Association Between Formative Assessment and Academic Performance for Undergraduate Medical Students in a Chinese Clinical Skills Training Course

Tianxin Zhu, Junyi Liang, Min Mao, Xintong Liu, Dandan Qian

Background: Previous research suggests that formative assessment (FA) enhances learning outcomes, but few studies have evaluated its impact on clinical skills training in China. We conducted this study in a clinical skills integral curriculum to further explore the educational value of FA.

Material/Methods: Sixty undergraduates from the Second Clinical Medical School of the Southern Medical University in 2016 were selected as the experimental group (consecutive FA), and 50 undergraduates in 2015 were selected as the control term (only final summative assessment, SA). Undergraduates in the FA group completed the after-class questionnaire at each lesson. Teachers, teaching content, assessment objectives, and topics are the same in both groups.

Results: The results of single-factor covariance (ANCOVA) analysis and Cochran-Mantel-Haenszel (CMH) analysis demonstrated that students of the FA group obtained better performance and higher success rates in summative examination than in the SA group. The students with relatively poor grades benefited more from FA, while the performance of students with higher grades was similar between the FA group and SA group. According to the results of questionnaire for students, the satisfaction of students with the course increased gradually, from 84.4% to 93.0%.

Conclusions: Proper use of FA is associated with better learning outcomes for students, especially for those with poorer grades. Our results, together with previous research, indicated that the use of FA may be of great benefit to students’ academic performance and satisfaction with the clinical skills training curriculum.

MeSH Keywords: Clinical Competence • Educational Measurement • Learning • Students, Medical

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/929068
Background

To strengthen the clinical skills and competency of undergraduate medical students, the Chinese government promotes medical education reforms such as “The implementation of distinguished doctor education and training plan” [1]. It has offered an opportunity to assess whether formative assessment (FA) can be effectively implemented despite the summative assessment (SA), which is perceived as being more useful in China. It is important to select an appropriate assessment of clinical skills in the undergraduate setting in China.

Both FA and SA can be used to identify competency in clinical skills training. FA, also known as assessment for learning (AFL), is contrasted to SA, which is often characterized as assessment of learning (AOL). Perceived feedback is essential, and FA provides feedback on performance to modify teaching and learning activities, to help students discover their strengths and weaknesses, and to guide their future learning [2]. FA is used as a regular, on-going assessment in the form of lecture quiz or routine training. Students can develop their skills in an environment that is low-stress, non-threatening, and non-judgmental [3]. SA, on the other hand, is undertaken at the end of a term, chapter, semester, or academic year to assess students’ knowledge or skills and to ensure students have reached goals and objectives, and it is usually presented as a numerical or letter grade score [4]. Correspondingly, SA is a high-stakes assessment with limited or no feedback to drive learning [5].

With SA used in the classroom, students can demonstrate their conceptual understanding as well as think critically when applying their understanding to solve new problems or to explain novel phenomena [6]. However, SA has been associated with undesirable learning strategies and lack of personal reflection or deep understanding of the lesson contents [7,8]. Educational specialists recommend the use of FA in addition to SA [9,10].

There is growing evidence that FA is a valuable tool for enhancing learning. FA has been found to promote student learning in a wide range of educational settings, including disciplinary areas, types of outcomes, levels, and so on [2]. As feedback is a core component in FA, FA provides feedback to students that allows them to make adjustments to their deficiencies and how they are learning [11]. Feedback motivates trainees to engage in appropriate learning and advises trainees regarding observed learning needs and resources available to facilitate their learning activities [12]. A study developed and evaluated a new feedback tool for formative assessment, which is useful in enhancing peer learning in nursing students [13]. One study found that students perceived personalized faculty feedback as beneficial after they received personalized formative feedback from faculty [14]. Apart from improving performance, FA is also a means to ensure trainees meet acceptable levels of competence and minimum safety standards [15]. Although in recent years FA has been significantly developed in China, only a few English-language studies have been performed to evaluate its impact on clinical skills training in China. Additionally, the application of FA in China has encountered problems, such as use of a single form, low implementation, and insufficient orientation. In conclusion, more effort is needed to confirm and explain the effect of FA, especially in the field of Chinese clinical skills training.

A study by Li et al. demonstrated that a pre-training evaluation and feedback group had higher scores than a control group in the post-training evaluation in basic life support (BLS), indicating that pre-training evaluation improved learner performance immediately after BLS training [16]. We therefore speculate that pre-training evaluation in FA may also strengthened learner memory retention, which needs more experiments to verify.

The purpose of this study was to explore the value of FA for use in Chinese clinical skills courses. To achieve this, we compared continuous FA with final summative examination. To determine if pretesting in the FA group could better reveal the shortcomings in student learning and reflect the students’ proficiency compared to the traditional assessment method, the present study applied a pretest immediately after the lecture in the FA group. We also surveyed student satisfaction with the quizzes at the end of each course. We hypothesized that the provision of FA would have a positive impact on student academic performance and satisfaction with our courses.

Material and Methods

Ethics statement

The study was conducted according to principles of the Helsinki and Declaration of Geneva, and was approved by the Ethics Committee of Zhujiang Hospital of Southern Medical University (2018-LCJNZX-001).

Participants and curriculum

We enrolled 110 students belonging to the bilingual class, Second Clinical Medical College of Southern Medical University. The participants were from 2 consecutive years, including 60 undergraduates in 2015 and 50 undergraduates in 2016. The reason for the inconsistency between the 2 grades was that the standard of undergraduate’s entering bilingual class was the same, but the number of students who met the standard was different. The distribution and grade-point average (GPA) of the students are shown in Table 1. GPA can be defined as the pretest score and the result of independent-sample t test was significant in the comparison of GPA between the 2 groups.
Table 1. Distribution and GPA of the students.

| Grade  | Students (No.) | Female/Male (F/M) | GPA mean (SD) | t   | p   |
|--------|----------------|-------------------|---------------|-----|-----|
| 2015   | 60             | 35/25             | 85.32 (2.07)  | 3.053 | .003 |
| 2016   | 50             | 32/18             | 83.95 (2.56)  |     |     |

Both groups received 40 min of theoretical teaching and 80 min of operational teaching of each skill. Among them, the lesson of the 4 puncture procedures involved 4 kinds of skills; therefore, students received 160 min of theoretical teaching and 80 min of operational teaching in this lesson, and the teaching time was evenly allocated to the 4 kinds of puncture. In the theory teaching, the teachers not only delivered theoretical knowledge, but also played standardized operation videos of these skills, focusing on the key and difficult points based on their former teaching experience, as well as pointing out the steps that were easy to forget or to perform incorrectly. In the operational teaching, the teacher provided 60 min of teaching and standard demonstration of these clinical skills, and then the students had 20 min of practice time after teaching. At the end of lessons, there were 5–10 min for after-class question and answer (Q&A) for students.

The major differences between 2 groups were as follow. The FA group (the undergraduates in 2016) immediately completed a baseline skills assessment (pretest), and completed feedback questionnaires after teaching or provided oral feedback after each practice to express their thoughts and confusion. Teachers made personalized comments and gave advice to students according to their academic performance on the pretest, and also improved their teaching content and method in accordance with the main comments and suggestions provided in questionnaires and oral feedback from students. Then, students in the FA group repeatedly corrected their mistakes and improved their operation under the targeted advice of teachers. In contrast, students in the SA group had no pretest after lessons. Although students in the SA group also had teacher guidance when they were practicing, teachers could not adjust their teaching plan in advance according to the performance of the students’ after-class pretest and could not give targeted guidance to students. Therefore, it is likely that students in the SA group were unable to discover their deficiencies or correct their mistakes after 4 weeks of practice. After a 4-week session featuring deliberate practice and feedback, students in both groups completed a posttest. The flowchart of the curriculum is shown in Figure 1.
Table 2. Summary table of ANCOVA analysis of the effects of FA on the achievement of final examinations (N=110).

| Sources                     | SS    | df | MS    | F     | P     |
|-----------------------------|-------|----|-------|-------|-------|
| CPR                         |       |    |       |       |       |
| PRE                         | 3.729 | 1  | 3.729 | .094  | .76   |
| FA                          | 63.769| 1  | 63.769| 1.614 | .21   |
| Change of dressing          |       |    |       |       |       |
| PRE                         | 65.368| 1  | 65.368| .630  | .43   |
| FA                          | .082  | 1  | .082  | .001  | .98   |
| Four puncture procedures    |       |    |       |       |       |
| PRE                         | 1207.685| 1  | 1207.685| 9.388 | .003**|
| FA                          | 3936.026| 1  | 3936.026| 30.598| .000**|
| Aseptic technique           |       |    |       |       |       |
| PRE                         | 7.634 | 1  | 7.634 | 1.116 | .73   |
| FA                          | 21.408| 1  | 21.408| .325  | .57   |
| Basal operative skill       |       |    |       |       |       |
| PRE                         | .083  | 1  | .083  | .002  | .97   |
| FA                          | 646.674| 1  | 646.674| 13.061| .000**|
| Total score                 |       |    |       |       |       |
| PRE                         | 28.098| 1  | 28.098| 1.293 | .26   |
| FA                          | 310.391| 1  | 310.391| 14.282| .000**|

* P<0.05, ** P<0.01. PRE – pretest score; FA – formative assessment

Assessment

The posttest assessment table is an improved version based on the Direct Observation of Procedural Skills (DOPS), and the scoring content of the pretest assessment table is the major part of the posttest assessment table. Each action in these tables was given equal weight using a dichotomous scoring system (done correctly vs. incorrectly). The checklist was designed and reviewed for accuracy and completeness by 8 authors with expertise in clinical skills, simulation-based education, and checklist design. The score table has been used in undergraduate teaching for 3 years, and has been applied to clinical skills competition at the hospital level and school level.

After the theoretical teaching and operational teaching, the FA group students were told to fill out the feedback questionnaires to grade the lessons and express their opinions concerning these lessons. The questionnaire for students consisted of 5 dimensions with 17 items, including Teaching Process (4 items), Course Contents (5 items), Skills Operation (3 items), Student Self-assessment (4 items), and Overall Evaluation (1 item). Thereinto, an open-ended response documented their perspective for the lesson as well as suggestions for improvement. Fifty undergraduates in the FA group completed the after-class questionnaires at the end of each lesson, and there were 5 lessons in this curriculum; therefore, a total of 250 questionnaires were distributed in the study. Among the 250 questionnaires in total that were distributed, 242 questionnaires were returned and all were usable for analysis. The response rate and effective rate were both 96.8%. The questionnaires were designed on a 5-point Likert scale (1=strongly disagree [SD] to 5=strongly agree [SA]).

Statistical analysis

We used SPSS 23.0 for all analyses, and P values of <0.05 were considered statistically significant. To examine the differences in learning achievement between different teaching strategies (FA and SA group), we carried out ANCOVA analysis using the pretest score (GPA) as the variable. Furthermore, we compared the percentage of excellent students who passed the established standard between the 2 groups by CMH analysis. The mark (equal or higher than 85 points) for excellent academic performance was consistent with the standard used in the Chinese university system. A descriptive table expressed the distribution of comments on the teaching questionnaire completed by students in the FA group after each course.

Results

Results of ANCOVA analysis

The results of ANCOVA in Table 2 showed that PRE (pretest score) as a covariate had a significant influence on the POST (posttest score) of 4 Puncture Procedures (P=0.003). Compared with the SA group, the FA group was effective in improving subject achievement of the 4 puncture procedures (paired t test; P=0.000), basal operative skill (P=0.000), and total score (P=0.000). However, no significant effects between FA strategy and subject achievement were found in the course of CPR (P=0.21), change of dressing (P=0.98), and aseptic techniques (P=0.967). Table 3 demonstrates the test standard error and adjusted means for the 2 groups after excluding the influence of basic ability. The data showed that the FA group (adjusted mean=91.534) did significantly better than the SA group (adjusted mean=78.994).
Table 3. Mean score and standard error of the 2 groups on final examination.

| Curriculum                  | Group   | Number | Mean* | Standard error |
|-----------------------------|---------|--------|-------|----------------|
| CPR                         | SA      | 60     | 92.641| .828           |
|                             | FA      | 50     | 94.237| .910           |
| Change of dressing          | SA      | 60     | 90.002| 1.342          |
|                             | FA      | 50     | 90.060| 1.476          |
| Four puncture procedures    | SA      | 60     | 78.994| 1.494          |
|                             | FA      | 50     | 91.534| 1.643          |
| Aseptic technique           | SA      | 60     | 88.989| 1.069          |
|                             | FA      | 50     | 89.914| 1.175          |
| Basal operative skill       | SA      | 60     | 90.060| 1.476          |
|                             | FA      | 50     | 89.909| 1.019          |
| Total score                 | SA      | 60     | 87.304| .614           |
|                             | FA      | 50     | 90.826| .675           |

* Adjusted with covariate: PRE (mean=84.70); FA – formative assessment; SA – summative assessment

Table 4. Number and percentage of the students (SA and FA group) who passed the established standard in the final examination.

| Curriculum                  | Group n (%) | OR    | 95% CI           |
|-----------------------------|-------------|-------|------------------|
|                             | SA (n=12)   | B (n=48) |                 |
| CPR                         | A           | 38     | (79.17)          |
|                             | B           | 38     | (80.00)          |
| Change of dressing          | A           | 8      | (80.00)          |
|                             | B           | 33     | (82.50)          |
| Four puncture procedures    | A           | 6      | (60.00)          |
|                             | B           | 33     | (82.50)          |
| Aseptic technique           | A           | 9      | (75.00)          |
|                             | B           | 26     | (65.00)          |
| Basal operative skill       | A           | 5      | (41.67)          |
|                             | B           | 33     | (68.75)          |
| Total score                 | A           | 10     | (83.33)          |
|                             | B           | 37     | (90.00)          |

Table 3. Mean score and standard error of the 2 groups on final examination.

Table 4. Number and percentage of the students (SA and FA group) who passed the established standard in the final examination.

Results of the CMH test

On average, participants achieved 84.70 (2.39) of the pretest score. When converting the pretest score into 2 group according to the Pareto principle (80/20 principle), 20% of them are defined as Group-A (“excellent”, score >87, n=22), including 12 students from the SA group and 10 students from the FA group. Eighty percent of the others were defined as Group-B (“fair”, score <87, n=88), consisting of 47 students from the SA group and 41 students from the FA group. The CMH tests were performed to examine the differences in the percentages of success rates (passing the established standards: the final examination score ≥85) for the 2 groups. The CMH tests (Table 4) revealed that there was a significantly higher success percentage of students within the FA group to pass the standard on the 4 puncture procedures (OR=9.00, 95% CI 3.40 to 23.79), basal operative skill (OR=4.61, 95% CI 1.99 to 10.72), and total score (OR=4.51, 95% CI 1.41 to 14.40), while the SA group had a higher rate than the FA group in aseptic technique (OR=0.39, 95% CI 0.18 to 0.88). In terms of CPR and change of dressing, the FA group had a slightly higher success rates, but the difference was not significant (Table 4).

Results of the questionnaires

For the students’ questionnaires, all of the students of grade 2016 were polled [32 females (64%), 18 males (36%)].
4131 valid evaluations were submitted. Cronbach’s alpha was 0.82 for the complete questionnaire (items 1–17), which confirmed the internal consistency of the questionnaire in our research. The value of the Kaiser-Meyer-Olkin test we obtained was 0.626 (>0.5) and the Bartlett’s Test of Sphericity was significant ($\chi^2=236.705$, $p<0.000$), which is considered appropriate for factor analysis to test the construct validity of the questionnaire. Analysis of the 17 items revealed 6 factors with eigenvalues above 1, accounting for 68.4% of the total variance. With respect to positive comments ($n=3666$), the mean was 73.32 positive comments per student. The distribution of comments in all of the 5 courses are indicated in Table 5. In the whole curriculum, students commented positively on aseptic techniques (84.4%), change of dressing (85.3%), CPR (89.3%), basal operative skill (91.7%) and 4 Puncture Procedures (93.0%).

### Discussion

The results of ANCOVA analysis demonstrated that students in the FA group obtained better performance in summative examination than the SA group. When it came to 5 clinical skills operations, there were significant differences in basal operative skill and 4 puncture procedures between the 2 groups, while the FA group students’ average scores for CPR, aseptic techniques, and change of dressing were slightly higher than in the SA group, but there the differences were not significant, and further investigations with larger sample sizes are needed to fully characterize the effects of FA on these 3 skills. On the whole, our results provide objective evidence of the positive effects of FA and agree with many previous studies [17–20], that teachers need to identify the areas where students might need support and feedback to avoid student dependence and to maximize learning. Teachers could inform students about gaps in performance and provide feedback to guide their learning, but not draw attention to problems. Rather than pointing out the mistakes immediately, it is better for teachers to focus on giving guidance gradually, which can provide learners with opportunities to recognize their weaknesses on their own and facilitate their learning competence.

According to the comments of questionnaire for students, this teaching model was well received by the FA group. This finding is in line with the literature on the FA educational model [22–24]. With continuous feedback from students and course improvement, the satisfaction of students with the course increased gradually, from 84.4% to 93.0%. The results not only indicated that FA contributed to improvement in student satisfaction with the courses, but also reflected their support for the formative style of assessment used in curriculum reform. For example, teachers could learn from student feedback that the practice time is not long enough and the grouping is unreasonable, then improved it in the next practice time, so as to improve student satisfaction. In summary, formative assessment greatly reduced students’ complaints and improved their objective assessment.

The role of assessment is not simply to assess students on the predetermined criteria, but also to stimulate students’ learning
through a process of feedback and provide them the opportunity to improve [25]. To keep pace with global advancements of FA used in medical education, Chinese curricular reform has been seeking a formative style of assessment, with more and more medical schools now utilizing an “FA system” rather than an “end-point assessment system” in medical education [26]. The end-point assessment system in medical education does not give the students any feedback on their individual responses, only the overall exam grade, while many questions from the lessons and practice for formative assessment do provide immediate feedback. However, the use of FA in clinical skills training is still in its infancy in China. Several medical schools in China have accumulated significant experience and considerable progress when undergoing exploratory reforms in using FA in clinical skills training; nevertheless, they are fragmented and not integrated into a system [27]. Therefore, in our study, a total of 5 clinical skills operations were integrated into the novel teaching strategy of clinical skills. Formative assessment was conducted in the undergraduates in 2016 (the FA group), while traditional summative assessment was conducted in the undergraduates in 2015 (SA group) to explore the value of FA in students’ academic performance and students’ views on this novel teaching mode. A major strength of our study is that the pretest occurred immediately after teaching, deepening the memory and filling a vacuum in theoretical knowledge and operating steps in the shortest time, consistent with the memory curve theory.

There are certain limitations of our study that need to be considered. First, we only conducted FA in a single grade of a single university, involving a relatively small group of students. Therefore, a multi-university study with a larger sample would produce further insights into formative assessment and learner-centered teaching for improving medical students’ skills. Second, we were not able to divide students in the same grade into an FA group and SA group so as to ensure they could equally access educational experiences. Therefore, a historical comparison group was used in this study. Although students in both groups were similar in demographics, there is also a need to perform the method of our study several times in larger samples and across various curricula to further confirm the results. Third, there were only 50 undergraduates in 2016; therefore, the FA and SA groups were of unequal sizes. Fourth, we did not limit or record the students’ practice time for each procedure during 1 month of practice time, so the time schedule of these clinical skills was not evenly distributed. The practice time is arranged by the students themselves after other courses, which may be related to the students’ target score and initiative. For example, some students want to get full marks in the pretest and posttest, while others’ target score is just passing the exam.

Although the study was limited to undergraduates and 5 clinical skills, this approach to formative assessment should be explored in other populations and clinical skills. For example, a similar assessment and teaching strategy could be used with interns and undergraduates in skills teaching. In addition, we expect that this work will encourage other researchers and educational workers to conduct studies exploring the subdivision of scoring contents in examinations as a novel method to better point out the shortcomings of students and reflect the students’ proficiency.

**Conclusions**

Medical education in China is facing a shift from SA to continuous FA. After thorough analysis of the results, we concluded that FA is of great benefit in promoting students’ academic performance and satisfaction with our courses. However, to make better use of the FA method, teachers should be required to be more qualified, while students need to cultivate self-learning and independent problem-solving abilities.

**Acknowledgments**

We would like to thank Jing Tian and Yueping Li for assistance and review of this manuscript.

**References:**

1. Jiang G, Chen H, Wang Q et al: National Clinical Skills Competition: An effective simulation-based method to improve undergraduate medical education in China. Med Educ Online, 2016; 21: 29889
2. Black P, Willam D: Assessment and classroom learning. Assesssm Edu, 1998; 5: 7–74
3. Rolfe I, McPherson J: Formative assessment: How am I doing? Lancet, 1995; 345: 837–39
4. Anziani H, Durham J, Moore U: The relationship between formative and summative assessment of undergraduates in oral surgery. Eur J Dent Educ, 2008; 12: 233–38
5. Epstein RM: Assessment in medical education. N Engl J Med, 2007; 356: 387–96
6. Dixon DD, Worrell FC: Formative and summative assessment in the classroom. Theory Into Practice, 2016; 55: 153–59
7. Smith G: How does student performance on formative assessments relate to learning assessed by exams? J Coll Sci Teach, 2007; 36: 28–34
8. Tshibalo AE: The potential impact of computer-aided assessment technology in higher education. S Afr J Educ, 2007; 21: 686–95
9. Nendaz MR, Tekian A: Assessment in problem-based learning medical schools: A literature review. Teach Learn Med, 1999; 11: 232–43
10. Harden RM, Crosby JR, Davis MH, Friedman M: AMEE Guide No. 14: Outcome-based education: Part 5-From competency to meta-competency: A model for the specification of learning outcomes. Med Teach, 1999; 21: 546–52
11. Riaz F, Yasmin S, Yasmin R: Introducing regular formative assessment to enhance learning among dental students at Islamic International Dental College. J Pak Med Assoc, 2015; 65: 1277–82
12. Nocini J, Burch V: Workplace-based assessment as an educational tool: AMEE Guide No. 31. Med Teach, 2007; 29: 855–71
13. Solheim E, Platth HS, Eide H: Nursing students’ evaluation of a new feedback and reflection tool for use in high-fidelity simulation – Formative assessment of clinical skills. A descriptive quantitative research design. Nurse Educ Pract, 2017; 27: 114–20
14. Valdez C, Shea L, Knutsen R, Hoody D: Facilitating skill development using student-directed activities and personalized formative feedback. Curr Pharm Teach Learn, 2014; 6: 826–33
15. Andreassen P, Malling B: How are formative assessment methods used in the clinical setting? A qualitative study. Int J Med Educ, 2019; 10: 208–15
16. Li Q, Ma EL, Liu J et al: Pre-training evaluation and feedback improve medical students’ skills in basic life support. Med Teach, 2011; 33: e549–55
17. Jain V, Agrawal V, Biswas S: Use of formative assessment as an educational tool. J Ayub Med Coll Abbottabad, 2012; 24: 68–70
18. Minder SP, Weibel D, Wissmuth B, Schmitz FM: Do students achieve the desired learning goals using open-book formative assessments? Int J Med Educ, 2018; 9: 293–301
19. Ismail MA, Ahmad A, Mohammad JA et al: Using Kahoot as a formative assessment tool in medical education: A phenomenological study. BMC Med Educ, 2019; 19: 230
20. Siau K, Crossley J, Dunckley P et al: Direct observation of procedural skills (DOPS) assessment in diagnostic gastroscopy: Nationwide evidence of validity and competency development during training. Surg Endosc, 2020; 34: 105–14
21. Carrillo-de-la-Peña MT, Bailés E, Caseras X et al: Formative assessment and academic achievement in pre-graduate students of health sciences. Adv Health Sci Educ Theory Pract, 2009; 14: 61–67
22. Valero G, Cárdenas P: Formative and summative assessment in veterinary pathology and other courses at a Mexican Veterinary College. J Vet Med Educ, 2017; 44: 331–37
23. Hendriks W, Bakker N, Pluk H et al: Certainty-based marking in a formative assessment improves student course appreciation but not summative examination scores. BMC Med Educ, 2019; 19: 178
24. Mogali SR, Rotgans JI, Rosby L et al: Summative and formative style anatomy practical examinations: Do they have impact on students’ performance and drive for learning? Anat Sci Educ, 2020; 13: 581–90
25. Prashanti E, Ramnarayan K: Ten maxims of formative assessment. Adv Physiol Educ, 2019; 43: 99–102
26. Huang L, Cai Q, Cheng L et al: Analysis of curricular reform practices at Chinese medical schools. Teach Learn Med, 2014; 26: 412–19
27. Zhang N, Henderson CN: Can formative quizzes predict or improve summative exam performance? J Chiropr Educ, 2015; 29: 16–21