, Elmore JG. Trends in study methods used in undergraduate medical education research, 1969–2007. *JAMA.* 2007;298:1038–1045.
14. Norman G. RCT = results confounded and trivial: the perils of grand educational experiments. *Med Educ.* 2003;37:582–584.
15. Smith S, Dunham L, Dekhtyry M, et al. Medical student perceptions of the learning environment: learning communities are associated with a more positive learning environment in a multi-institutional medical school study. *Acad Med.* 2016;91:1263–1269. doi:10.1097/ACM.0000000000001214.
16. Cruess RL, Cruess SR, Boudreau JD, Snell L, Steinert Y. A schematic representation of the professional identity formation and socialization of medical students and residents: a guide for medical educators. *Acad Med.* 2015;90:718–725. doi:10.1097/ACM.0000000000000700.