CONFIDENCE-BASED ASSESSMENTS WITHIN AN ADULT LEARNING ENVIRONMENT

Paul Novacek

University of North Texas – Learning Technologies, Denton, Texas, USA

ABSTRACT

Traditional knowledge assessments rely on multiple-choice type questions that only report a right or wrong answer. The reliance within the education system on this technique infers that a student who provides a correct answer purely through guesswork possesses knowledge equivalent to a student who actually knows the correct answer. A more complete assessment is needed to eliminate guesswork and offer greater efficiency in managing the remediation process. Confidence-based assessment techniques integrate the selection of multiple-choice answers with the student's self-perceived level of certainty and offer a middle ground between the traditional multiple-choice answer and a lengthy essay response. Research has discovered that confidence-based assessments provide a more comprehensive measure of a person's knowledge, increases the retainability of learned material and identifies topics in which people are misinformed.

1. INTRODUCTION

The traditional use of multiple-choice questions reward a student for guessing. Students are often told when preparing for an exam that even if they are unsure of the correct answer, they should answer it anyway, because with a multiple-choice selection there is a 20% (5-choice) to 25% (4-choice) chance of guessing the correct answer. Their odds are even better when guessing on a true/false choice; hence the reason students' poke fun at the process, calling it a "multiple-guess" exam. There is an effort to maximize the score instead of gaining an understanding of the course material. But in this world of number-crunching rationalizations, it requires much less effort to assign a number (numeric test score) to represent a level of understanding that can be quantified, studied and managed. Is this really an effective way of measuring a student's comprehension of a complex subject?

Some research has been performed with explorations into the implementation of confidence-based assessments across various academic disciplines. Most notably, the work of Darwin Hunt, James Bruno and Tony Gardner-Medwin has provided exhaustive research in the field, with many of the assessment techniques moving into more focused research and commercial operations. Dr. Darwin Hunt began his research in the early 1980s, discovering highly correlative measures between human self-assessment and learning. His work since then has continued to provide evidence that a confidence-based approach to knowledge assessment provides a more comprehensive measure of a student's knowledge, including the retainability of learned material.

Hunt (2003) found evidence of the obvious, that when a student is given a selection of answers for a simple arithmetic problem, he or she still has a chance to select the correct answer even if they don't know how to add two numbers. The regrettable part is that the reliance within the education system on this technique infers that a student who provides a correct answer purely through guesswork possesses knowledge equivalent to a student who actually knows the correct answer. This situation is even more damaging (Adams & Ewen, 2009), as it presents numerous problems for academic institutions in their attempt to offer a fair and representative evaluation of a student's knowledge that can be compared against a standard.

Guessing on a few questions in Math-101 may be perceived as somewhat benign, but taken to its eventual limits, the most damaging aspect of this guesswork is within a safety-critical environment.

Many industrial accidents, injuries and even deaths have been caused by misinformed operators who held a steadfast belief that a factoid of knowledge was correct when it was indeed wrong.

2. CONFIDENCE-BASED ASSESSMENTS

Traditional knowledge assessment methods attempt to focus on the recall of previously-presented information. Often these assessments rely on multiple-choice exams, tests or quizzes that only measure a right or wrong answer. But exploring the real assessment of knowledge needs to eliminate any guesswork involved. This issue is generally mitigated with a short list of techniques that are often ineffective or burdensome. One of the more popular techniques is to give the student a very long test, with upwards of 50, or more, multiple-choice questions. This method is somewhat effective in reducing the negative aspect of missing a few questions, and also offers a more uniform assessment of knowledge because multiple questions can be asked about a single topic from different directions, but the long test takes a lot of time to complete and is quite overwhelming to apprehensive students. Another method of mitigating guesswork is to administer questions that require fill-in, or essay type answers, but these techniques are again daunting to students and burdensome to grade with large class sizes, not to mention the problems associated with a subjective evaluation grading scheme.

Some studies have identified the lack of knowledge retention even after a student successfully passes traditional assessments. Adams et al. (2009) postulated that, "…even when students pass these [traditional] assessments, they lack the necessary skill set to perform well in the work place (p. 1)." This infers that even a high test score is not a sufficient indicator of subsequent job performance, although as a society, we generally hold this view. Many variables are involved when attempting to predict job performance, with knowledge retention being a major contributor.

Adams continues, "For years we have been teaching and assessing using traditional models that encourage guesswork in the testing process. There are even strategies on test taking provided to students to maximize their score, rather than gain an understanding of their knowledge acquisition, skills and competency to perform (p. 1)." An entire cottage industry has evolved around the "just get me through the test" cram course.

A common practice among entrance exam or certification test takers is to purchase test-prep books or software that are just reprints of test question banks. This rather defeats the purpose of an exam in the first place, since answers can be easily memorized. It does not promote even the slightest hint of understanding, much less correlation of the course material. The question beckons, do we really want question memorizers to operate the machinery of our industrialized society?

2.1 Confidence-Based Assessment Method

Short of turning every exam into a drawn-out creative writing exercise, the implementation of a confidence-based testing methodology is a way to effectively gain a quality assessment of a student's knowledge retention while still being able to quantify the results. Confidence-based assessments offer a middle ground between the traditional multiple-choice answer and a lengthy essay response. While taking a multiple-choice test, students indicate which answers they believe are correct while also indicating how confident they feel with their selections. A combined composite score is the result, with a rating scale technique used to reduce the variables.

The combined score assigns greater weight to more confident right answers, and penalizes highly-confident wrong answers. This scheme effectively eliminates overconfident learners who assign a high confidence level to all their answers. Upon analysis by the instructor, or self-realized by the student, a confident wrong answer would deserve special attention and greater emphasis during follow-up learning sessions. The effect is to eliminate, or severely reduce, any guesswork to reveal a true assessment of a student's knowledge.

2.1.1 Marking System

The instrument of a confidence-based assessment scheme uses a double-tiered marking system to gain two distinct but interrelated scores in response to a specific question, although its effectiveness is dependent on

the type of knowledge being assessed. Within a testing scheme of clearly black and white options, such as math, hard science or medical subjects, the confidence-based assessment model provides easily-quantifiable results. Whereas, in the soft sciences or art related subjects, the confidence-based assessment model may only lend support to traditional (subjective) evaluations.

Once a test taker marks the answer which they believe is correct, they are asked to mark their level of confidence. Previous studies have used either a three-level choice (Gardner-Medwin & Curtin, 2003a & Bruno, 1995) or a five-level choice (Hunt, 2003), providing ease of use and sufficient data spread for statistical analysis. Both scales have been used in experimental trials, but in a study that specifically monitored the range of certainty markings, Gvozdenko et al. (2007) discovered that 85% of his student subjects preferred the shorter scales. Additional studies still under investigation are experimenting with variations on this methodology.

2.1.2 Grading Scheme

Table 1 shows the grading scale used by Gardner-Medwin & Gahan (2003b) within an active biomedical student curriculum. The cumulative grade using the certainty-based grading scale assigns a grade straight across for low certainty, but in the medium and high certainty levels, bonus points are awarded for correct answers, while a greater proportion of points are deducted for wrong answers.

Table 1. Grading scales used by Gardner-Medwin & Gahan (2003b) in a biomedical student curriculum

Degree of Certainty	C=1 Low certainty	C=2 Medium certainty	C=3 High certainty	No reply
Mark if Correct	1	2	3	0
Penalty if wrong	0	-2	-6	0

Gardner-Medwin successfully employed this three-level confidence scale, where the upper two certainty levels apply negative weights for wrong answers, effectively assigning a greater penalty for each higher level of misinformation. This gradation is critical, and weighs heavily in the test taker's decision for each answer, offering motivation for a more reliable self-assessment.

An important element within this confidence-based assessment scheme is that test takers know and understand the marking methodology, especially that points will be taken away from their final score for highly-confident wrong answers. If the confidence-based scheme is going to be used as a formative assessment, then the test taker must be informed of its use as a self-assessment tool and how it could be used effectively in their own motivation and appraisal of future learning parameters.

2.1.3 Benefits

From the students' perspective, when offered the use of an interactive assessment that integrated confidence awareness, the instant feedback allows them to quickly confirm their beliefs in selecting the right answer or to correct their perception of wrong answers. This scheme offers a learning tool that would greatly enhance their knowledge absorption and retention. (Gardner-Medwin et al. 2003a, 2003b & 2006)

Likewise, the teacher can obtain a grade report across an entire class, which can then be used to determine the teaching effectiveness of the courseware materials or teaching techniques across groups. Although that effectiveness is dependent on close control of group variables, it can easily determine that all-important check of a student's, or group, knowledge level either at the beginning of a course, or at pivotal points within the course. The information can then be used to modify the course curriculum or place special emphasis on weak areas.

The eventual goal is that the individual student completes the course with a higher retention rate and is fully prepared for the next course in the series. Adams et al. (2009) concluded that, "…confidence and knowledge are correlated and are both critical determinants in evaluating future performance (p. 2)." Hunt (2003) also supports that self-assessment testing provides measurements more closely related to a person's later performance than the common multiple-choice test. But, there are limitations to the process, as Khatibi et al. (2010) found gender differences in a retrospective study of biochemistry students' summative exam scores. The female students' confidence-level scores were significantly higher than the males, suspecting different psychological characteristics as a factor.

Yen et al, (2010) compared the addition of a confidence-weighted component to a computer-administered, multiple-choice exam, finding a correlation between confidence and individual abilities. Additionally, their study found that confidence-based exams became overall more efficient, requiring fewer individual questions to estimate examinees' ability.

The use of a confidence-based assessment scheme appears to be very popular with students, especially when offered as a formative tool during a course, as they see it as a method to improve their overall grade. Gvozdenko et al. (2007) quantified this when monitoring the use of the certainty scale in formative and summative testing. The study found that, "…the majority of students (67%) on a summative test and nearly all students (96%) on a formative test chose to use the certainty scale as an additional tool in the testing procedure (p. 215)."

3. CONCLUSION

It is safe to conclude that the implementation of a confidence-based assessment scheme is beneficial in achieving the objectives of a course when used as a formative evaluation tool. Confidence-based assessments offer a middle ground between the traditional multiple-choice answer and a lengthy essay response, resulting in a quality measure of a student's knowledge retention while still being able to quantify the results against a standard. Additionally, the confidence-level marking scale is easily understood and accepted by students when it was used in secondary and post-secondary schools.

Much of the functionality of a confidence-based assessment platform is already integrated into popular Learning Management Systems (LMS). A few well-established academic LMS platforms (Blackboard, Moodle, Schoology, etc.) contain assessment modules, but Moodle (www.Moodle.org) has already implemented a robust confidence-based assessment module, offering a wide variety of development options, which has been used by Gardner-Medwin in both his formal research studies.

The researchers throughout this review primarily explored the effects of a confidence-based assessment scheme within secondary and post-secondary school environments. Little research has been performed on the effects within a purely adult-oriented, corporate or skills-based technical learning environment. Further studies outside of academia may uncover additional benefits or limitations associated with these non-traditional situations. Additionally, it is suspected that there may also be cultural implications when implementing a confidence-based assessment tool, especially within Asian cultures who generally find it unnatural to rate themselves above the average.

REFERENCES

Adams T. M. & Ewen G. W., 2009. The Importance of Confidence in Improving Educational Outcomes. *Proceedings of 25th Annual Conference on Distance Teaching & Learning*. Aug. 4-7, Madison, Wisconsin.

Bruno J. E., 1995. Information reference testing (IRT) in corporate and technical training programs; UCLA

Gardner-Medwin T. & Curtin N., 2003a. Certainty-Based Marking (CBM) for Reflective Learning and Proper Knowledge Assessment. *Assessment design for learner responsibility.* May 29-31, 2007.

Gardner-Medwin A. R. & Gahan M., 2003b. *Formative and Summative Confidence-Based Assessment* Proc. 7th International Computer-Aided Assessment Conference, Loughborough, UK, July 2003, pp. 147-155

Gardner-Medwin A.R., 2006. Confidence-Based Marking - Towards Deeper Learning and Better Exams. In Bryan, C. & Clegg, K. (eds) *Innovative Assessment in Higher Education*, Routledge, Taylor and Francis Group Ltd, London.

Gvozdenko E. & Chambers D., 2007. Applying Computerized Testing & Certainty Based Assessment to Reveal More about Student Learning. *International Journal of Learning*. Vol. 13, No. 12.

Hunt D., 2003. The concept of knowledge and how to measure it. *Journal of Intellectual Capital*, 4 (1), pp. 100-113

Khatibi R. et al, 2010. Assessment of the Relation between Students' Gender and Their Scores on Selecting Confidence Choices in Confidence-Based Exams. *Iranian Journal of Medical Education*. 2012; 11 (8) :926-932

Yen, Y. C. et al, 2010. Development and Evaluation of a Confidence-Weighting Computerized Adaptive Testing. *Educational Technology & Society*, 13 (3), pp. 163–176.