

# Multiple Medical Students on a Clerkship Team: Do They Affect Each Others' Grades?

Krishan K. Sharma

Massachusetts General Hospital

Yuchiao Chang

Massachusetts General Hospital

Eli Michael Miloslavsky (✉ [emiloslavsky@mgh.harvard.edu](mailto:emiloslavsky@mgh.harvard.edu))

Massachusetts General Hospital <https://orcid.org/0000-0003-0008-8105>

---

## Research article

**Keywords:** Medical students, clerkship grades, team, medical school

**Posted Date:** July 22nd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-44283/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background** Medical students are often paired together on clinical teams during their clerkships, but the effect of this practice on student performance is unknown. The primary objectives of this study were (1) to retrospectively assess whether students paired together on a medical team during their Internal Medicine sub-internship affected each other's grade and (2) to survey medical students' perceptions on the impact of pairing on their evaluations.

**Methods** We examined clerkship grades of 186 student-pairs at 3 sub-internship hospital sites of Harvard Medical School from 2013-2017. To evaluate student perceptions we administered a survey to the graduating class of 2018.

**Results** There was no significant deviation between the expected and observed distribution of student grades ( $p=0.39$ ) among 186 student pairs, suggesting that pairing had no meaningful effect on the sub-internship grade. We also saw no effect when controlling for prior internal medicine clerkship performance ( $p=0.53$ ). We then surveyed students in the 2018 graduating class assessing student perceptions on pairing. Of the 99 respondents (59% response rate), 90% and 87% of respondents felt that being paired affected their evaluations by resident and attending physicians, respectively.

**Conclusions** Our analysis suggests that paired medical students do not meaningfully affect each others' grades, despite the majority of surveyed students believing that being paired affects their evaluations. Awareness of student perceptions regarding pairing can inform clerkship structure and be utilized to address student concerns.

## Background

The core clerkship year of medical school, consisting of required clinical rotations in a subset of specialties, is widely regarded as the most educational and transformative year in undergraduate medical education.<sup>1</sup> Outside of the standardized elements of a taught curriculum, clerkship education is situated within the learning environment, taking place within the relationships between students, patients, and physicians, supported by informal elements of a learned curriculum.<sup>2</sup> At some medical schools, medical students are often paired together during their clerkships, where two or more students will work with the same medical team and be evaluated by the same group of resident and attending physicians. Despite extensive research on the preceptor-student relationship, there is limited data describing the effect of student pairing on their experience in clinical clerkships.<sup>3,4</sup>

Pairing students on the same clinical team can have potential benefits, as peer relationships support the moral development of medical students, peer-led teaching improves knowledge acquisition in both clinical and pre-clinical settings, and peer observation from students improves clinical skills.<sup>5,6,7</sup> Furthermore, group-based learning modalities, such as team-based learning and case-based collaborative learning, utilized for pre-clinical curricula, have demonstrated that grouping medical students of different

performance levels can often benefit low-performing students without hindering high-performing students.<sup>8,9</sup> However, pairing can also serve as an element of increased stress and anxiety for students, particularly as it relates to clerkship evaluation.<sup>10</sup> Medical student evaluations are subject to a wide degree of heterogeneity, and pairing may allow for further cognitive bias and implicit comparison between students if evaluated by the same medical team.<sup>11,12</sup> To our knowledge, whether paired students affect each others' grades and student perceptions on the impact of this practice have not been previously investigated.

We sought to examine whether paired students on a medical sub-internship team had a measurable effect on each other's clerkship grade as well as the students' perceptions of how pairing affects their evaluations. In alignment with the available literature for preclinical teaching modalities, we hypothesized that a given student would perform better than expected from probability alone when paired with a high-performing student.<sup>13</sup> We also suspected that majority of students would perceive pairing as affecting their evaluations.

## Methods

### Analysis of the effect of pairing on clerkship grade

We performed a retrospective study examining students' grades for the medicine sub-internship at Harvard Medical School (HMS) from 2013–2017. The medicine sub-internship at HMS is a four-week clinical experience where students in their final year of medical school are assigned to a general medical ward and function in a capacity similar to that of a first-year intern. At all 3 sites, medical students are paired with one or two other students. This pairing generally replaces a single intern. "Pairing" is defined as students working with the same medical team of resident and attending physicians, in addition to participating in didactics together. Paired students would separately assume responsibility for their own patients, with the exception of one site (Massachusetts General Hospital [MGH]), where medical teams use a team-based care model, and responsibility for patient care is shared among all members of the team. Students in a pair are evaluated by the same team of resident and attending physicians. Our analysis utilized student grades, as they represented the summation of all evaluations, and were thus a proxy for student clinical performance. Students are randomly assigned the same partner student for the duration of the medicine sub-internship, making the effect of student pairing suitable for investigation. By contrast, students rotating on other core clerkships at HMS are inconsistently paired and rotate on multiple teams, making the effect of pairing difficult to study. The sub-internship grade is a global rating based on the students' clinical performance which is comprised of the following 10 items: history taking, physical examination, fund of knowledge, patient management, clinical evaluation and management skills, interpersonal skills, presentation skills, professionalism, cultural/social/systems awareness and initiative and desire to learn. There is no written examination. The internal medicine clerkship grade (taken by students in their 3rd year) is also based on a global rating comprised of the same 10 items.

However, in order to achieve the highest grade, the student had to meet a minimum threshold on the National Board of Medical Examiners (NBME shelf) examination during the time period examined.

We analyzed retrospective data of fourth-year medical student pairings for the medicine sub-internship from 2013–2017 at 3 HMS sub-internship sites: Beth Israel Deaconess Medical Center (BIDMC), Brigham and Women’s Hospital (BWH), and MGH. Our null hypothesis was that student pairing, regardless of their prior internal medicine clerkship performance, has no effect on their medicine sub-internship grades. Therefore, we excluded pairs containing visiting students from medical schools outside of HMS, as well as MD-PhD students who completed their internal medicine clerkship prior to entering their PhD (N = 46 pairs). Among the 46 instances of three medical students paired together on the same medical team, we only included the student with the highest sub-internship grade and student with the lowest grade among the triad for our primary analysis.

## **Statistical analysis**

During the period of the study HMS awarded the following final grades to clerkship students during third and fourth year: Honors with Distinction (HD), Honors (H), Pass (P), and Fail (F). The grading schema was changed for the graduating class of 2019. Because less than 5% of students receive a grade of either P or F, we dichotomized student grades into “High” and “Low” categories, with High defined as a grade of HD, and Low defined as a grade of either H, P, or F. To test our hypothesis, we compared the observed distribution of High-High, High-Low, and Low-Low student pairs to the expected distribution based on chance alone.

To further address our research objective of determining how high or low-performing students affect each others’ performance when paired together, we conducted a secondary analysis of sub-internship grades that accounted for student performance from their third year internal medicine clerkship. Compared to other core clerkships at HMS, the internal medicine clerkship is most similar in experience and content to the medicine sub-internship, and thus serves as an appropriate method of defining students as either “high” or “low” performers entering into their sub-internship. We defined “high-performing” students as those who received a grade of HD on their internal medicine clerkship, and “low-performing” students as those who received a grade of H, P, or F. To test our hypothesis, we compared the observed distribution of High-High, High-Low, Low-High and Low-Low student pairs to the expected distribution based on chance alone. In contrast to the previous analysis where High and Low students were analyzed in a single distribution category, the order of pairing in this analysis is relevant because the probability of receiving a sub-internship High grade is different between a student with a Low internal medicine clerkship grade and a student with a High internal medicine clerkship grade. It is also important to note that the grade distribution for the medicine sub-internship is different than the medicine clerkship with more High grades awarded in the sub-internship.

Our main analyses were conducted using chi-square goodness-of-fit tests. The chi-square goodness of fit test differs from the chi-square test in that it is used specifically to analyze how the observed value differs from the expected value. We first evaluated the correlation between students’ sub-internship grade

within their respective pairs. We compared the observed numbers to the expected numbers when there was no correlation. We examined the randomness of sorting of high and low medicine grade students into sub-internship pairs, comparing the observed numbers to the expected numbers when pairing was random. To investigate whether the effect of pairing students was dependent on prior performance, we compared the observed and expected distribution of sub-internship grade pairs stratified by students' performance on their internal medicine clerkship. We calculated the expected number of sub-internship grading pairs conditional on their internal medicine clerkship grade, and compared these expected values of pairings with our observed data. Additionally, we compared the probability of getting a High sub-internship grade between those paired with a High third-year internal medicine clerkship grade and those paired with a Low third-year internal medicine clerkship grade, stratified by their third-year internal medicine clerkship grade using Pearson chi-square tests.

We conducted two subgroup analyses: (1) stratifying analyses by different types of pairing (pairs vs. triads) and (2) stratifying by site (team-based vs. non team-based care model). In addition, we conducted a sensitivity analysis removing students with Pass or Fail internal medicine clerkship or sub-internship grade from the analyses. We used SAS version 9.4 (Carey, NC) for all analyses, with statistical significance defined as a two-sided  $P < 0.05$ .

## **Survey of student perceptions of pairing**

We developed a questionnaire in order to assess students' perceptions of how pairing affects their evaluations. Utilizing a four-choice Likert scale, we developed a three-item instrument that assessed if students perceived pairing on their clerkships as affecting how they were evaluated by both resident and attending physicians, along with describing the nature of this effect (positively or negatively influencing evaluation). The survey was revised based on expert input and cognitive testing with four students.

The anonymous questionnaire (Appendix I) was distributed to the HMS graduating class of 2018 via Qualtrics (Provo, UT). Due to the potential of differing grading schema, which was introduced for the graduating class of 2019, influencing student perceptions about the effect of pairing on their evaluations, we excluded subsequent classes from the study. Descriptive statistics were utilized for analysis using Microsoft Excel 2007 (Microsoft Corp., Redmond, Washington).

The study protocol was approved by the institutional review board of HMS.

## **Results**

We analyzed sub-internship grades from 372 students comprising 186 sub-intern pairs. During the sub-internship, 234 students received a High grade (63%), while during the medicine clerkship 152 students (41%) received a High grade. We then conducted two main analyses. First, we examined whether grades were different when students were paired compared to if student grades were completely independent of each other. Second, we analyzed whether taking into account their prior performance in a third year

Internal Medicine clerkship affected the results, testing the hypothesis that when low performing students are paired with a high performing student their grade may improve.

To evaluate the correlation between students' sub-internship grade within their respective pairs, we found that the distribution of students' sub-internship grade pairs was similar to the expected number based on random pairing ( $p = 0.39$ , Table 1), suggesting that students' grades during the sub-internship are not affected by their partners' grades when third year clerkship grades are not taken into account.

Table 1  
Paired student grades for the medicine sub-internship

Pairing	Observed	Expected*	P value
Both High	78	73.6	
One High, One Low	78	86.8	
Both Low	30	25.6	
Total	186		0.39**
* Expected numbers were calculated based on $\text{Pr}(\text{High sub-internship grade}) = 0.63$			
** Chi-square goodness-of-fit test with 2 degrees of freedom			

To investigate whether students' prior performance affected student grades when paired, we utilized third-year internal medicine clerkship grades as a proxy for prior performance. We first examined whether high and low performing students were randomly paired together on sub-internship teams and found that the pairing was indeed random as would be expected ( $p = .65$ ). We then compared the observed and expected distribution of sub-internship grade pairs stratified by students' performance on their internal medicine clerkship. We examined High internal medicine clerkship grade paired with another High grade student (N = 28 pairs), High internal medicine clerkship grade paired with a Low grade student (N = 45 pairs), Low internal medicine clerkship grade paired with a High grade student (N = 51 pairs), and Low internal medicine clerkship grade paired with another Low grade student (N = 62 pairs). Of note, the order of pairing in this analysis is important to determine its effect because a high performing student's effect on a low performing student and vice versa differ. We found no significant deviation from our observed pairings to the expected values based on conditional probability ( $p = 0.53$ , Table 2). To further corroborate this result, we found that among those with high internal medicine clerkship grade, 77% had a high sub-internship grade when they were paired with another high grade student compared to 83% when students were paired with a low grade student ( $p = 0.39$ , Table 3). Among those with low internal medicine clerkship grade, 45% had a high sub-internship grade when they were paired with a high grade student compared to 55% when students were paired with another low grade student ( $p = 0.17$ ), further suggesting no meaningful effect of pairing.

Table 2

Contingency table comparing the expected distribution of paired grades to observed data, stratified by third year internal medicine clerkship performance

Observed no. (Expected no.)*		Sub-internship grade pairs				Total	P value
		High-High	High-Low	Low-High	Low-Low		
<b>Internal medicine clerkship grade pairs</b>	High-High	17 (18.3)	5 (4.3)	4 (4.3)	2 (1.0)	28	0.76**
	High-Low	19 (18.4)	19 (18.0)	1 (4.3)	6 (4.3)	45	0.34**
	Low-High	20 (20.8)	3 (4.9)	22 (20.4)	6.0 (4.8)	51	0.76**
	Low-Low	22 (15.8)	9 (15.5)	15 (15.5)	16 (15.2)	62	0.16**
	Total	78	36	42	30	186	0.53§
* Expected numbers were calculated based on $\Pr(\text{High sub-internship grade} \mid \text{High internal medicine grade}) = 0.81$ and $\Pr(\text{High sub-internship grade} \mid \text{Low internal medicine grade}) = 0.50$ . For example, the probability of getting a High-High sub-internship grade pair from a High-Low internal medicine grade pair is $0.81 \times 0.51 = 0.41$ and expected number of pairs is $45 \times 0.41 = 18.4$ .							
** Chi-square test with 3 degrees of freedom							
§ Chi-square test with 12 degrees of freedom							

Table 3

Sub-internship High grades by Internal Medicine clerkship performance

Internal Medicine Clerkship Grade of Student 1	Internal Medicine Clerkship Grade of Student 2	N	Sub-internship High grade
High	High	56	43 (77%)
	Low	96	80 (83%)
Low	High	96	43 (45%)
	Low	124	68 (55%)
Total		372	234

In the sensitivity analysis stratified by different types of pairing, we found similar results between students paired in two and students paired in three. When stratified by site, there were only 39 pairs of

students from the site with the team-care based model and no significant differences were found. We observed similar results when the students with Pass or Fail grades were removed from analysis.

In order to assess students' perceptions regarding the effect of pairing on their evaluations, we surveyed the 2018 graduating class of HMS. Ninety nine out of 168 students responded (58.9%). Of those who responded, 10.1% felt that being paired did not affect their evaluations by residents, while 40.4%, 37.4%, and 12.1% felt that being paired affected their evaluations slightly, moderately, and strongly by resident physicians, respectively. Similarly, 13.1% of students felt that being paired did not affect their evaluations by attending physicians, while 34.3%, 35.4%, and 17.1% felt that being paired affected their evaluations slightly, moderately, and strongly by attending physicians, respectively (Fig. 1). When asked to describe the nature of this effect on their evaluations, 10.1% of students felt that being paired had a mostly positive effect on their evaluations, while 8.8% felt that the effect was mostly negative. 77.8% of students felt that this effect varied, dependent on whom the student was paired with.

## Discussion

We examined whether students paired on the same sub-internship team affected each others' grade and student perceptions on the impact pairing had on their evaluations. We demonstrate that medical student perception is not consistent with the grading patterns observed in our study. Our analysis suggests that there is no significant effect of a given student's clinical performance on their partner's performance during the sub-internship. However, the vast majority of students perceive that the student they are paired with has an effect on their clerkship evaluation. Despite student pairing being common on clinical clerkships, this is the first study to our knowledge assessing the impact of this practice.

Our results indicate that pairing does not have an immediate impact on performance, however students' perceptions towards their evaluations suggest that the practice of pairing may potentially impact the learning environment for a significant subset of students. The educational environment of clerkships encompasses the physical, social, and psychological contexts in which students learn, including interactions with faculty and peers, along with informal and hidden curricula.<sup>14</sup> School-related stressors, including academic pressure and grading, have been shown to negatively impact the learning environment and students' well being.<sup>15</sup> Educators must continue to work towards optimizing the educational setting of clerkships, as the learning environment influences students' professional development and identity.<sup>2</sup> Unsupportive learning environments, including non-collaborative and competitive settings, have been independently associated with student burnout and distress, and can potentially alter critical components of student motivation, including their sense of safety, belonging, and self-esteem.<sup>16,17</sup> Increased concern over evaluations and grading have also been associated with student burnout, and may potentially impair student learning by increasing extraneous cognitive load.<sup>18,19</sup> Therefore, efforts to address the perceptions of pairing may have a positive impact on the clerkship learning environment.

Our findings differ from prior investigation of classroom-based group learning during medical school, where grouping students of different performance levels has demonstrated a benefit to low-performing students without hindering high-performing students.<sup>8,9</sup> One possible explanation is that there may be limited peer interaction and peer-assisted learning within the clinical environment, when compared to formalized, classroom-based group learning. Because peer-assisted learning can benefit students on clinical clerkships, consideration can be given to interventions that optimize peer teaching in clerkships where students are paired and limit competition amongst students as it relates to patient exposure and evaluations.<sup>6</sup>

Our results also inform the issue of evaluator cognitive bias. Substantial variability exists among clinical evaluators with respect to the reliability, accuracy, and validity of assessments made when directly observing trainees.<sup>20,21,22</sup> Moreover, the halo effect and multiple other biases have been described affecting how a given individual may be evaluated.<sup>23,24</sup> If two students in a given pair are of different performance levels, students may be reluctant to be evaluated by the same medical team due to fear of comparison.<sup>10</sup> While our analysis puts into question the role of the halo effect and other comparator biases on evaluations of paired students, we cannot rule out the possibility that a comparator bias exists and is balanced by other factors, such as peer-teaching and peer support within a student pair. We suggest that evaluators who have had limited experience with medical students, including residents and junior attending physicians, receive training to be made better aware of their own implicit cognitive biases, including contrast and grouping biases.

We acknowledge several limitations to our study. First, this study represents the findings of a single medical school and was retrospective in nature, with no true control group of students who were unpaired. In addition, the findings of this study are dependent on the interplay between two students on a medical sub-internship, and thus may not be fully generalizable to other clerkships or medical schools, where the interactions of paired students may differ. Unfortunately, we were not able to examine the effect of pairing in other clerkships due to each clerkship having multiple clinical experiences within it where students were unpaired or had multiple pairings. Sixty three percent of the students received the top grade on the sub-internship, thus potentially limiting the sensitivity of our study. However, in the majority of U.S. medical schools the sub-internship grades skew towards a high percentage of students getting the top grade as compared to core clerkships and the HMS sub-internship has a relatively lower percentage of students receiving the top grade than other U.S. medical schools. Furthermore, we cannot exclude the possibility that a small effect size of pairing on grades exists. However, our study included over 370 students and we feel the study was of sufficient size to exclude a meaningful effect. Additional data outside of our study period could not be included due to data availability and curriculum changes. The survey evaluating student perception of pairing referred to general perception of clerkship experiences and was not limited to the sub-internship, leaving open the possibility that perceptions of pairing differ between the medicine sub-internship and other clerkships. The survey results are also subject to recall, selection and acquiescence bias. It is notable that students' perception of the pairing effect (positive versus negative) was highly dependent on the other student. This suggests that there are

multiple factors, such as student characteristics as well as student and team dynamics that may impact the learning environment within a student pair. A qualitative study is currently underway to address this question. The strengths of our study include multiple years of grading data from three different hospitals in a clerkship with consistent student pairing throughout its duration. We believe this work serves as a critical first step towards evaluating the impact of pairing students together on clerkships. Additional studies are needed to investigate the effect of pairing students together on different clerkships, and to further explore students' perceptions of the evaluative process with regards to pairing.

## **Conclusions**

Our analysis suggests medical students paired on teams during their medicine sub-internship do not meaningfully affect each other's grade, a finding that may extend to other clerkships. This observation held true even when pairing students of differing academic performance. Nonetheless, the vast majority of students surveyed believed that being paired affected their evaluations by both resident and attending physicians. We believe this gap offers insight into the current learning environment of clinical clerkships, warranting further study. Awareness of student preferences regarding pairing can inform clerkship structure and be utilized to address student concerns.

## **Abbreviations**

BIDMC - Beth Israel Deaconess Medical Center

BWH - Brigham and Women's Hospital

HMS - Harvard Medical School

MGH - Massachusetts General Hospital

## **Declarations**

### **Ethics approval and consent to participate:**

The study protocol was approved by the institutional review board of Harvard Medical School. Consent for participation in the survey part of the study was implied when participants completed the questionnaire.

### **Consent for publication:**

Not applicable

### **Availability of data and materials:**

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## Competing interests:

The authors declare that they have no competing interests

## Funding:

This study was supported by a grant from the Massachusetts General Hospital Center for Education Innovation and Scholarship which supported funding for gift cards and publication fees.

## Authors' contributions:

KKS and EMM developed the study protocol, collected and analyzed the data, interpreted the results and wrote and approved the final manuscript. YC analyzed and interpreted the results and wrote and approved the final manuscript. All authors have read and approved the manuscript

## Acknowledgements:

The authors would like to acknowledge Alberto Puig, Meredith Atkins, Erik Alexander, Alex Carbo, Benjamin Davis, Mary Montgomery, Subani Chandra, Holly Gooding, David Hirsh, and Grace Huang for their expert input.

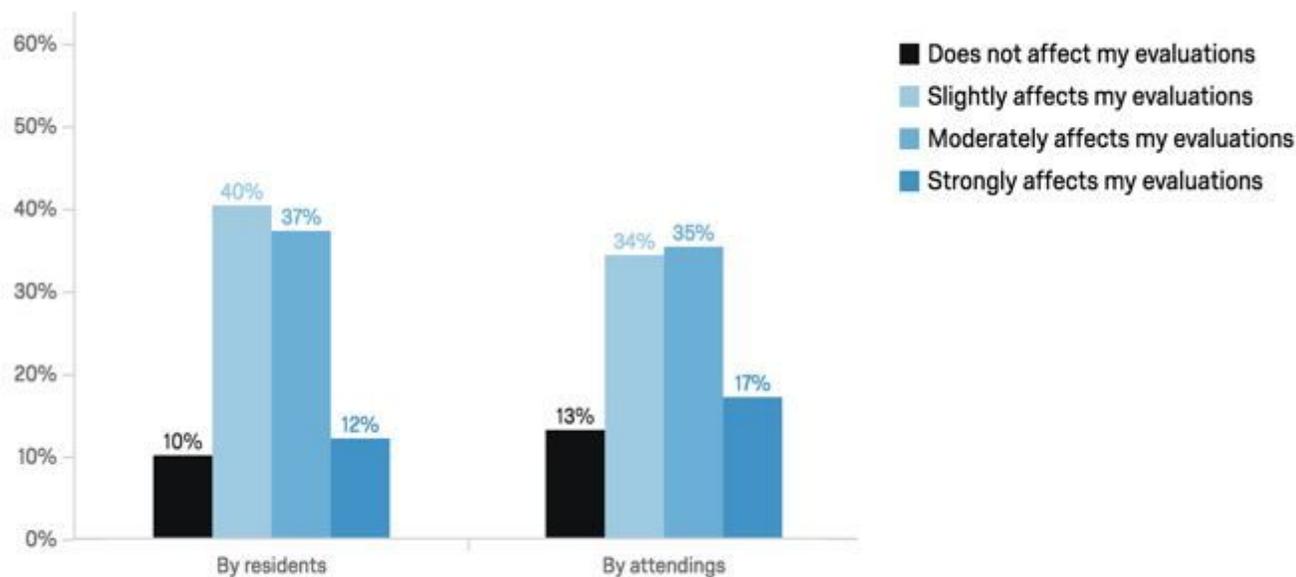
## References

1. Barrett J, Trumble SC, McColl G. Novice students navigating the clinical environment in an early medical clerkship. *Med Educ.* 2017;51(10):1014-1024. doi:10.1111/medu.13357
2. Dornan T, Tan N, Boshuizen H, et al. How and what do medical students learn in clerkships? Experience based learning (ExBL). *Adv Heal Sci Educ.* 2014;19(5):721-749. doi:10.1007/s10459-014-9501-0
3. Hauer KE, O'Brien BC, Hansen LA, et al. More is better: Students describe successful and unsuccessful experiences with teachers differently in brief and longitudinal relationships. *Acad Med.* 2012;87(10):1389-1396. doi:10.1097/ACM.0b013e31826743c3
4. van der Zwet J, Zwietering PJ, Teunissen PW, van der Vleuten CPM, Scherpbier AJJA. Workplace learning from a socio-cultural perspective: Creating developmental space during the general practice clerkship. *Adv Heal Sci Educ.* 2011;16(3):359-373. doi:10.1007/s10459-010-9268-x

5. Branch WT. Supporting the moral development of medical students. *J Gen Intern Med*. 2000;15(7):503-508. doi:10.1046/j.1525-1497.2000.06298.x
6. Burgess A, McGregor D, Mellis C. Medical students as peer tutors: A systematic review. *BMC Med Educ*. 2014;14(1). doi:10.1186/1472-6920-14-115
7. St-Onge C, Martineau B, Harvey A, Bergeron L, Mamede S, Rikers R. From See One Do One, to See a Good One Do a Better One: Learning Physical Examination Skills Through Peer Observation. *Teach Learn Med*. 2013;25(3):195-200. doi:10.1080/10401334.2013.797342
8. Krupat E, Richards JB, Sullivan AM, Fleenor TJ, Schwartzstein RM. Assessing the effectiveness of case-based collaborative learning via randomized controlled trial. In: *Academic Medicine*. Vol 91. Lippincott Williams and Wilkins; 2016:723-729. doi:10.1097/ACM.0000000000001004
9. Koles PG, Stolfi A, Borges NJ, Nelson S, Parmelee DX. The impact of team-based learning on medical students' Academic Performance. *Acad Med*. 2010;85(11):1739-1745. doi:10.1097/ACM.0b013e3181f52bed
10. Asprey D. In sickness and in hangover: pitfalls in choosing a clinical partner. *Student BMJ*. 2010;(18).
11. Alexander EK, Osman NY, Walling JL, Mitchell VG. Variation and imprecision of clerkship grading in U.S. medical schools. *Acad Med*. 2012;87(8):1070-1076. doi:10.1097/ACM.0b013e31825d0a2a
12. Williams RG, Klamen DA, McGaghie WC. Cognitive, Social and Environmental Sources of Bias in Clinical Performance Ratings. *Teach Learn Med*. 2003;15(4):270-292. doi:10.1207/S15328015TLM1504\_11
13. Kalaian SA, Kasim RM. Effectiveness of various innovative learning methods in health science classrooms: a meta-analysis. *Adv Heal Sci Educ*. 2017;22(5):1151-1167. doi:10.1007/s10459-017-9753-6
14. Karani R. Enhancing the Medical School Learning Environment: A Complex Challenge. *J Gen Intern Med*. 2015;30(9):1235-1236. doi:10.1007/s11606-015-3422-3
15. Reed DA, Shanafelt TD, Satele DW, et al. Relationship of pass/fail grading and curriculum structure with well-being among preclinical medical students: A multi-institutional study. *Acad Med*. 2011;86(11):1367-1373. doi:10.1097/ACM.0b013e3182305d81
16. Dyrbye LN, Thomas MR, Harper W, et al. The learning environment and medical student burnout: A multicentre study. *Med Educ*. 2009;43(3):274-282. doi:10.1111/j.1365-2923.2008.03282.x
17. Hutchinson L. Hutchinson L. Educational Environment. In: Cantillon P, Hutchinson L, Woods D. (eds.) The ABC of learning and teaching in Medicine. *BMJ* 2003;326:810-812. *BMJ*. 2003;326(7393):810-812. doi:10.1136/bmj.326.7393.810
18. Dyrbye LN, Thomas MR, Shanafelt TD. Medical student distress: causes, consequences, and proposed solutions. *Mayo Clin Proc*. 2005;80(12):1613-1622. doi:10.4065/80.12.1613
19. Van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: Design principles and strategies. *Med Educ*. 2010;44(1):85-93. doi:10.1111/j.1365-2923.2009.03498.x

20. Schott M, Kedia R, Promes SB, et al. Direct observation assessment of milestones: Problems with reliability. *West J Emerg Med.* 2015;16(6):871-876. doi:10.5811/westjem.2015.9.27270
21. McGill DA, van der Vleuten CPM, Clarke MJ. A critical evaluation of the validity and the reliability of global competency constructs for supervisor assessment of junior medical trainees. *Adv Health Sci Educ Theory Pract.* 2013;18(4):701-725. doi:10.1007/s10459-012-9410-z
22. Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Acad Med.* 2014;89(5):721-727. doi:10.1097/ACM.0000000000000221
23. Cockayne TW, Samuelson CO. Halo effect and medical student evaluation of instruction. *Acad Med.* 1978;53(4):364. doi:10.1097/00001888-197804000-00016
24. Thomas MR, Beckman TJ, Mauck KF, Cha SS, Thomas KG. Group assessments of resident physicians improve reliability and decrease halo error. *J Gen Intern Med.* 2011;26(7):759-764. doi:10.1007/s11606-011-1670-4

## Figures



**Figure 1**

Student perceptions regarding whether pairing affects evaluation Does being paired with a medical student have any effect on your evaluations (both positive and negative) by the following team members?

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AppendixI.docx](#)