

ORIGINAL ARTICLE

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

Deborah A. Barr, ScD and Jeanmarie R. Burke, PhD

Objective: Confidence-based marking (CBM), developed by A. R. Gardner-Medwin et al., has been used for many years in the medical school setting as an assessment tool. Our study evaluates the use of CBM in the neuroanatomy laboratory setting, and its effectiveness as a tool for student self-assessment and learning.

Methods: The subjects were 224 students enrolled in Neuroscience I over a period of four trimesters. Regional neuroanatomy multiple choice question (MCQ) quizzes were administered the week following topic presentation in the laboratory. A total of six quizzes was administered during the trimester and the MCQ was paired with a confidence question, and the paired questions were scored using a three-level CBM scoring scheme.

Results: Spearman's *rho* correlation coefficients indicated that the number of correct answers was correlated highly with the CBM score (high, medium, low) for each topic. The χ^2 analysis within each neuroscience topic detected that the distribution of students into low, medium, and high confidence levels was a function of number of correct answers on the quiz ($p < .05$). Pairwise comparisons of quiz performance with CBM score as the covariate detected that the student's level of understanding of course content was greatest for information related to spinal cord and medulla, and least for information related to midbrain and cerebrum.

Conclusion: CBM is a reliable strategy for challenging students to think discriminately-based on their knowledge of material. The three-level CBM scoring scheme was a valid tool to assess student learning of core neuroanatomic topics regarding structure and function.

Key Indexing Terms: Chiropractic; Educational Measurement; Learning

J Chiropr Educ 2013;27(1):21–26 DOI 10.7899/JCE-12-018

INTRODUCTION

Developing reliable and valid assessment tools of student performance that reflect student learning accurately is difficult. Determining what skills and concepts students have mastered led to the development and revision of many educational methods to include more recently team-based learning (TBL), problem-based learning (PBL), and best evidence among others. The current study evaluated confidence-based marking (CBM) as a feedback tool for students learning neuroanatomical structure and function. Over the past 15 years, Gardner-Medwin et al. at University College London developed and applied CBM to the basic sciences courses for medical students using true/false and multiple choice question (MCQ) formats with confidence ratings.^{1,2}

The selection of confidence ratings is based upon mathematical theory in which students should select a rating of moderately confident if greater than 67% sure of their answer, and a rating of highly confident if greater

than 80% sure of their answer.² CBM takes into account how students judge the reliability of their answer to a question; that is, their confidence about their knowledge. Incorrectly rating their confidence results in a penalty score as the goal of medical training is to promote confidence, and accurate clinical diagnoses and referrals in response to acknowledged deficiencies.² In this way, the process of selecting a confidence rating requires the student to review and justify their answers, and cross-reference other knowledge while justifying their answers.² This process emphasizes the comprehension and synthesis of knowledge, for example understanding and learning by the student. CBM provides benefits to students who have a sound understanding of their knowledge base. Through the process of justification and reservation, high confidence ratings gain the best marks, the greatest penalty is assessed for not acknowledging guesswork, while low confidence ratings minimize or eliminate grading penalties by acknowledging reservations, and grading rewards by not being able to assess the reliability of their answer.²

Table 1 - Three-Level CBM Scoring Scheme²

	Confidence Levels		
	Low	Medium	High
Correct answer	1	2	3
Incorrect answer	0	-1	-4

Thus, CBM format of questioning addresses the role of students guessing in the performance assessment, which now allows the instructor to assess student learning more accurately.¹

The educational training of chiropractic students also must promote confidence, and accurate clinical diagnoses and referrals in response to acknowledged deficiencies. CBM is the testing methodology that may serve as an effective self-assessment tool for students to identify better their strengths and weaknesses in decision-making situations as well as identify their gaps in knowledge. CBM also may influence how students synthesize and comprehend knowledge, that is facilitate or promote learning. The purpose of our study is to evaluate CBM as it applies to the educational training of chiropractic students in a first trimester neuroanatomy laboratory.

METHODS

Subjects

The subjects were 224 first trimester students enrolled in Neuroscience I over a period of four trimesters. According to the standard operating procedures of the Institutional Review Board at New York Chiropractic College, data collection was exempt from the policy of US Department of Health and Human Services, Code of Federal Regulations, Title 45, Public Welfare, Part 46, Protection of Human Subjects, because the research was conducted in a commonly accepted educational setting with data collection evaluating the effectiveness of an instructional technique and with minimal risks to the students.

Testing Format of the Laboratory Quizzes

Using an MCQ format, quizzes were administered the week following neuroanatomy topic units of the spinal cord, medulla, pons, midbrain, diencephalon, and cerebrum, for a total of six quizzes over the trimester. Each laboratory consisted of an instructor-directed prosection followed by small group work by the students using anatomic material. At the start of the next week's laboratory, a five-question multiple choice quiz was administered covering the previous week's material. Each content question was paired with a confidence question, and the paired questions were scored (see

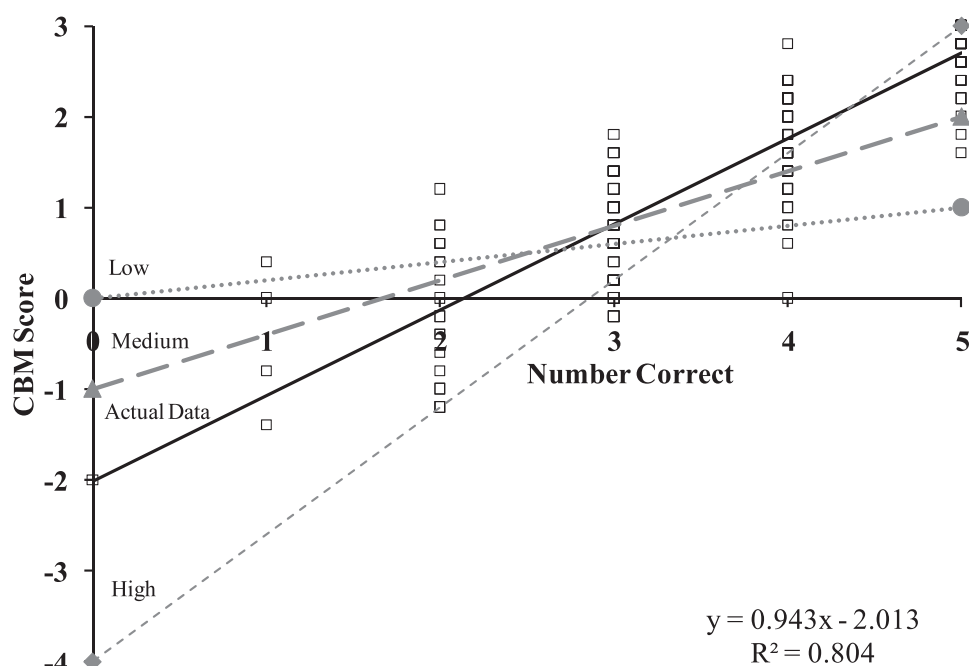


Figure 1 - Scatter diagram between the number of correct answers and CBM score for the spinal cord. The solid black line is the regression line of the actual data for the spinal cord quizzes with data points represented by open squares ($\square$). The regression equation is for the actual data with the R^2 value of the Pearson's correlation coefficient. Superimposed on the scatter diagram were the proper CBM scoring lines for low (gray dotted line with $\bullet$ symbols), medium (gray long dashed line with $\blacktriangle$ symbols), and high (gray short dashed line with $\blacklozenge$ symbols) confidence levels according to the three level CBM scoring scheme. The scatter diagram was constructed for each neuroscience topic to verify the validity of the data according to the three level CBM scoring scheme (not shown).

Table 2 - Spearman's ρ Correlation Coefficients

Neuroscience Topics	Correlation Coefficient
Spinal cord	0.91
Medulla	0.90
Pons	0.88
Midbrain	0.88
Diencephalon	0.86
Cerebrum	0.85

Appendix). Students were given five minutes to complete the quiz, the quiz was collected, and the questions were reviewed orally and answered immediately by the instructor.

Three-Level CBM Scoring Scheme

The CBM scoring scheme was adopted from Gardner-Medwin's experimental work using MCQ questions.² The scoring system requires the students to assess their confidence level as low, medium, or high for each test question. Based upon the answer being correct or incorrect and their confidence rating, a CBM score was assigned to each question according to Table 1. In our laboratory environment, CBM was intended to complement other traditional assessments, such as pinned examinations, and was used as a means to earn bonus points. The total number of correct answers on the six quizzes was the grading criterion.

Data Analyses

For each quiz on the assigned neuroscience topic, the average CBM score and number of correct answers were recorded for each student. The data were compiled across four consecutive trimesters ($n = 224$ students). A scatter diagram of number of correct answers with CBM score was constructed for each neuroscience quiz topic. Regression lines of the actual data were calculated for each scatter diagram. Superimposed on each scatter diagram were the proper CBM scoring lines for low, medium, and high confidence according to the three-level CBM scoring scheme. Then, classification of students into mutually exclusive groups of low, medium, and high confidence was determined using the three-level CBM scoring lines. As a function of the number of correct answers, the three-level CBM scoring lines set the threshold limits for defining the corresponding CBM score into a classification category of low, medium, and high confidence (Fig. 1) is representative of the scatter diagram between the number of correct answers and CBM score that was constructed for each neuroscience topic.

Statistical Analysis

The following statistical procedures were conducted for each neuroscience quiz topic. Spearman's ρ correlation coefficients were calculated between the number of correct answers and CBM score. A single factor ANOVA was performed to compare changes in CBM score as a function of the number of correct

Table 3 - Descriptive Statistics for CBM Scoring (Mean $\pm$ SD, Frequency)

Neuroscience Topics	Number Correct on Quiz				
	0	1	2	3	4
Spinal cord ($n = 224$)	-2.00 $n = 1$	-0.45 ± 0.806 $n = 4$	-0.12 ± 0.760 $n = 21$	-0.74 ± 0.534 $n = 54$	1.74 ± 0.518 $n = 70$
Medulla ($n = 230$)	-0.20 $n = 1$	-0.09 ± 0.540 $n = 7$	0.319 ± 0.530 $n = 19$	0.93 ± 0.451 $n = 55$	1.74 ± 0.441 $n = 63$
Pons ($n = 229$)	-0.90 ± 0.990 $n = 2$	-0.48 ± 0.825 $n = 25$	0.16 ± 0.559 $n = 49$	0.89 ± 0.458 $n = 66$	1.54 ± 0.434 $n = 57$
Midbrain ($n = 217$)	-0.47 ± 0.599 $n = 12$	0.01 ± 0.372 $n = 32$	0.43 ± 0.339 $n = 52$	0.93 ± 0.506 $n = 49$	1.67 ± 0.511 $n = 51$
Diencephalon ($n = 197$)	- $n = 0$	-0.15 ± 0.705 $n = 17$	0.41 ± 0.421 $n = 33$	0.76 ± 0.410 $n = 49$	1.49 ± 0.500 $n = 59$
Cerebrum ($n = 190$)	-0.14 ± 0.313 $n = 10$	-0.28 ± 0.920 $n = 23$	0.34 ± 0.519 $n = 47$	0.86 ± 0.422 $n = 53$	1.55 ± 0.483 $n = 41$
					2.75 ± 0.300 $n = 74$
					2.60 ± 0.379 $n = 85$
					2.61 ± 0.486 $n = 30$
					2.48 ± 0.465 $n = 22$
					2.36 ± 0.535 $n = 39$
					2.53 ± 0.526 $n = 16$

Table 4 - Means $\pm$ SD for Quiz Performance and CBM Score, Mean Rank Order of Confidence Levels across Neuroscience Topics¹ and the Distribution of Students by Confidence Levels

Neuroscience Topics	Number Correct	CBM Score	Mean Rank Order ^a	Confidence Levels (n)			
				Low	Medium	High	Total ^b
Medulla	3.98 $\pm$ 1.073	1.82 $\pm$ 0.881	4.27	27	72	130	229
Spinal cord	3.96 $\pm$ 0.993	1.67 $\pm$ 1.063	4.26	25	59	139	223
Diencephalon	3.41 $\pm$ 1.193	1.19 $\pm$ 0.923	3.34	62	63	72	197
Pons	3.18 $\pm$ 1.194	1.08 $\pm$ 1.029	3.33	52	98	77	227
Midbrain	2.88 $\pm$ 1.392	1.04 $\pm$ 0.925	2.96	77	69	59	205
Cerebrum	2.81 $\pm$ 1.292	0.85 $\pm$ 0.961	2.84	71	64	45	180

^a Mean Ranked Order determined by the Friedman Analysis for group membership into low, medium, and high confidence levels across the six neuroscience topics.

^b The algorithm used to distribute students into low, medium, and high confidence levels did not include students earning a zero on the quiz. The total number of students is slightly less in Table 4 than in Table 3 by the number of students earning a zero on the quizzes, per neuroscience topic.

answers. A χ^2 analysis, test of independence, was performed to determine changes in the proportion of students in the three levels of the CBM scoring scheme (low, medium, and high confidence classification) as a function of the number of correct answers, 3×5 contingency table. The purposes of these statistical procedures were to confirm that the three-level CBM scoring scheme was a valid assessment of students' level of understanding.

The following statistical procedures were conducted to compare quiz performance and CBM scores among the neuroscience topics. A one-way repeated measures ANOVA model was used to reveal differences in quiz performance and CBM scores among the neuroscience topics. Post-hoc, pairwise comparisons among all of the neuroscience topics were conducted, using mean differences and the Bonferroni adjustment for multiple comparisons. The purpose of these statistical procedures was to confirm that the three-level CBM scoring scheme was a reliable assessment of the student's level of understanding across different performance outcomes as function of course content.

Secondary analyses to confirm that the three-level CBM scoring scheme was a valid assessment of the students' level of understanding included one-way repeated measures ANCOVA model for quiz performance across neuroscience topics, with CBM scores as the covariate, and the Friedman test to compare the distribution of low, medium, and high confidence levels (group classification) by neuroscience topic.

All statistical procedures were conducted using SPSS, version 17, at a level of significance of 0.05 (SPSS, Inc., Chicago, IL).

RESULTS

The number of correct answers was correlated highly with the CBM score for each neuroscience topic (Table 2). Within each neuroscience topic, pairwise differences in CBM scores were detected across the number of correct answers ($p < .05$, Table 3). The χ^2 analysis within each neuroscience topic also detected that the distribution of students into low, medium, and high confidence levels was a function of number of correct answers on the quiz ($p < .05$). The

majority of students who correctly answered from 4–5 quiz questions met the criteria for group membership reporting high confidence levels. The majority of students who correctly answered three questions met the criteria for group membership reporting medium confidence levels. The majority of students answering less than three questions correctly met the criteria for group membership reporting low-to-medium confidence levels.

Among the neuroscience topics, pairwise comparisons of course content detected that quiz performance and CBM score were greatest for information related to spinal cord and medulla ($p < .05$, Table 4). Rank order of group membership by confidence levels revealed that students had the highest confidence in their understanding of course content on spinal cord and medulla compared to all other neuroscience topics ($p < .05$, Table 4). Pairwise comparisons of quiz performance with CBM score as the covariate detected that the students' level of understanding of course content decreased in linear manner from course content on spinal cord and medulla to course content on diencephalon, pons, midbrain, and cerebrum ($p < .05$, Fig. 2). In summary, the student's level of understanding of course content was greatest for information related to spinal cord and medulla, and least for information related to midbrain and cerebrum.

DISCUSSION

Assessment by confidence or certainty-based marking (CBM) is not a new concept.^{3,4} However, since the 1990s the work by Gardner-Medwin et al. in the medical school at University College London has furthered this methodology in the medical education setting.^{5–8} Regarding the benefits of CBM, their conclusions are that CBM is a simple strategy for challenging students to think discriminately, based on sound knowledge of the material they are learning. Additionally, it provides self-assessment checks that help students tie together different facets of knowledge that enhances student learning. Further, these objectively marked tests have proven validity and reliability.^{2,3} To our knowledge, this is the first report regarding the CBM approach to self-assessment in the chiropractic education literature.

Health professionals must assess their level of knowledge accurately to provide quality patient care. Accurately

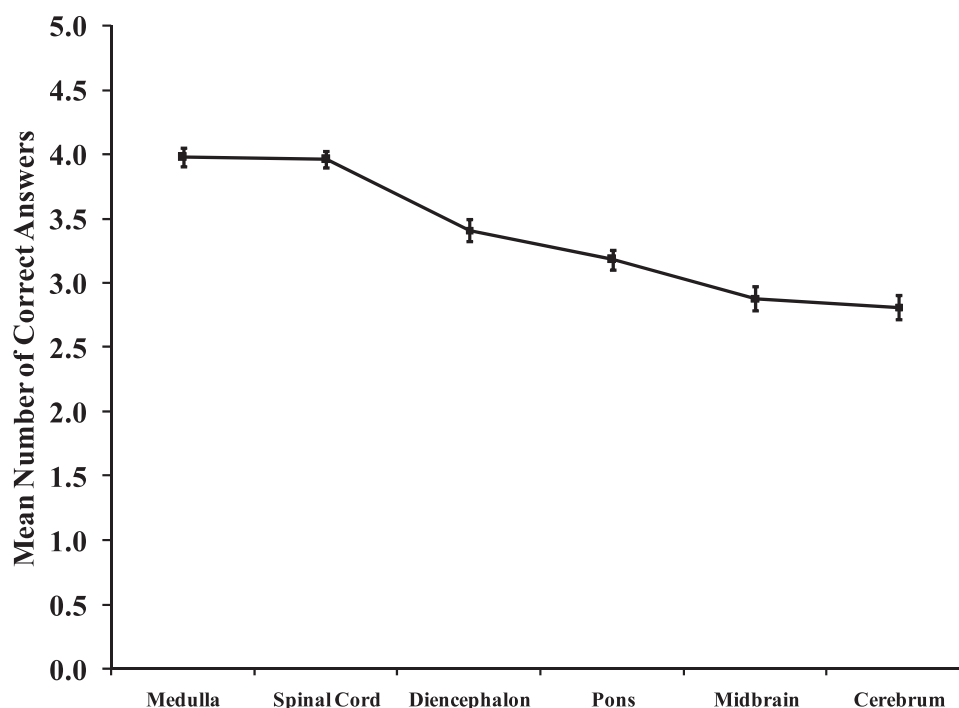


Figure 2 - Pairwise comparisons of quiz performance with CBM score as the covariate. Students' performance was: (1) best for materials on medulla and spinal cord with high confidence, (2) acceptable for materials on diencephalon and pons with medium confidence, and (3) worst for materials on midbrain and cerebrum with low confidence. Error bars are the standard errors of the means.

assessing their level of knowledge will allow them to minimize treatment mistakes and pursue appropriate continuing education hours to increase their clinical competencies. As health professional educators, we must develop critical thinking skills among our students, which include critical assessment of their level of understanding of core content knowledge underlying their scope of practice. The results of our study indicated that the three level CBM scoring scheme was a valid and reliable tool to assess students' level of their ability to recall and apply core neuroscience topics. There were high correlations between the quiz performance and CBM score among the neuroscience topics. Based upon algorithms developed from the proper CBM scoring lines for low, medium, and high confidence, the rank order of group membership by the three confidence levels was congruent with quiz performance across the neuroscience topics. The CBM score as covariate explained the significant quadratic component of descending changes in quiz performance across the neuroscience topics. In addition, the students agreed—strongly agreed that the quizzes (93%) and the three-level CBM scoring scheme (79%) were helpful to them in learning neuroscience.

There were some limitations to this study. The order of quizzes was not random as course content was presented within the normal hierarchy of neuroscience topics from spinal cord to brainstem to midbrain to diencephalon to cerebrum. The materials on the cerebrum and midbrain may be more difficult than the other course materials. However, the timing of quiz within the 1st trimester curriculum may have impacted quiz performances on the midbrain, administered at midterm, and cerebrum, administered immediately

before finals week. Some of the completed data sets were collected from students who withdrew from a previous trimester and then successfully retook the class. These students may have had a better understanding of the three-level CBM scoring scheme. As the three-level CBM scoring scheme was used to accumulate bonus points and not as a grading criterion, any potential biases that may confound the three-level CBM scoring scheme as an effective grading criterion are unknown. Regardless of these potential limitations, the positive relationships between quiz performance and CBM scoring among the neuroscience topics indicated that the students reported their level of understanding of course content accurately.

CONCLUSION

The three-level CBM scoring scheme was a valid and reliable tool to assess students' level of factual recall, and application of core neuroanatomic topics regarding structure and function.

CONFLICTS OF INTERESTS

The authors reported no conflicts of interest.

About the Authors

Deborah Barr is in the basic sciences department and Jeanmarie Burke is in the research department, both at New York Chiropractic College. Address correspondence to Deborah Barr, New York Chiropractic College, Department of Basic Sciences, 2360 State Route 89, Seneca Falls, NY, 13148; dbarr@nycc.edu. This article was received July 25, 2012, revised November 27, 2012, and accepted December 1, 2012.

© 2013 Association of Chiropractic Colleges

REFERENCES

1. Cornwell R, Gardner-Medwin T. Perspectives on certainty-based marking: an interview with Tony Gardner-Medwin. *Innovate* [Internet] 2008 Feb/Mar [cited 2012 Jul 25]; Available online at http://www.innovateonline.info/pdf/vol4_issue3/Perspective_on_Certainty-Based_Marking_-_An_Interview_with_Tony_Gardner-Medwin.pdf.
2. Gardner-Medwin AR. Confidence-based marking—towards deeper learning and better exams. In: Bryan C, Clegg K, eds. *Innovative Assessment in Higher Education*. London: Routledge, Taylor and Group; Francis 2006:141–149.
3. Ahlgren A. Reliability, predictive validity, and personality bias of confidence-weighted scores. In: Gardner-Medwin AR, ed. *Proceedings from American Educational Research Association symposium "Confidence on Achievement Tests- Theory, Applications"*. 1969. Available online at <http://www.ucl.ac.uk/lapt/ahlgren.pdf>.
4. Good IJ. 'Proper fees' in multiple choice examinations. *J Stat Comput Simul*. 1979;9:164–166.
5. Gardner-Medwin AR. Confidence assessment in the teaching of basic science. *Assoc Learn Technology J*. 1995;3:80–85.
6. Gardner-Medwin AR, Gahan M. Formative and summative confidence-based assessment. *Proceedings of the 7th International Computer-Aided Assessment Conference* [Internet] 2003 Jul; Loughborough, UK [cited 2012 Jul 25]; Available from: <http://caaconference.co.uk/pastConferences/2003/proceedings/gardner-medwin.pdf>
7. Gardner-Medwin AR. Confidence-based marking: encouraging rigour through assessment: *Proceedings of the Physiological Society, University of Bristol*. *J Physiol*. 2005;567P, WA10.
8. Gardner-Medwin T, Curtin N. Certainty-based marking (CBM) for reflective learning and proper knowledge assessment. *Proceedings of the REAP International Online Conference: Assessment Design for Learner Responsibility* [Internet] 2007 May 29-31; United Kingdom [cited 2012 Jul 25]; Available from: http://www.reap.ac.uk/reap/reap07/Portals/2/CSL/t2%20meta-cognition/Certainty_based_marking_for_reflective_learning_and_knowledge_assessment.pdf

APPENDIX: SAMPLE QUESTIONS

MC questions (a) were posed followed by the paired confidence question (b) for each question. Initial quiz questions (1, 2) asked basic recall of knowledge, and over time progressed to include comprehension of basic structure/function questions (3, 4).

1a. Which of the following cranial nerves is located in the pre-olivary sulcus?

- A. CN IX Glossopharyngeal
- B. CN X Vagus
- C. CN XI Accessory
- D. CN XII Hypoglossal

1b. My level of confidence in the answer I chose in 1a is:

- A. 1 (Low—not sure)
- B. 2 (Medium—pretty sure)
- C. 3 (High—very sure)

2a. When looking at a specimen of the posterior brainstem and the diencephalon, what structure is found between the pineal gland and the pulvinar of the thalamus?

- A. Superior colliculus
- B. Inferior colliculus
- C. Habenula
- D. Lateral geniculate body

2b. My level of confidence in the answer I chose in 2a is:

- A. 1 (Low—not sure)
- B. 2 (Medium—pretty sure)
- C. 3 (High—very sure)

3a. A lesion of the Crus Cerebri on one side results in which patient presentation?

- A. Ipsilateral flaccid paralysis
- B. Ipsilateral spastic paralysis
- C. Contralateral flaccid paralysis
- D. Contralateral spastic paralysis

3b. My level of confidence in the answer I chose in 3a is:

- A. 1 (Low—not sure)
- B. 2 (Medium—pretty sure)
- C. 3 (High—very sure)

4a. A lesion to the Precentral Gyrus results in which patient presentation?

- A. Contralateral flaccid paralysis
- B. Contralateral spastic paralysis
- C. Contralateral loss of deep sensibility
- D. Contralateral loss of pain and temperature

4b. My level of confidence in the answer I chose in 4a is:

- A. 1 (Low—not sure)
- B. 2 (Medium—pretty sure)
- C. 3 (High—very sure)