

Successful Use of a Competency Step Exam in a Perfusion Education Program

Jeffrey B. Riley, MHPE, CCT; Philip D. Beckley, PhD; Richard D. Tallman, PhD;
Allison S. Spiwak, MSBME

Circulation Technology Division, School of Allied Medical Professions, College of Medicine and Public Health, and School of Educational Policy and Leadership, College of Education, The Ohio State University, Columbus, Ohio

Abstract: The perfusion education program at The Ohio State University uses a step exam to rank students and identify incompetent students in regard to the program learning objectives. The step exam determines student progress from the didactic to the clinical phase. Each student must pass the competency step exam to gain entry to the clinical rotations. The development, use, and results of the step exam are reported. The design and knowledge matrix establish the content validity of the exam. Single test question discrimination and difficulty statistics identify valid exam items. Examples of the exam's predictive ability are presented. The step exam is a 200-question exam using multiple choice items. The exam is modeled after several health-related

national certification exam processes. The exam has content validity based on the published, written objectives for the education program. Each item on the exam has a history of use and meets criteria for difficulty, discrimination, and distraction. The use of a high-stake competency exam in clinical science and medical education programs is controversial and technically challenging. A step exam to have high-stake consequences must be reliable, meet requirements for content validity, and hopefully exhibit predictive validity. **Keywords:** perfusion education, knowledge examination, certification, licensure. *JECT. 2006;38:38-43*

Tests may serve several important purposes in health professions education, namely to measure student mastery and proficiency achievement or to search for and rank talent. While the specific questions (items) chosen for these two types of purposes might vary, the measurement criteria used in their selection have some commonalities: namely, a consideration of their difficulty, discrimination, and distraction. Mastery tests must have basic level difficulty and discriminate only at this required basic skills level. Thus, such tests are designed to separate those who have achieved knowledge mastery from those who have not or those who are competent from those who are not. Talent searches are different, in that they are designed to challenge specifically at the highest-level students, thus discriminating students only at this high level, while frustrating all others below this level.

The testing purpose affects the range of item difficulty, at what ability level each item discriminates, and therefore the resulting distribution of scores. To meet these purposes, there are criteria to evaluate an exam, namely the

distribution of the scores and the test's reliability and validity (1,2). For an exam to meet these criteria, the individual items composing the exam may be evaluated for the criteria of difficulty, discrimination, and distraction (AG D'Costa, personal communication, November 12, 2004). There are three steps to designing a cognitive mastery exam: 1) develop a specification matrix based on a job analysis, 2) develop test items using the "three Ds" (difficulty, discrimination, and distraction) to select appropriate items, and 3) measure test reliability and validity (1,2).

A step exam is a high-stake exam to be mastered before moving forward in a professional education program or to advance in a candidate's professional employment. In the perfusion profession, the American Board of Cardiovascular Perfusion (ABCP; www.abcp.org/) basic science and clinical science exams are examples of professional step exams because many employers and some state licensure applications require exam mastery.

In 1999, the Circulation Technology (CT; www.amp.osu.edu/CT/) Division in The Ohio State University School of Allied Medical Professions (SAMP) adopted the use of a step exam at the end of the last three didactic academic quarters in the perfusion education (PE) curriculum.

For a step exam to have such high-stake consequences, the exam and its items must be reliable, meet require-

Address correspondence to: Jeffrey B. Riley, 152 Atwell Hall, 1583 Perry Street, Columbus, OH 43210. E-mail: riley.267@osu.edu
The senior author has stated that authors have reported no material, financial or other relationship with any healthcare-related business or other entity whose products or services are discussed in this paper.

ments for content validity, and hopefully exhibit predictive validity as well (1). The purpose of this communication is to present the development, reliability, and validity for a step exam used in a perfusion education program.

METHODS

A January 2005 e-mail survey of 19 perfusion program directors asked two questions: 1) does the program use a step exam process, and 2) if the program has a step exam process, does the program stop the student’s progress and require the student to remediate if they do not pass the exam?

The CT step exam measures the knowledge competency of the students just before entering their two clinical rotation quarters. Figure 1 shows the PE curriculum in the CT Division. The CT curriculum fits an activity theory model for a constructivist learning environment (3). Step exams have been used by the medical school at The Ohio State University. The CT exam follows the models of our medical school, health personnel state licensure agencies, and the ABCP national certification exam process. The step exam is composed of multiple choice questions with four options; only one option is correct, and there are three distracters. Although difficult to construct, multiple choice questions are versatile and can be used in settings with large numbers of students (4). A single CT course (Circ Tech 510) in the fifth quarter awards the student’s grade based on the results of the step exam. Successful

completion of the exam is a score greater than 70%. The passing score was set consistent with University grading policy.

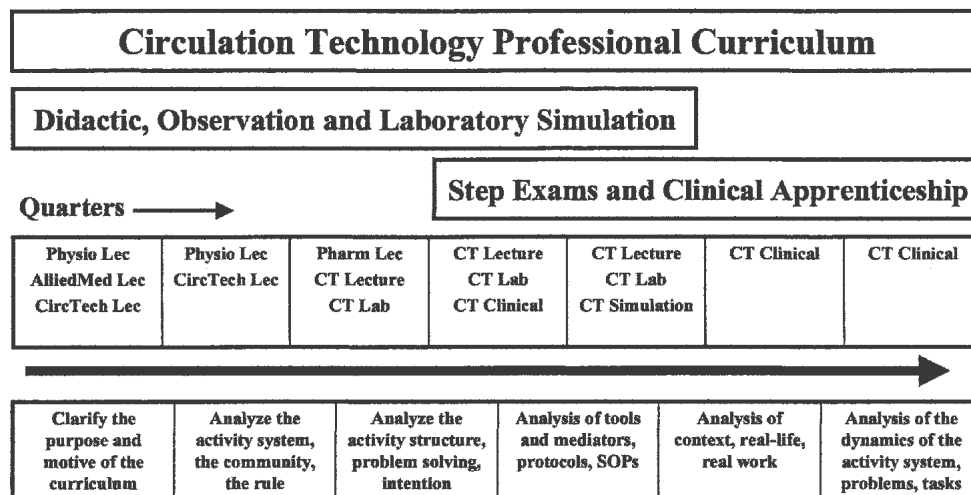
The SAMP curriculum committee approved the use of the step exam for a course score. A defined process of remediation for students who fail to pass the exam was also approved and is published in the CT Student Handbook (http://amp.osu.edu/CT/download/CT_Handbook_2004-05.pdf). Remediation is repeating CT 510 or completing an independent study course to retake the step exam. After remediation and successfully passing the step exam, the student is allowed to enter the clinical rotation clerkships for the last two academic quarters.

A 1968–1972 job analysis for the Circulation Technologist led to the development of the curriculum and has historically focused on nine knowledge and clinical skill areas (5). The nine job areas are shown in Table 1. These curriculum knowledge areas serve as the scales or subsections for the competency step exam score. Written educational objectives for the nine scales are also published in the CT Student Handbook and are well known to the students. The knowledge scales lead to the test specifications in regard to content validity and the learning taxonomy levels used by the test.

A large bank of test items has been developed and used in the CT program. Test items are evaluated using a modified Angoff methodology (6). The Angoff method captures the professional judgment of the faculty to establish defensible mastery standards. Test items are evaluated

OSU Circulation Technology Program Curriculum Activity Theory Framework

Figure 1. The OSU perfusion education professional curriculum is seven academic quarters. The first five quarters focus on lectures, laboratories, and simulation. The last four quarters focus on laboratories, competency check-offs, and clinical rotations.



“Adopted from: Jonassen, D.H. & Rohrer-Murphy, L. (1999). Activity theory as a framework for designing constructivist learning environments. *ETR&D*, Vol 47, No 1. ISSN104201629.

For a list of specific OSU courses goto: <http://amp.osu.edu/CT/download/CTCurriculum.pdf>

Table 1. Circulation technology curriculum knowledge areas.

Biomedical monitoring (BioMed)
Cardiovascular anatomy (CV Anat)
Cardiovascular physiology (CV Physio)
Extracorporeal technology (ECT)
Mechanical cardiac assist (MCA)
Pacemakers and analysis (Pacemaker)
Pathology and surgery (Path & Surg)
Perioperative blood management (PBM)
Pharmacology (Pharm)

The nine competency areas from the Circulation Technologist job description. These areas serve as the sections or scales for the competency step exam. Retrieved from http://amp.osu.edu/CT/download/CT_Handbook_2004-05.pdf.

for difficulty, distraction, and discrimination using simple statistics by a software package designed by Dr. D'Costa (personal communication, November 27, 2004).

Measures of exam reliability were calculated by test-retest and parallel forms methods (1). In late 2004, students from the perfusion education programs at The State University of New York (SUNY Upstate Medical University; <http://www.upstate.edu/chp/cp/>) and Ohio State University had the opportunity to take each other program's perfusion basic science step exam to measure the two tests' concurrent validity.

RESULTS

Ten of 19 program directors responded to the e-mail survey. Three programs (30%) reported to use a step exam process, and two programs (20%) required students

Table 2. Step exam content matrix: knowledge area, domains and item distributions.

Knowledge Scales	Comprehension	Application	Analysis	No. of Items (%)
BioMed	2	3	5	10 (5.0)
CV Anat	25	6	2	33 (16.5)
CV Physio	4	8	7	19 (9.5)
ECT	11	16	22	49 (24.5)
MCA	4	3	7	14 (7.0)
Pacemaker	8	1	2	11 (5.5)
Path & Surg	12	11	0	23 (11.5)
PBM	2	7	6	15 (7.5)
Pharm	18	6	2	26 (13.0)
No. of Items (%)	86 (43.0)	61 (30.5)	53 (26.5)	200 (100.0)

The content matrix is an important tool to establish content validity. The matrix presents the distribution of test items chosen by the program faculty.

to remediate before advancing to clinical rotations in the curriculum if not successful on the exam.

Table 2 shows the content specification matrix for the most recent version of preclinical step exam. The test items are primarily composed of cognitive learning domain items.

Figure 2 presents the frequency distribution for the step exam scores for all students since 1999. Table 3 reports a sample of the item analysis report used to evaluate each exam offering for each class. Figure 3 shows a sample item characteristic analysis for test item 115 taken from the step exam. The number of students selecting a particular item option for the multiple choice question is presented as a measure of distraction.

Distribution of Student Step Exam Scores

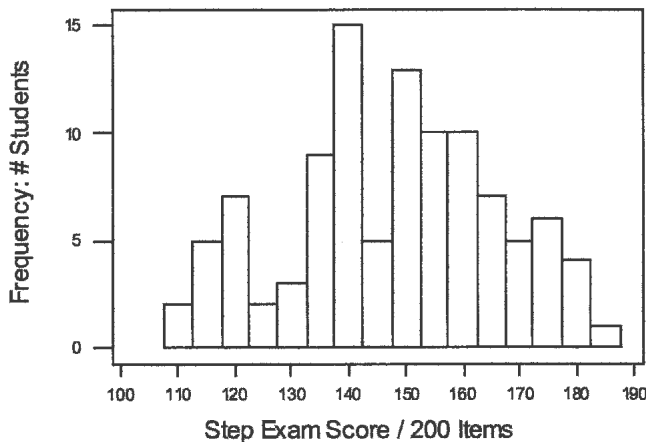


Figure 2. Descriptive statistics for 104 first-time student test taker scores for the CT 510 step exam for classes 1999 to 2004. The student scores are normally distributed.

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Step	104	148.03	149.50	148.24	18.36	1.80
Variable	Minimum	Maximum	Q1	Q3		
Step	110.00	183.00	137.00	160.75		

Anderson-Darling (adjusted) Goodness of Fit for normal distribution = 0.589

Table 3.

Item Number	Subscale	<i>p</i> Value	SD	<i>d</i> Discr	<i>R</i> (Total)	<i>R</i> (Subscale)
112	CV Physio	0.96	0.09	0.22	0.33	0.08
113	Pharm	0.38	0.41	-0.15	0.00	0.14
114	Pharm	0.82	0.28	0.65	0.82	0.52
115	Pharm	0.72	0.19	0.50	0.65	0.51
116	CV Physio	0.68	0.35	0.24	0.45	0.62
117	Pharm	0.82	0.16	0.16	0.42	0.44

Table provides the item analysis report for items 112 to 117 presenting the fraction correct (*p* value), SD, discrimination (*d* Discr), point biserial correlation relative to the total score (*R* Total), and relative to the subscale (*R* Subscale) to which the item belongs.

Test-retest reliability (giving repeated subtests of the exam) measures yielded minimally acceptable values of 0.45–0.55. Cronbach α for the last offering of the test was 0.62 (1). The Pearson correlation between Ohio State University step exam scores ($n = 48$) and the ABCP basic science exam scores is currently 0.38 ($p < .05$).

Table 4 shows the results for the cross-testing of the SUNY and Ohio State University perfusion students with each program’s basic science step exam. Figure 4 is a sample student exam score report that includes scores for each subscale or cognitive area.

DISCUSSION

The CT didactic step exam has high curriculum content validity (Table 2) in that the academic and clinical faculty members have worked to assure that the exam items reflect the curriculum content. Continuous faculty review of the CT job description, the curriculum objectives, and annual feedback from students, employers, and graduates

maintains both the curriculum validity and the content validity of the step exams (7). In recent years, clinical critical incident analysis has resulted in the addition of specific test items (8). The addition of test questions based on critical clinical incidents has increased the difficulty of the exam.

Student performance on items and the exam overall is evaluated by subscales and a mix of varying difficulty questions is used to rank student knowledge. The ideal test item for a step-exam has the characteristics shown in Figure 3. The question difficulty score drops off rapidly for the low scoring quintiles of exam takers.

The Accreditation Committee–Perfusion Education (AC-PE; <http://www.ac-pe.org/>) policies require the completion of prescribed annual graduate and employer surveys with minimum satisfactory threshold scores. The CT curriculum and objectives are continually adjusted based on the feedback from the surveys.

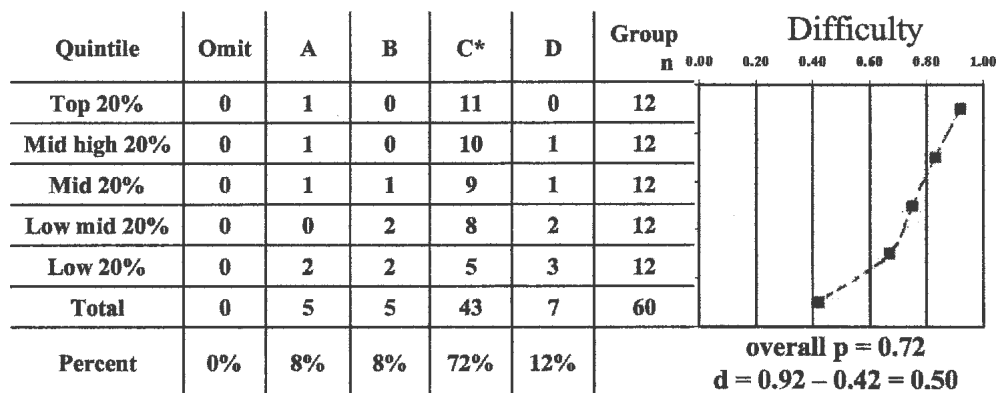
The exam currently shows moderate internal and external reliability. There is room for improvement in Cronbach α , which is a function of the number of test items and the average intercorrelation between the items that measure internal consistency on the exam. The maintenance of a high-stake exam is continuous in that the faculty evaluates items for deficiency—failure to include important items, contamination—items students have not been exposed to in coursework, and distortion—inappropriately assigning weight to an unimportant scale area and for cueing effects (9). Each new version of the exam, depending on the mix of items, increases in regard to internal validity as the faculty improves the question structure and construct.

The step exam development and results process are an

Item Characteristic Report

Item # 115: Subscale = pharmacology

Figure 3. Example item characteristic report for distraction, difficulty, and discrimination from Item Analysis Software by D’Costa. Sixty students’ selection rates for distracters on one pharmacology item are presented: *p* is the fraction correct and *d* is distraction index = top quintile *p* – low quintile *p*.



Personal communication: Item Analysis Software® by D’Costa, A.G.

Table 4. Step exam scores from two perfusion education programs.

Program	Program 1 Exam Score	Program 2 Exam Score	ANOVA <i>p</i> Value
Program 1 (<i>n</i> = 6)	76 ± 3	66 ± 6	.007
Program 2 (<i>n</i> = 14)	61 ± 6	82 ± 7	<.001
ANOVA <i>p</i> value	<.001	<.001	
Regression	Program 1	Program 2	All Students
Exam 1 vs. Exam 2 <i>r</i> ²	.719	.106	.205 (negative)
<i>p</i> Value	.033	NS	.045

Each perfusion programs' students took both programs' exams. The exam scores are the mean ± SD. The regression (Pearson *r*) for all students between exams yielded a negative correlation (Exam 1 = 95.6 - 0.384 Exam 2; *p* = .045). NS, not significant.

important part of an educational program total quality management approach (10–12). How can a student successfully pass four academic quarters of courses and then fail to show mastery on a cognitive step exam? It happens despite the fact that the CT coursework is cumulative and subsequent courses build on knowledge from prerequisite courses. Trends in student performance on step exams offer immediate feedback to faculty. In addition to knowledge content areas, exam subscale scores (Figure 4) may be calculated according to specific CT course objectives, ABCP knowledge areas, specific CT courses, and even CT faculty members. The subscale report for an individual student is useful to identify areas of strength and weakness to focus future study or remediation.

The process of constructing a mastery exam allows the faculty to adjust the test reliability and level of difficulty

by altering the mix of test items. Adding test items with low *p* values (percent correct) to the item mix reduces the potential total score for an individual student but increases the reliability of the exam and its ability to rank student performance. The items with discrimination scores less than 0.20 and *p* values less than .40 are reviewed by the faculty after every exam offering. Decisions to edit a specific item and whether to continue to use of the item were made based on information from the specific item's characteristic analysis (Figure 4). Multiple choice item distracters and effectiveness are evaluated with the same visual aid.

The opportunity for perfusion education programs to share their high-stake examination processes with each other can only lead to quality improvement and increased success in the AC-PE quality outcome measures. If the step exam has high content validity with the AC-PE Consensus Curriculum for Perfusion and has concurrent correlation with the ABCP exam, use of the step exam to remediate students will improve the AC-PE program outcome of success on the ABCP exam.

The method and results to develop a consistent (reliable) and valid step exam for a perfusion education program are presented to show the successful use of a controversial exam with high consequence validity. The program faculty has high confidence in the exam to measure student mastery of the curricular objectives. As the exam continues to evolve and more data to support validity and reliability are collected, perhaps other uses for the exam may be discovered.

ACKNOWLEDGMENTS

The contributions of prior Circulation Technology faculty

One-way ANOVA: *p* value versus OSU_Cat
Analysis of Variance for *p* value

Source	DF	SS	MS	F	P
OSU_Cat	8	0.7170	0.0896	2.32	0.021
Error	191	7.3648	0.0386		
Total	199	8.0818			

Subscale Report

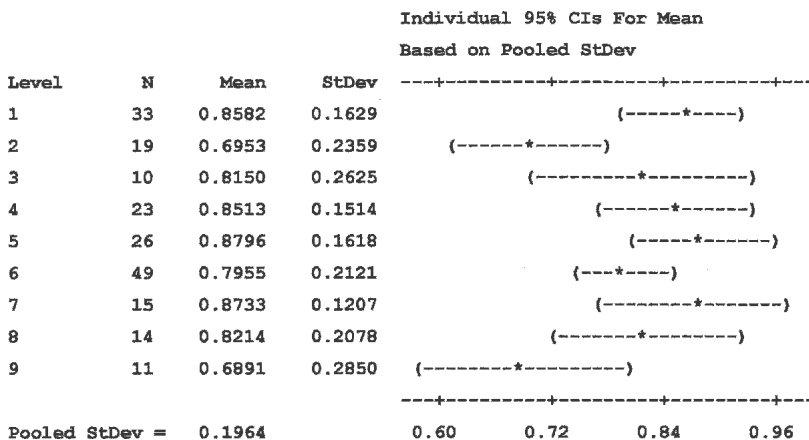


Figure 4. The nine knowledge area mean student scores for the step exam plan from Table 2.

members David W. Holt and Paul Shinko in the past development of the step exam items are gratefully acknowledged. The authors thank Dr. Ayres G. D'Costa for instruction and help in preparing this manuscript and Dr. Larry Sachs for statistical consultation. The cooperation of Bruce Searles and the SUNY program faculty and students is important to establish exam validity.

REFERENCES

1. Hopkins KD. Educational and Psychological Measurement and Evaluation. 8th ed. Boston, MA: Allyn and Bacon; 1998.
2. D'Costa AG. The validity of credentialing examinations. *Evaluation Health Professions*. 1986;9:137-69.
3. Jonassen DH, Rohrer-Murphy L. Activity theory as a framework for designing constructivist learning environments. *Education Technology Research and Development*. 1999;47:61-79.
4. Terry MA. Writing a multiple-choice test question. *J Am Osteopath Assoc*. 1992;92:112-4.
5. Toth LS. From vision, to here, to where? *J Extra Corpor Technol*. 1989;21:40-51.
6. Angoff WH. Scales, norms and equivalents scores. In: Thorndike RL, ed. *Educational measurement*. 2nd ed. Washington, DC: American Council of Education. 1971:580-600.
7. Frankenberg D, Foegelle W. Developing a perfusion technology curriculum using the DACUM process. *J Extra Corpor Technol*. 1983;14:97-100.
8. Flanagan JC. The critical incident technique. *Psychol Bull*. 1954;5:327-58.
9. Schuwirth LWT, van der Vleuten CPM, Donkers HJLM. A closer look at cueing effects in multiple-choice questions. *Med Educ*. 1996;30:44-9.
10. Stammers AH. Perfusion education in the United States at the turn of the century. *J Extra Corpor Technol*. 1999;31:112-7.
11. McCrea C. The quality delivery process: a useful framework for quality improvement initiatives in training? *Med Teacher*. 1996;18:300-3.
12. Dolmans DHJM, Wolfhagen HAP, Scherpbier AJJA. From quality assurance to total quality management: How can quality assurance result in continuous improvement in health professions education? *Educ Health*. 2003;16:17-21.