Confidence Testing for Knowledge-based Global Communities

Brady Michael Jack
National Kaohsiung Normal University, Taiwan R.O.C.
Chia Ju Liu
National Kaohsiung Normal University, Taiwan R.O.C.
Houn Lin Chiu
National Kaohsiung Normal University, Taiwan R.O.C.
James A. Shymansky
University of Missouri at St. Louis, MO.

Please address all correspondence to:
Brady Michael Jack
National Kaohsiung Normal University
Graduate Institute of Science Education
No.62, Shenjhong Rd., Yanchao Township, Kaohsiung County 824, Taiwan (R.O.C.)
Phone number: 011-886-7-780-5824
Fax number: 011-886-7-780-5823
E-mail address: bradyjack@gmail.com

Paper presented at the
Annual Meeting of the American Educational Research Association
San Diego, April 13–17, 2009

Confidence Testing for Knowledge-based Global Communities

Brady Michael Jack
National Kaohsiung Normal University, Taiwan R.O.C.
Chia Ju Liu
National Kaohsiung Normal University, Taiwan R.O.C.
Houn Lin Chiu
National Kaohsiung Normal University, Taiwan R.O.C.
James A. Shymansky
University of Missouri at St. Louis, MO.

This proposal advocates the position that the use of confidence wagering (CW) during testing can predict the accuracy of a student's test answer selection during between-subject assessments. Data revealed female students were more favorable to taking risks when making CW and less inclined toward risk aversion than their male counterparts. Student comments suggested CW was a good way of self-regulating point value of selected answers, assisting teachers and students on revealing which areas needed further instruction and practice and where the student's confidence was incorrectly placed. CW could be modified to assess a student's level confidence of domain knowledge in other courses and may be of assistance to U.S. educators interested in researching confidence accuracy of reflective knowledge during testing.

Keywords: assessment, confidence testing, knowledge-based economy, reflective knowledge testing

It is well-known that confidence bolsters activity in the market place and the lack of such assurance slows such activity (Taniuchi, 2000). The same is true in a knowledge-based economy. Knowledge within such an economy is only valuable to the extent it is found creditable, useful and dependable. The driving force behind each of these three characteristics is confidence. To merely acknowledge an idea as being creditable without confidently trusting in it, useful without confidently using it, or dependable without confidently relying upon it are all somewhat superficial forms of acknowledgement. Instilling confidence in the knowledge that one can learn new and different ways of doing things is a key component for overcoming feelings of alienation and loneliness (Holt, 2005). The strength of an individual's confidence is directly correlated to the judgments that are made during times of decision (Cowley, 2004) and within the familiar world that is well-known to us (Wulf, 2004). What happens when our, "well-known" world within the classroom changes or is in a constant state of flux? The United States Center for Education Statistics reported the percentage of U.S. public school students who were considered to be part of a racial or ethnic minority group increased from 22 percent in 1972 to 31 percent in 1986 to 43 percent in 2006 (Planty et al., 2008). With such a continual infusion of ethnic diversity, the circle of knowledge within U.S. public schools is not reflective of a mono-western cultural but of a multifaceted knowledge-based global community. This proposal introduces a new practical method of analyzing how confident participants within such a community are toward their reflective knowledge during testing. This utilitarian form of analysis uses confidence wagering to evaluate how accurate middle-school students are in their confidence regarding answer selection during multiple-choice tests. Such wagering encourages the student to make honest evaluations of his/her reflective knowledge by giving control of the point value of an answer to the student. This paper proposes confidence wagering as a non-invasive method of evaluating a student's confidence toward reflective knowledge during testing.

Much has been done in developing testing strategies to measure knowledge and confidence. Of debate among many is the degree self-assessed levels of confidence can accurately and objectively reflect knowledge during testing. Some speculate the major problem of such assessment is a tendency toward exaggerated perceptions of personal accomplishments (Ross, 2006). Others believe assessing confidence during testing is potentially disruptive because it adds another source of variance, which is likely to negatively affect the reliability and validity of a test (Bar-Hillel, Budescu, & Attali, 2005).

Contrary to such positions are studies that view confidence as an important factor in predicting accuracy performance. Several factors have emerged as having an important influence upon the relationship between confidence and accuracy during testing. These factors include reflection and self-regulation (Liddell & Davidson, 2004), immediate feedback (Isaacson & Fujita, 2006), time between what is learned and what is assessed (Brewer, Sampaio, & Barlow, 2005; Lindsay, Nilsen, & Read, 2000; Perfect, Hollins, & Hunt, 2000), degree of experience or the level of mastery over what has been learned (Borgmeier & Horner, 2006; Isaacson & Fujita, 2006; Liddell & Davidson, 2004) level of distractions surrounding what is to be learned (Izaute & Bacon, 2006), and the method by which accuracy and confidence are measured (Patterson, Foster, & Bellmer, 2001). This proposal introduces a new practical method of analysis which uses the utility of wager to evaluate the confidence and accuracy of middle-school science and mathematics students during multiple-choice tests.

Theoretical framework

Research has shown that as the time between a memory recognition task and the time of confidence judgment regarding that task increases, the accuracy of such a confidence judgment is negatively affected (Brewer & Sampaio, 2006). To obtain a state where the accuracy of a confidence judgment and a memory recognition task is unaffected by lag time, the variable used to denote a memory recognition task must be meaningfully connected to the confidence judgment. This study uses the utility of wager to meaningfully connect the level of confidence a student has toward an answer selection at the moment of decision. Each wager measures the student's inclination toward risk seeking or risk aversion. In order to mathematically calculate such inclination, a Risk Inclination Model (RIM) (Jack, Liu, Chiu, Hung, & Shymansky, submitted) was created. RIM uses Varignon's Theorem to calculate the nth factorial moment of probability among the distribution of student wagers of risk within a test set. Using RIM, a Risk Inclination Index (RII) can be constructed to define all possible combinations of risk inclination a test taker could exhibit at the moment of answer selection within a predefined number of point-wagers.

Methodology

Rationale

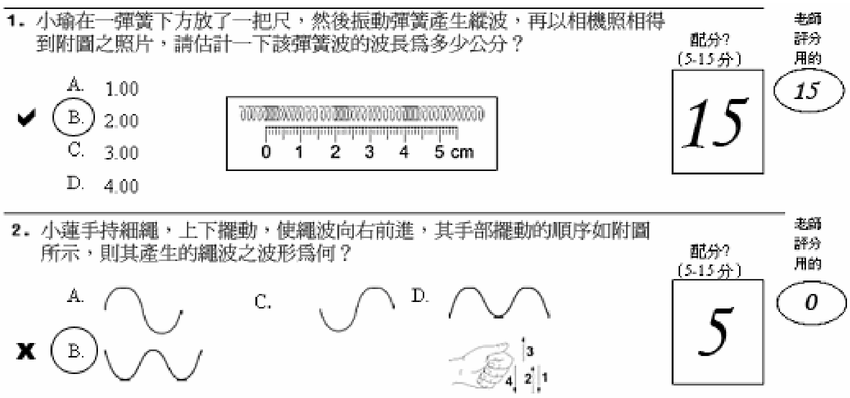
Confidence wagering (CW) evaluates the confidence and accuracy of middle school science and mathematics students at the moment of answer selection during multiple-choice tests. The results from this exploratory study provide information regarding how confidence assessments affect such accuracy. Research questions used during this study: (1) Can CW reveal a relationship between confidence and accuracy during testing? (2) How do students feel about the benefits and disadvantages of such wagering?

Participants

Eight-five Taiwanese students from two mathematics classes and three science classes participated in this study. The mathematics teacher responsible for teaching these two classes and the science teacher responsible for teaching the three science classes also participated. At the end of each week of instruction, students were given a 10-question multiple-choice test. The question and answer selection content were controlled by the teachers and reflected each week of instruction.

Testing format

The design format of each CW test required the student to write how many points he/she felt his/her answer select was worth. This method of allowing students to make honest confidence judgments about the worth of an answer selection was used by James A. Shymansky during his years of teaching mathematics and science. Shymansky (2007) discovered when students were allowed to make such judges; they performed better than on traditional teacher-determined



correct; x incorrect

FIGURE 1. *Test question components*

score-valued tests. The CW design layout has four basic components: 1) test question/statement, 2) multiple choice answer selections, 3) a box where the student writes 5, 10, or 15 indicating the point value he/she thought his/her answer was worth, and 4) a circle area where the teacher writes the student’s assigned answer value if the answer is correct or zero if the answer is wrong. The sum of these circled values equals the final quiz grade. Figure 1 shows a two-question example design layout of each test paper.

TABLE 1. *Collated teachers’ statement with two different examples*

IN	Likert scale statement from teachers	
11.	Questions (Example 11) that match theoretical concepts with examples in real-life or vise-versa are easier to answer than questions that require me to identify what is true or not true about a concept or misconception.	
	Example used on Form M	Example used on Form S
	Example 11: A person takes $400NT to a fruit market. If he buys 3 apples and 5 pears, he is left with $30NT. If he buys 5 apples and 4 pears, he has no money left over. Would the following equation help the person know the cost of 5 apples and 4 pears? $\begin{cases} 3x + 5y = 370 \\ 5x + 4y = 400 \end{cases}$	Example 11: While holding a pen point down, you stick half of it into a glass of water. The part of the pen that is in the water appears not to be in a straight line with the part of the pen that is out of the water. This appearance is caused by light A. reflection B. refraction C. diffusion D. transmission

IN: Item Number; NT: New Taiwan dollars (roughly $32NT per $1US dollar)

Assessment

At the end of three weeks using CW, students and teachers were asked to write down their thoughts and feelings about its use during testing. Their comments were analyzed using a four-step approach: 1) collection of fragmented responses, 2) collating and separating them according to intent, 3) writing collated responses into a Likert scale format and separating them into two forms, Form S (science) and Form M (math), and 4) passing these forms out to the students with respect to their specific course setting and asking them to respond to the collated items. Sixteen items reflected students’ views and nine items reflected teachers’ views. The items representing student and teacher reactions were the same on both forms. The only difference between the science and math forms was the examples given by each teacher to help their respective students better understand their statements. Table 1 shows an example of this difference.

Results

A significant standardized regression coefficient (.235**) existed between the risk inclination of female students the accuracy of their selection. The actual scores of female 8th grade science students was significantly predicted by their risk inclination, $F_{(1, 135)} = 7.917$, $p < .01$. The Adjusted R Squared value was .048. This indicates that 4.8% of the variance of the actual score was explained by the female risk inclination. A regression equation for females students was constructed as FSscore = 46.844 + 3.838 (risk inclination). The ‘FSscore’ represents the female students’ final score. The constants 46.844 and 3.838 were the unstandardized coefficients.

The significant correlations of risk inclination and answer accuracy between genders represent the degree of risk inclination students were willing to make during answer selection. The above formula which used RII during data analysis revealed measured confidence in the form of risk inclination. Among female students, such risk inclination could predict to a certain degree the accuracy of their final score. However, no such evidence was found among male students.

A Multiple-regression was conducted on 45 female cases. The combination of two variables: 1) quiz scores and 2) Risk Inclination Index ratings (RII) significantly predicted the student's semester grade, $F_{(2, 42)} = 64.48$, $p<.000$. The adjusted R squared value of .75 indicated that 75% of the variance in the semester grade was explained by this model.

Of the 85 students, analysis was conducted on 30 cases using Form M (mathematics) representing the mathematics students and on 55 cases using Form S (science) representing the

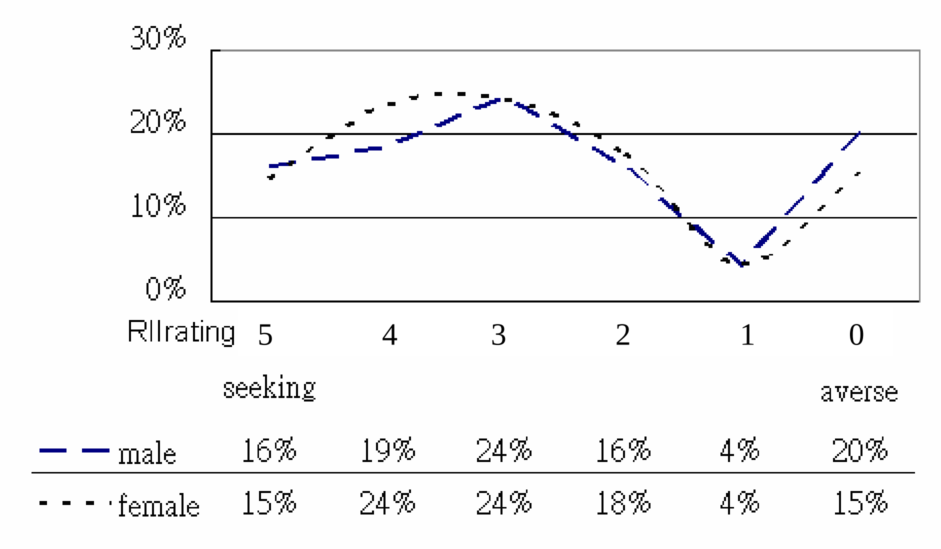


FIGURE 2. *Risk inclination comparison between genders*

science students. On Form M, eleven of the twenty-five statements attained a corrected item-total correlation of 0.40 and above. The remaining fourteen items were omitted. After the eight items were omitted, the Alpha for Form M increased from 0.83 to 0.90

Reliability analysis of the items from Form S produced an Alpha of 0.89. On this form, seventeen of the twenty-five statements attained a corrected item-total correlation of 0.40 and above. After the eight items were omitted, the Alpha for Form S increased from 0.89 to 0.93.

A risk inclination percentage analysis was done among the confidence wagering male and female students from mathematics and science classes. Figure 2 shows female students appear to be more risk seeking in predicting how accurate their answers are and less prone to risk aversion than their male counterparts.

Conclusion

Using confidence wagering among Taiwanese mathematics and science students showed the risk predictions of females to be more accurate than males. Using the Risk Inclination Index (RII) revealed female students to be more risk seeking and less risk adverse in mathematics and science classes than their male counter parts. Over 80% of the students stated that they favored this method of testing. The main reason for this favorable view was the ability to control the points they could earn on test items. Other reasons are as follows: Confidence wagering lets me and the teacher know how much I know, how much I really don’t know, how confident I am in my knowledge, where I need to change and gives me control over how much I can earn. Correctness increases confidence, and incorrectness lowers it.

Educational importance

The empirical data of this study have shown that the use of confidence wagering during testing appears to be a valid and reliable way of measuring to what extent Taiwanese mathematics and science students were willing to show differentiation toward their answer responses to test items and how such differentiation could predict their test answer accuracy. The confidence wagering structure could easily be modified to assess student level domain knowledge and level of reflective confidence in other courses and contexts and may be of assistance to U.S. educators and other researchers interested in studying the confidence accuracy of students toward their reflective knowledge during testing.

REFERENCE LIST

Bar-Hillel, M., Budescu, D., & Attali, Y. (2005). Scoring and keying multiple-choice tests: a case in irrationality. Mind & Society, 4, 3-12.

Borgmeier, C., & Horner, P. H. (2006). An evaluation of the predictive validity of confidence ratings in identifying functional behavioral assessment hypothesis statements. Journal of Positive Behavior Interventions, 8(2), 100-105.

Brewer, W. F., & Sampaio, C. (2006). Processes leading to confidence and accuracy in sentence recognition: A metamemory approach. Memory, 14(5), 540-552.

Brewer, W. F., Sampaio, C., & Barlow, M. R. (2005). Confidence and accuracy in the recall of deceptive and non-deceptive sentences. Journal of Memory and Language, 52, 618-627.

Cowley, E. (2004). Recognition confidence, recognition accuracy and choice. Journal of Business Research, 57, 641-646.

Holt, H. (2005). E-learning and student support in global environment and development studies. In L. Wong (Ed.), Globalization and Education for Sustainable Development Sustaining the Future (pp. 126-129). Nagoya, Japan: UNESCO.

Isaacson, R. M., & Fujita, F. (2006). Metacognitive knowledge monitoring and self-regulated learning: academic success and reflections on learning. . Journal of the Scholarship of Teaching and Learning, 6(1), 39-55.

Izaute, M., & Bacon, E. (2006). Effects of the amnesic drug lorazepam on complete and partial information retrieval and monitoring accuracy. Psychopharmacology, 188(4), 472-481.