EDUCATION
RESEARCH SERVICE

LETTER TO THE EDITOR

Using confidence-based marking in a laboratory setting: A tool for student self-assessment and learning

Dr. Kieran Walsh
Clinical Director, BMJ Learning, BMJ Publishing Group, London.

© 2013 Association of Chiropractic Colleges

To the Editor:

Barr and Burke present an intriguing study on the use of confidence-based marking in a laboratory setting, and the correlation between multiple choice questions (MCQ) scores and confidence-based marking scores.[1] They conclude that learners are able to self-assess their knowledge accurately, but it is worth exploring to what degree their conclusion is based on the evidence presented.

First of all, the MCQ assessments that were completed each week had not been validated. If the MCQ assessments had been shown to be valid and reliable, and to have a positive effect on learners' behaviors, and the confidence-based marking scores correlated with such validated MCQ assessments, then we would place much more weight upon such a correlation.[2] However, the authors do not report any prior validation of the MCQ assessments.

Secondly the sample MCQs that the authors list at the end of the study are questionable. The first two sample questions are almost purely tests of academic anatomic knowledge. The last two questions appear to turn best practice in MCQ construction on its head. Good MCQs generally should start with a clinical scenario and then a question, and then a number of possible answers. However, the last two sample questions start with an anatomic lesion and ask the candidate to link this lesion with a clinical scenario. This is counterintuitive – it is the opposite of the typical clinical encounter – and immediately calls into question the validity of these questions.[3]

There is one final strategic point to make. Even if we put these doubts aside, and assume that the assessment was valid and reliable, what should we do with the positive results discovered? Are we so convinced that we can conclude that MCQs are not necessary, and that we can dispense with them and rely solely on learner's confidence scores? I suspect that the answer for most of us is probably not – and if the majority of readers were to agree with this, then it calls into question the core purpose of the research – insofar as it is unlikely to change educational or assessment practice.

REFERENCES

1. Barr DA, Burke JR. Using confidence-based marking in a laboratory setting: a tool for student self-assessment and learning. *J Chiropr Educ*. 2013;27(1):21–26.
2. Schuwirth LW, van der Vleuten CP. General overview of the theories used in assessment: AMEE Guide No. 57. *Med Teach*. 2011;33(10):783–797.
3. Case S, Swanson D. *Constructing Written Test Questions For the Basic and Clinical Sciences*. Philadelphia, PA: National Board of Medical Examiners; 2002.

In Reply

To the Editor:

The reliability and validity of the multiple choice questions would impact the interpretation of results on confidence-based marking. Calculations of point-biserial correlation coefficients for every multiple choice item provide an index of the reliability and validity of the multiple choice questions within a sample. More than 90% of point-biserial correlation coefficients for the multiple choice questions from each quiz administered were greater than .30 (Table 1). A matrix of the distributions of point-biserial correlation coefficients by the percentage of students answering correctly from each quiz administered revealed that 80% of multiple choice questions were acceptable, 17% may require review and modifications, and 3% should be eliminated (Table 2). The distributions of the Kuder-Richardson formula 20 (KR_{20}) coefficients from the quizzes administered over the four trimesters met

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-09-19 via free access

the criterion of adequate test reliability and validity (Table 3). In summary, the reliability and validity of the multiple choice questions used to measure student performance in the neuroscience laboratory course were adequate.

The face validity of the multiple choice questions to assess factual recall, and application of core neuroanatomic topics regarding structure and function was adequate for this first trimester introductory laboratory course. This course directs its efforts primarily at the knowledge and comprehension level of Bloom's and Krathwohl's taxonomies. Therefore, if a student can take either a patient presentation and identify where the lesion might be in the neuroaxis, or be given the level of the lesion and explain what the patient presentation could be, they begin to develop confidence in their knowledge and reasoning skills that reinforces structure/function comprehension. As students progress through the 10-trimester curriculum, they receive training on clinical scenarios that increase in complexity with appropriate correlation to their level of educational training. At these time points in the curriculum, tests of learning objectives for clinical scenarios are presented within the context of clinical encounters.

The authors did not generalize their data findings beyond the current descriptive research to imply any change in educational or assessment practices. The authors adequately presented their results on confidence-based marking to allow readers to assess internal validity of the current descriptive research. With the information presented in the Letter to the Editor and the authors' response, readers have insights on the impact of reliability and validity of the multiple choice questions on confidence-based marking.

Deborah A. Barr, ScD
Jeanmarie R. Burke, PhD
New York Chiropractic College, Seneca Falls, New York

© 2013 Association of Chiropractic Colleges

Table 1 - Distributions of Point-Biserial Correlation Coefficients for Each of the Items From Each Quiz Administered

Point-Biserial Correlation Coefficient	Frequency (Count)
<0	1
0–.10	0
.11–.20	0
.21–.30	7
.31–.40	11
.41–.50	24
.51–.60	37
.61–.70	30
.71–.80	10
>.80	0

Table 2 - Distributions of Point-Biserial Correlation Coefficients by the Percentage of Students Answering Correctly From Each Quiz Administered

	Point-Biserial Correlation Coefficients		
Percent Correct	**>.30**	**≥.15 ≤.30**	**0 < .15**
≤30%	*3*	**2**	**0**
>30% <50%	*17*	*1*	**0**
≥50% ≤80%	77	0	*0*
80%	15	4	0

Matrix Cell Formatting: Bold, Eliminate; Italic, Review; No formatting, Acceptable. Eliminate the one item with negative point-biserial correlation coefficient (not included in the matrix).

Table 3 - Distributions of the KR_{20} Coefficients

	Quiz Topic					
Trimesters	**Spinal Cord**	**Medulla**	**Pons**	**Midbrain**	**Diencephalon**	**Cerebrum**
1	0.56	0.69	0.70	0.72	0.61	0.62
2	0.64	0.68	0.69	0.69	0.67	0.73
3	0.74	0.66	0.66	0.74	0.71	0.67
4	0.65	0.74	0.68	0.70	0.72	0.64

KR_{20} coefficients ≥.50 may be considered acceptable for short item tests (<10–15 items) as compared to the typical criterion of .70 or .80.

Downloaded from https://prime-pdf-watermark.prime-prod.pubfactory.com/ at 2025-09-19 via free access