CONFIDENCE WAGERING DURING MATHEMATICS AND SCIENCE TESTING

This proposal presents the results of a case study involving five 8th grade Taiwanese classes, two mathematics and three science classes. These classes used a new method of testing called confidence wagering. This paper advocates the position that confidence wagering can predict the accuracy of a student's test answer selection during among-subject assessments. Quantitative analysis of data using the Risk Inclination Model (Jack, Hung, Liu, & Chiu, 2009) revealed that female students were more prone to taking risks when making confidence wagering predictions and less prone toward risk aversion as compared to their male counterparts. Qualitative analysis of student comments revealed a positive acceptance of confidence wagering as a good way to self-regulate the point value of selected answers, to assist the teacher and student in refining which areas need more instruction and practice, and to reveal where the student's confidence is incorrectly placed.

Brady Michael Jack, National Kaohsiung Normal University, Kaohsiung, Taiwan R.O.C.
Chia-Ju Liu, Ph.D., National Kaohsiung Normal University, Kaohsiung, Taiwan R.O.C.
Hoan-Lin Chiu, Ph.D. National Kaohsiung Normal University, Kaohsiung, Taiwan R.O.C.
James A. Shymansky, Ph.D. University of Missouri at St. Louis, MO. U.S.A.

Confidence is the key to the quality and extent to which acquired knowledge is acted upon within a knowledge-based community. Being confident in what one knows allows a person to take initiative and be creative. The strength of such confidence is directly correlated to judgments made during times of decision (Cowley, 2004). Of debate among many is the degree to which 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 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 perspectives are studies which view confidence as an important factor in predicting performance accuracy. 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 (W.F. 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

Brewer and Sampaio (2006) proposed that as the time between a memory recognition task and the time of confidence judgment regarding that task increased, the accuracy of memory was negatively affected. This study uses the utility of wager to capture the level of confidence at the moment of answer selection. Each wager measures the student's inclination toward risk or aversion. In order to mathematically calculate such inclination, a Risk Inclination Model (RIM) (Jack, Hung et al., 2009) 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) was 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

Purpose

This paper introduces a new method of analysis called confidence wagering which evaluates the confidence and accuracy of Taiwanese middle-school science and mathematics students at the moment of answer selection during multiple-choice tests. The results from this exploratory study will provide important information regarding how confidence assessments affect and reflect such accuracy. Research questions: (1) Can confidence wagering 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 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. Question and answer selection content were controlled by the teachers and reflected what had been taught each week.

Testing

The design format of each test used Confidence Wagering (CW) (Jack, Liu, Chiu, & Shymansky, 2009). This structure required the student to predict how many points he/she felt his/her answer was worth. This method of allowing students to make honest confidence judgments about their answers was used by James A. Shymansky during years of teaching. Shymansky (2007)

TABLE 1. Collated teachers' statement with two different examples

<table>
<tr><th>IN</th><th colspan="2">Likert scale statement from teachers</th></tr>
<tr><td>11.</td><td colspan="2">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.</td></tr>
<tr><td></td><td>Example used on Form M</td><td>Example used on Form S</td></tr>
<tr><td></td><td>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?
$\begin{cases} 3x+5y=370 \\ 5x+4y=400 \end{cases}$</td><td>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</td></tr>
</table>

IN: Item Number; NT: New Taiwan dollars (roughly $32NT per $1US dollar) discovered that when students were allowed to make such judgments they performed better than they did on traditional teacher-determined score-valued tests. The CW design layout had four basic components: 1) test question/statement, 2) multiple choice answer selections, 3) a box where the student wrote 5, 10, or 15 indicating the point value he/she thought his/her answer was worth, and 4) a circle area where the teacher wrote the student's predicted answer value if the answer was correct or zero if the answer was wrong. At the end of correction, the sum of these circled values equaled the final quiz grade.

Evaluation

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 these two forms was the clarifying examples given by each teacher in order to help their respective students better understand their statements. Table 1 above shows an example of this difference.

Results

Quantitative analysis

A significant standardized regression coefficient (.235**) existed between the risk inclination of female students and 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 analysis 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.

Qualitative/quantitative analysis

Of the 85 students, analysis was conducted on 30 cases using Form M representing the mathematics students and on 55 cases using Form S representing the 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. Table 2 shows the corrected factor loadings of responses given by mathematics students. 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

TABLE 2. Corrected factor loadings of Form M (mathematics)

Item	Pos-Neg. Wording	Item-Total Correlations	Factor Loadings #1	Factor Loadings #2
S 01	Positive	.61	.56	
S 03	Positive	.75	.78	
S 07	Positive	.65	.65	
S 08	Positive	.59	.69	
S 09	Positive	.65	.61	
S 19	Positive	.63	.65	
S 22	Positive	.80	.87	
S 23	Positive	.66	.58	
S 25	Positive	.46	.48	
S 02	Positive	.64		.93
S 10	Positive	.65		.62
		Alpha:	.90	

S: Statement

above. The remaining eight items were omitted. After the eight items were omitted, the Alpha for Form S increased from 0.89 to 0.93 (see Table 3).

TABLE 3. Corrected factor loadings of Form S (science)

Item	Pos-Neg. Wording	Item-Total Correlations	Factor Loadings #1	Factor Loadings #2
S 01	Positive	.64	.60	
S 02	Positive	.68	.76	
S 03	Positive	.69	.60	
S 05	Positive	.58	.52	
S 06	Positive	.78	.70	
S 07	Positive	.75	.82	
S 08	Positive	.64	.64	
S 09	Positive	.79	.67	
S 10	Positive	.74	.71	
S 11	Positive	.58	.70	
S 17	Positive	.62	.47	
S 19	Positive	.68		.59
S 21	Positive	.52		.67
S 22	Positive	.72		.71
S 23	Positive	.59		.63
S 24	Positive	.55		.61
S 25	Negative	.44		.56
		Alpha:	.93	

S: Statement

A risk inclination percentage analysis was done among the confidence wagering male and female students from mathematics and science classes. Figure 1 show 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.

Discussion

The use of confidence wagering among Taiwanese mathematics and science students has produced favorable results. Data show 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 counterparts. Over 80%

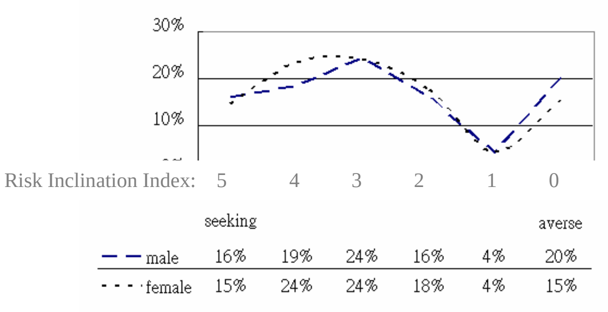

Figure 1. Risk inclination comparison between genders.

of the students stated that they favored this method of testing. The main reason for this favorable view was that the students indicated that they enjoyed having the ability to control the points they could earn. Other reasons they liked confidence wagering 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 it gives me control over how many points I can earn. Correctness increases confidence, and incorrectness lowers it.

From a local school classroom context, the empirical data of this study have shown the use of confidence wagering during testing among the participants of this study 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 accuracy. Researchers of future studies could consider using confidence wagering structure to assess the student's confidence and accuracy level of reflective confidence in other courses.

References

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.

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.

Jack, B. M., Hung, K. M., Liu, C. J., & Chiu, H. L. (2009, April 13-17). *Utilitarian Model of Confidence Testing for Knowledge-based Societies.* Paper presented at the American Education Research Association (AERA), San Diego, CA.

Jack, B. M., Liu, C. J., Chiu, H. L., & Shymansky, J. A. (2009, April 13-17). *Confidence Testing for Knowledge-based Global Communities.* Paper presented at the American Education Research Association (AERA), San Diego, CA.

Liddell, M. J., & Davidson, S. K. (2004). Student attitudes and their academic performance: is there any relationship? *Medical Teacher, 26*(1), 52-56.

Lindsay, D. S., Nilsen, E., & Read, J. D. (2000). Witnessing-condition heterogeneity and witnesses' versus investigators' confidence in the accuracy of witnesses' identification decisions. *Law and Human Behavior*, *24*(6), 685-697.

Patterson, M. L., Foster, J. L., & Bellmer, C. D. (2001). Another look at accuracy and confidence in social judgments. *Journal of Nonverbal Behavior*, *25*(3), 207-219.

Perfect, T. J., Hollins, T. S., & Hunt, A. L. R. (2000). Practice and feedback effects on the confidence accuracy relation in eyewitness memory. *Memory, 8*(4), 235-244.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation*, *11*(10), 1-13.

Shymansky, J. A. (2007). Making the most of assessments. Kaohsiung & Chang Hua, Taiwan: Faculty and graduate student workshop given at the National Taiwan Normal University and National Chang Hua University of Education.