

Self-monitoring: Confidence, academic achievement and gender differences in physics

Manjula Devi Sharma

Faculty of Science
The University of Sydney, Australia
manjula.sharma@sydney.edu.au

James Bewes

School of Physics
The University of Sydney, Australia

Abstract

Metacognition is the higher-order monitoring that deals with a person's regulation of thought processes and governs learning strategies and understanding in an instructional setting. The ability to appraise and judge the quality of one's own cognitive work in the course of doing it is self-monitoring. If the work needs to be done within a short time frame then rapid assessments of how confident a person is that their answer is accurate provide means of self-monitoring. The aim of this study was twofold, first, to investigate physics students' self-monitoring, and second, to investigate gender differences in self-monitoring. The study was carried out with 490 first year university physics students who were administered an online mechanics quiz that contributed to assignment marks. Results indicate that classes with higher academic achievement exhibit better self-monitoring capability. Gender differences were found on confidence but not on self-monitoring. Theoretical models of self-monitoring are explored, as are implications for teaching and learning.

Keywords

Self monitoring, physics education, gender differences in physics.

Introduction

Metacognition has several facets, all of which deal with a person's regulation of thought processes. The focus of this study is a primary component of metacognition known as self-monitoring. Self-monitoring is defined as the ability to watch, check, appraise and judge the quality of one's own cognitive work in the course of doing it (Kleitman & Stankov, 2001). The capability to monitor cognitive processes effectively can lead to improved learning strategies, as well as a superior ability to recognise and rectify one's deficiencies in knowledge (Karabenick, 1996). For this reason, self-monitoring is critical in achieving optimal learning levels.

While literature concerning self-monitoring is quite substantial, most of the research has been conducted in tightly controlled experimental settings and not in authentic educational settings. When self-monitoring is studied, it is often to trial a self-monitoring strategy in an intervention rather than to examine how self-monitoring exhibits itself. For example, a study of first year university students' self-monitoring of reading-comprehension of physics text and laboratory manuals, concluded that the self-monitoring exercise improved student learning (Koch, 2001).

Similar results were found using self-questioning during notetaking from scientific text as a self-monitoring strategy with fifth and sixth graders (Laidlaw, Skok & McLaughlin, 1993). This study is an attempt to examine whether self-monitoring as operationalised in educational psychology is applicable to physics education. If we are able to measure self-monitoring capability simply and effectively, then we may provide the educator with another tool for gauging student learning. Furthermore, as gender issues are important in physics (Kost, Pollock & Finkelstein, 2009, McCullough, 2004; Hazari, Tai & Sadler, 2007) we decided to investigate this feature as well.

The confidence paradigm

Self-monitoring has been operationalised using the confidence paradigm (Kleitman & Stankov, 2001; Pallier, 2003). The confidence paradigm requires participants to report how confident they are in the accuracy of their responses as they progress through a test. *Bias* is obtained by subtracting the mean test score (accuracy) from the mean confidence for each participant.

$$\text{Bias} = \text{Confidence (\%)} - \text{Accuracy (\%)}$$

The term *calibration* refers to how closely a person's reported confidence in the accuracy of their answers corresponds to their test scores. A person who gets every question correct on a test and reports a mean confidence of 100%, has a bias score of zero and is considered perfectly calibrated.

The advantages of using bias in this study are twofold. First, the creation of a bias score for each participant provides a simple and transparent way of measuring self-monitoring. *Miscalibration* refers to over- or under-confidence and is represented by a bias score that is not zero. A person with a positive bias score is over-confident in self-monitoring and a negative bias indicates under-confidence. The second advantage is that it is established in the metacognition literature and has a theoretical as well as an empirical foundation (Kleitman & Stankov, 2001; Pallier, 2003; Juslin, 1994; Pallier, Wilkinson, Danthiir, Kleitman, Knezevic & Stankov, 2002).

Theoretical models

Two predominant theoretical explanations of the nature of self-monitoring are the *Ecological Approach* and the *Individual Differences Approach*.

The Ecological Approach

The Ecological Approach proposes that students' use a probabilistic mental model to determine their answer and how confident they are of their answer. Students draw on cues from the environment, taking into account the relative frequency of events to solve mental problems (Gigerenzer, Hoffrage & Kleinbolting, 1991). If students have frequently observed an event and have a viable interpretation that explains that event, then they are more confident that they have selected the correct answer. If students have predominantly chosen the correct answer, then the question is said to be *representative*. In other words students have valid explanations and interpretations and the correct answer represents students' understandings. When the correct answer does not represent student understandings, more students select incorrect answers and the question is said to be *non-representative*. The latter category includes questions that elicit alternate conceptions (misconceptions) or are counter-intuitive. The *cue validity* of the question does not correspond with what the students understand as the *real* state of affairs in the natural environment, its *ecological validity* (Pallier, 2003). In physics education, it is reasonable to explore such mismatches arising from alternative conceptions. Within the framework of this model, there is a discrepancy between ecological and cue validities in the students probabilistic mental models that leads to miscalibration. The Ecological Approach predicts that calibration and students' self-monitoring capability should be severely impaired by non-representative questions.

Individual Differences Approach

The roots of the Individual Differences Approach are in the field of differential psychology and the crux of the approach is that the accuracy of self-monitoring is an independent metacognitive trait (Pallier et al. 2002). The Individual Differences Approach proposes first that students are inclined to report a consistent confidence level; and second that the confidence level is *relatively* decoupled from accuracy. The prediction is that students would show little variation in their reporting of confidence values. Even if a distribution of bias is heavily skewed towards over-confidence, there should be a percentage of students that exhibit under-confidence. Another prediction is that the percentage of students exhibiting under-confidence would depend on students' prior experience with physics learning. This study explores the influence of high school backgrounds on confidence and self-monitoring, an issue emerging as an important factor influencing attitudes and beliefs (Gray, Adams, Wieman & Perkins, 2008; Gire, Jones & Price, 2009).

Gender differences and the stereotype threat

There is a common perception that males outperform females in mathematics and the sciences (Halpern & LaMay, 2000; Mullis, Martin, Fierros, Goldberg & Stemler, 2000). Of equal concern is the gender ratio in the sciences, with the greatest observed gender inequality in physics (Ivie & Stowe, 2000). There is extensive literature regarding gender differences in the physics classroom, (Kost, Pollock & Finkelstein, 2009; Hazari, Tai & Sadler, 2007; Seymour, 1995) however none of those studies address the role that self-monitoring plays in learning. This is surprising, because studies have shown that gender differences *do* exist in confidence, and that metacognition and self-monitoring mediate learning (Pallier, 2003).

Furthermore, the existence of an inhibitor of performance termed *stereotype threat* has been postulated (Steele, 1997). This phenomenon is summarized as: 'the added demands felt by members of stereotyped groups... in situations where their behavior can confirm... that their group lacks a valued ability' (Aronson, Lustina, Good, Keough, Steele & Brown, 1999). Two key conditions in such studies are that individuals are aware that their group tends to perform poorly and that the test is diagnostic in nature. Studies found that women aware of the stereotype threat and told that they were doing diagnostic mathematics tests performed worse than both men and women not aware of the stereotype threat doing the same test but told that the test was non-diagnostic in nature (Martens, Johns, Greenberg & Schimel, 2006).

Purposes of this study

A broad objective was to examine how the predictions of the two theoretical models of self-monitoring emerge in first-year university physics students. Each model provides different reasons for observed calibration or miscalibration (Pallier, 2003). The models are complementary and both models can be used for understanding the situation examined in this study. Which features emerge can shed light on how students monitor their learning during physics tasks. As the intent was to investigate if self-monitoring as understood in educational psychology emerges in university physics education, self-monitoring data was collected only once early in first semester of studies and comparisons made with high school and first semester academic achievement. In future, one can investigate how self-monitoring changes during the course of physics studies. For now we are interested in whether accuracy and self-reported confidence gathered during a test are meaningful. This study investigates a way of measuring self-monitoring.

Our specific aim was to address the following research questions on the sampled students' self-monitoring as operationalised by the confidence paradigm:

- Can we identify misconceptions through non-representative questions?
- What are the trends between the students' self-monitoring and their academic achievement in physics?
- Are there gender differences in physics students' self-monitoring?

Method

Participants and procedure

The participants were from three first year undergraduate physics classes, Fundamentals, Regular and Advanced, at a metropolitan university in Australia. The Fundamentals class was for students who had not done senior high school physics, the Regular for students who had successfully completed senior high school physics, and Advanced for students who had performed very well at the senior high school level *and* had also successfully completed high school physics.

In their lectures, the students were instructed to complete a mechanics quiz online for assessment. This occurred in week 2 of first semester for the Regular and Advanced classes and in week 4 for the Fundamentals class. Students then had one week to log in and complete the mechanics quiz. They could complete the task at their home computer or in a computer centre at the university. As a student moved through the test, responses and the time taken were recorded in a MySQL database. Following completion of the test, students were given automated feedback on their performance. All students participated with informed consent and a total of 490 students were included in the study.

Materials

Student understanding of 10 topics relating to Newton's first and second laws of motion were measured using a 26-question multiple-choice quiz. The quiz contained questions from two conceptual tests: 22 questions from the Force Concept Inventory and three from the Force and Motion Conceptual Evaluation (Hestenes, Wells & Swackhamer, 1992; Thornton & Sokoloff, 1998). Questions on the quiz are both established and validated (Coletta, Phillips & Steinert, 2007; Thornton, Kuhl, Cummings & Marx, 2009). One question was created by the authors. The questions deliberately target qualitative, conceptual knowledge and do not require calculation. Students were asked to rate how confident they were that their response(s) were correct on a seven-point Likert scale, with 1 representing *uncertain* and 7 representing *certain*.

Analysis

Accuracy was the percentage of correct answers for each student for the 26 questions. Confidence values were computed in line with conventions set out in the confidence paradigm (Kleitman & Stankov, 2001). Uncertain was assigned the percentage value of selecting the correct answer by chance, certain was assigned 100% confident and subsequent Likert scale values were equally divided between those values. Reported confidence on all questions was averaged for each student. In accordance with the confidence paradigm, a bias score for each student was derived by subtracting the accuracy from the reported confidence.

Two measures of senior high school academic achievement were used to examine the students' academic background: senior high school physics marks and the Universities Admissions Index (UAI) – a ranking of all students who completed a state-wide examination at the end of their schooling. Not all measures were available for all students and n , the number of students, is provided for the relevant statistics. All data were checked for normality using the K-S test and in one case non-parametric statistics is reported.

Results

To establish that the mechanics quiz and the end of semester examination marks provided similar measures of physics knowledge for the students in this study, performance on the quiz was correlated with end of semester physics examination marks, see Table I. Pearson's correlations (Table I) are sizeable and statistically significant ($p < 0.01$) for all three classes, indicating that there is internal consistency between the two measures.

Table I: Correlations between accuracy on the mechanics quiz and end of semester physics examination marks

Class	n	Pearson's correlations
Fundamentals	140	0.442**
Regular	227	0.450**
Advanced	123	0.555**

** Statistically significant at $p=0.01$ (2-tailed)

Misconceptions as non-representative questions

Accuracy on the quiz was compared with reported confidence for each of the 10 topics tested. Figure 1 shows a plot of the mean accuracy against mean confidence. Each point represents one topic. The closer a point is to the ideal calibration line, the better the match between the cue and ecological validities for that topic. As we go from Fundamentals to Regular to Advanced, the match between the cue and ecological validities increases, suggesting students with better calibration. For one specific topic we find a good match between the cue and ecological validities for all three streams, demonstrating that it is possible to have well calibrated student learning.

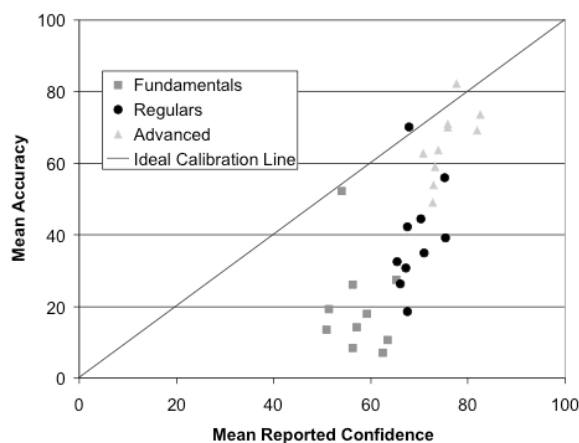


Figure 1: Mean accuracy versus mean reported confidence for the different classes across the ten topics. The ideal calibration line represents a cohort which has a perfect match between accuracy and reported confidence.

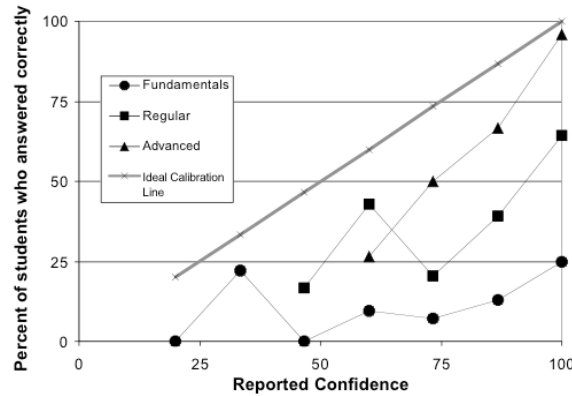


Figure 2: The percentage of students who were correct for each confidence level. The ideal calibration line represents a cohort which has a perfect match between accuracy and reported confidence. Since only the percentage of students who were correct for each confidence level is plotted, the number of students in each bin is not equal. Bins with less than 10 students are not plotted.

Another perspective is provided by investigating calibration for a single question as shown in Figure 2. The ideal calibration line represents a cohort which has a perfect match between accuracy and confidence. If students are overconfident, the calibration curve falls below the ideal calibration line. That is, the students anticipate being more accurate than they are. It is apparent that Fundamentals students exhibit the greatest overconfidence in their responses, while the Advanced students demonstrate the least. Since only the percentage of students who were correct for each confidence level is plotted, the number of students in each bin is not equal. Bins with less than 10 students are not plotted.

Trends in self-monitoring and academic achievement

Accuracy and confidence

The means and standard deviations for accuracy, and reported confidence for each class are presented in Table II. Little variation in the reporting of confidence was observed. The spread of reported confidence values is small and similar across classes. In contrast, variation in accuracy differs markedly across classes with the standard deviation in the Advanced being nearly double that of the Fundamentals. Further, the highest mean values of both reported confidence and accuracy were reported in the Advanced class, followed by the Regular and Fundamental physics classes respectively. This was not unexpected as the Advanced students are likely to have the greatest prior experience in physics.

Table II: Means and standard errors of the means of reported confidence, accuracy and bias. Standard deviations are provided in parentheses

Class	n	Mean reported confidence in %	Mean accuracy in %	Mean bias in %
Fundamentals	140	58.0±1.5 (18)	19.8±1.3 (15)	38.2±1.7 (20)
Regular	227	69.4±1.2 (17)	39.5±1.7 (25)	29.8±1.6 (24)
Advanced	123	76.3±1.5 (16)	66.1±2.5 (28)	10.3±2.1 (23)
Total	490	67.9±0.8 (19)	40.6±1.3 (29)	27.3±1.4 (25)

The relationship between self-monitoring and achievement in physics

Is self-monitoring capability, as measured through bias scores, linked with physics performance?

Table II reveals that the classes with higher levels of physics experience have on average, lower bias scores, indicating that they are better calibrated. A comparison of the variances in the bias for students in the three classes using one-way ANOVA showed a significant difference ($F=11.7$, $df=2$, $p<0.05$). Correlations between measures of academic achievement and bias using Spearman's rho are shown in Table III. All correlations between achievement measures and bias are negative, indicating that students with higher levels of physics experience also exhibit better self-monitoring capability. Using Cohen's (1988) interpretation of correlation coefficients, there is a medium to high correlation between both high school physics mark and bias, and end of semester physics examination mark and bias for the Advanced and Regular classes.

Table III: Correlations between bias and physics marks

Class	Spearman's rho correlations between bias and physics marks in	
	high school	end of semester examination
Fundamentals	Not available	-0.17 (n=140)
Regular	-0.19*(n=148 [#])	-0.31**(n=227)
Advanced	-0.43**(n=111 [#])	-0.48**(n=123)

* Statistically significant at $p=0.05$ (2-tailed)

** Statistically significant at $p=0.01$ (2-tailed)

[#] Only one measure - marks for the state wide examination from the previous year were included. Students with marks from other states, international and from other years were excluded.

Trends in bias across classes

To further examine trends in bias across the physics classes, distributions of bias scores were plotted, see Figure 3. The *ideal* bias score of zero represents good calibration. We note three features. Firstly, all distributions were approximately normally distributed and the peak of the distribution tends towards a bias score of zero as we go from the Fundamentals, Regular to Advanced classes. Secondly, there is a general trend of over-confidence, with a combined mean bias score for all three classes of +37. Thirdly, some students do exhibit under-confidence; of the 490 participants, 64 had a negative bias score. That number represents 13% of total participants in this study, which is a number consistent with previous studies in the confidence paradigm (Pallier, 2003). When analyzed by physics class, the number of students who demonstrated under-confidence increased from Fundamentals, Regular to Advanced.

Gender differences in self-monitoring

Gender differences before the quiz

UAI and senior high school physics marks for men and women were compared to see if any differences were present before testing (Table IV). Using t -tests, males had no better performance than females on these measures. In fact, females in the Regular physics class had higher UAIs than males ($t=2.175$, $df=150$, $p=0.031$). Gender differences in senior high school mathematics performance were also investigated as a possible confounding factor, however no statistically significant differences were found. Students in the same class, regardless of gender, had equivalent prior academic achievement on average.

Table IV: Means and standard errors of the means of UAI and high school physics marks. Standard deviations are provided in parentheses.

	University Admissions Index, UAI		High school physics mark	
	n	Mean in %	n	Mean in %
<i>Fundamentals</i>				
Female	64	93.7±0.7 (5.2)	-	-
Male	21	91.1± 1.8 (7.9)	-	-
<i>Regular</i>				
Female	54	92.1±0.7 (5.3)*	51	82.9±0.7 (5.2)
Male	98	90.0±0.6 (6.0)*	97	83.3±0.6 (5.6)
<i>Advanced</i>				
Female	23	98.0±0.4 (1.8)	24	90.5±0.8 (3.9)
Male	81	97.3± 0.3 (2.3)	87	91.3±0.3 (2.6)

* Statistically significant at $p=0.05$ (2-tailed t -test)

Gender differences on the quiz

The means for accuracy, reported confidence and bias on the mechanics quiz across genders are presented in Figure 4. In all 3 physics classes, females scored lower on both accuracy and confidence. The mean differences in accuracy and confidence were significant using t -tests ($p<0.05$) and as evident from the standard errors of the means shown in Figure 4, in all but one case. Only in the Regular class, there was no statistically significant difference in mean accuracy scores. Interestingly, mean bias scores were *not* statistically significantly different between genders, again confirmed using t -tests and evident from standard errors of the means shown in Figure 4.

Gender differences after the quiz

The means on the end of semester physics examination for men and women were compared using t -tests to see if any differences were present after testing. In all classes, males had no better performance than females on this measure.

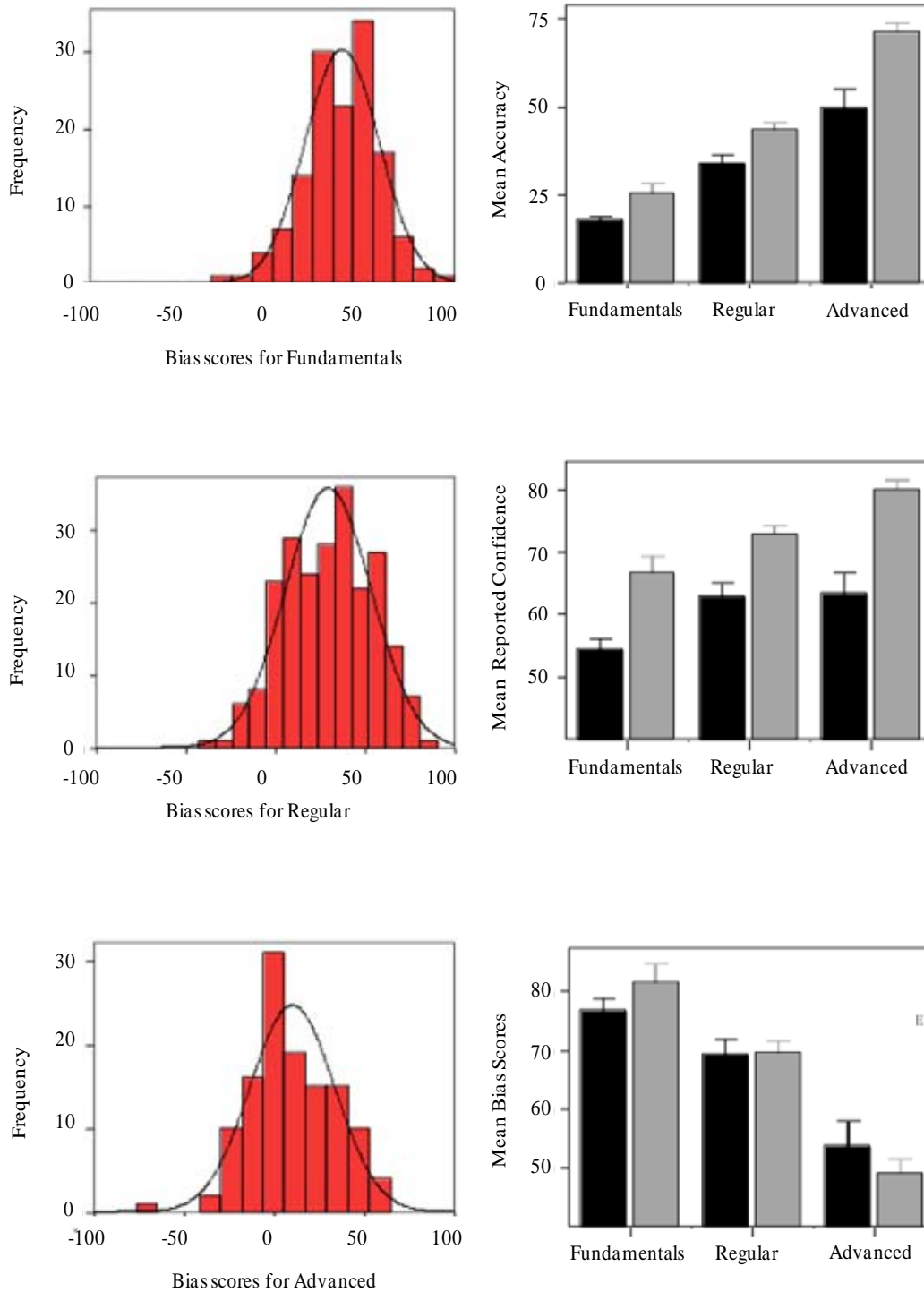


FIG. 3. Histograms of bias scores by class. The percentage of underconfident students with negative bias scores increases going from Fundamentals, Regular to Advanced.

FIG. 4. Gender differences in mean reported confidence, accuracy and bias scores by class. Female Male Error Bars show Mean +/- SE

Discussion

In order to extend current theoretical models of self-monitoring, we have applied the *confidence paradigm* to an authentic tertiary education setting. We found that trends predicted by the models do indeed emerge in our data and allow for meaningful interpretations within this educational context. The sampled physics students were, in general, overconfident in their self-monitoring of performance on the mechanics quiz, as predicted by the *Ecological Approach* (Gigerenzer, Hoffrage & Kleinbolting, 1991). Furthermore, a statistically significant correlation was found between students' calibration and physics academic achievement (Kleitman & Stankov, 2001). Students who were better calibrated generally ranked higher in measures of physics academic achievement. Finally, little spread was found in participant's judgements of confidence, as predicted by the *Individual Differences Approach* (Kleitman, 2003) and a significant percentage of students actually expressed under-confidence in their self-monitoring.

The implications of our finding that facets of different theoretical models emerge are twofold for physics education. The implications in themselves are not novel, but provide different perspectives on current understandings. Firstly, the non-representative nature of many concepts in physics, lead to a mismatch between 'cue' and 'environmental' validities. A prime example is that our everyday experience of forces is counter-intuitive to the currently accepted Newtonian interpretation. Studies have documented that students are confused with the mismatch (Muller, Sharma & Reimann, 2008) while others have shown that students disagree with scientists (Gray, Adams, Wieman & Perkins, 2008). The self reported confidence measured in our study tentatively indicates how embedded the mismatch is.

If online quizzes such as that employed in our study are used than instructors would have a quick and easy indication of which concepts are non-representative, and could use appropriate measure to address the mismatch. One would need to consider students' prior conceptions and seek to actively counteract them, which is the basis of conceptual change research and some educational interventions. At the same time, how students' confidence is affected as the mismatch is realized, will need to be considered; an area that has not been given much attention. We do note that in our study, for one topic there is a good match between cue and ecological validities for students in all three classes, and this is what we should aim for across all topics. The extensive literature on misconceptions and alternative conceptions also acknowledge the need to address alternative conceptions but rarely, if ever, acknowledges self-monitoring as done in this study.

The second implication is that a student-centered approach based on confidence may produce better learning outcomes. A central tenet of the Individual Differences Approach is that each individual has a unique metacognitive trait (Pallier et al., 2002). Accordingly, this trait predisposes one to report consistent confidence levels, which subsequently do not vary to the same extent as accuracy. The ensuing mismatch results in miscalibration. Consequently, teaching and learning strategies could be specifically designed for students with varying levels of self-monitoring determined through the use of the confidence paradigm. At the level of individual students in large university classes, such techniques may not be feasible. However, clear trends in self-monitoring maybe identifiable for particular classes. For example, self-monitoring in a class of prospective primary school teachers could be very different to self-monitoring amongst physics majors. A difference in students' views has been noted when comparing engineering students with physics majors (Gire, Jones & Price, 2009) and there is no doubt that more research needs to be done. Students possessing poor self-monitoring capabilities could be given additional scaffolding in terms of both techniques for self-monitoring and physics knowledge as pre-empted by Karabenick (1996).

An interesting finding of this study was the significant trend in the correlations between measures of academic achievement and calibration. No causal conclusion can be made from this preliminary study, but it is apparent that higher-achieving students also exhibit better self-monitoring capability. On one hand, maybe good calibration is a consequence of a thorough understanding of physics. On the other hand, perhaps better-calibrated people understand physics concepts more readily. The notion of self-monitoring in physics education has been implemented sporadically (Koch, 2001; Laidlaw, Skok & McLaughlin, 1993), and needs to be further researched.

Gender differences in self-monitoring

To demonstrate the applicability of the confidence paradigm, the acknowledged problem of gender inequalities in physics performance was examined. Prior to, and after testing, there were no apparent differences in achievement scores between males and females. In the case of the Regular physics class, females actually entered with *higher* UAIs. However, in all but one case, males exhibited higher levels of accuracy and confidence ratings on the mechanics quiz than their female peers. The resulting mean bias scores were the same suggesting that the women somehow acknowledged that they were not as accurate on the mechanics quiz. We present two possible explanations for this finding.

First, there may be gender differences in conceptual understanding of mechanics or the nature of the test instrument. This explanation is based on the assumption that the mechanics quiz is different to measures such as the UAI and end of semester examinations on which gender differences do not emerge. Studies have shown a correlation between high school results and academic performance at university (Dobson, 1999). There is evidence that the questions used in the mechanics quiz are biased towards men. McCullough (2004) attributes gender differences on the Force Concept Inventory to contextual differences in the questions, specifically that the questions deal with subject matter oriented towards, and of interest to males. Other pertinent factors favoring men, such as the type of questions, multiple choice or extended responses, and difference in levels of everyday experience with some physics topics have also been found (Hazari, Tai & Sadler, 2007). There is also the possibility that the test questions involve visualization and spatial reasoning known to favor men (Pallier, 2003) or that the type of complex reasoning raises the issue of risk aversiveness that women are more prone to (Wittmann, 2005). How each of these explanations relates to self-monitoring and why they do not emerge in examinations is for further study.

The second explanation relates to a phenomenon known as *stereotype threat*. According to Steele (1997), stereotype threat surfaces when a group which possesses a stereotyped quality, trait or characteristic performs worse than control groups on a diagnostic test. It is possible that the mechanics quiz delivered in the first few weeks of semester captured the conditions that implicitly elicit stereotype threat. The transition from school to university physics exposes girls to environments with more boys and male academics in their first few weeks of semester. The gender ratio imbalance in physics is amongst the largest from the science and mathematics subjects at this University. For girls from single-sex girls schools this difference is even more pronounced. McCullough (2004) in attributing gender differences on the Force Concept Inventory also acknowledges that stereotype threat may play a mediating role. In a study investigating why women leave science early in their undergraduate years, Seymour (1995) notes that unintentional behaviors can result in the following.

Young women tend to lose confidence in their ability to “do science,” regardless of how well they are actually doing, when: they have insufficient independence in their learning styles, decision making, and judgments about their own abilities; to survive denial of motivational support and performance reassurance by faculty, the refusal of male peers to acknowledge that they belong in science. Women who persist: enter with sufficient independence to adjust quickly to the more impersonal pedagogy; bond to the major through intrinsic interest and a strong sense of career direction; and develop attitudes and

strategies (including alternative avenues of support), in order to neutralize the effects of male, peer hostility. (p. 470).

Aspects of curricula and pedagogy that are not amenable to women have been examined (Hazari, Tai & Sadler, 2007). The department where this study has been carried out has pastoral care, collaborative learning environments and assessment practices (Sharma, Mendez & O'Byrne, 2005; Sharma, Sefton, Cole, Whymark, Millar & Smith, 2005) that possibly sufficiently counterbalance negative features over the course of the semester. However in the first few weeks when the mechanics quiz was administered the effect of such strategies was not evident. There is no doubt that further studies are necessary to explore such assertions.

In this study, it is significant that no gender differences were observed in bias scores. Both genders were *equally* overconfident when it came to analysing their own work rather than males being generally more overconfident than females (Pallier, 2003). The observation that males and females did not differ in self-monitoring suggests that they are actually equally-performing groups; there must be additional variables that are reducing female performance on our measures.

Conclusion

In conclusion, self-monitoring provides a different perspective for physics education. This preliminary study demonstrates that variables, such as bias, allow for insights into student's monitoring of their knowledge and into gender issues in physics learning and teaching.

References

- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, 35(29).
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Coletta, V. P., Phillips, J. A., & Steinert, J. J. (2007). Interpreting force concept inventory scores: Normalized gain and SAT scores. *Phys. Rev. ST Phys. Educ. Res.* 3, 010106.
- Crawford, J., & Stankov, L. (1996). Age differences in the realism of confidence judgments: A calibration study using tests of fluid and crystallized intelligence. *Learning and Individual Differences*, 6(84).
- Dobson, I. (1999). Performance indicator in the measurement of institutional success in transition, Higher Education Series, Report 2b, Canberra.
- Gigerenzer, G. Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence, *Psychological Review*, 98, 506 (1991).
- Gire, E., Jones, B., & Price, E. (2009). Characterizing the epistemological development of physics majors. *Phys. Rev. ST Phys. Educ. Res.* 5, 010103.
- Gray, K. E., Adams, W. K., Wieman, C. E., & Perkins, K. K. (2008). Students know what physicists believe, but they don't agree: A study using the CLASS survey. *Phys. Rev. ST Phys. Educ. Res.* 4, 020106.
- Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review*, 12, 229.
- Hazari, Z., Tai, R. H., & Sadler, P. M. (2007). Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors. *Science Education*, 91, 847.
- Hestenes, D., Wells, M., & Swackhamer, G. A. (1992). Force Concept Inventory. *The Physics Teacher*, 30, 141.
- Ivie, R., & Stowe, K. (2000). *Women in Physics, 2000, AIP report R-430* (American Institute of Physics, College Park, MD, 2000).
- Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes*, 57, 226.

- Karabenick, S. A. (1996). Social influences on metacognition: effects of colearner questioning on comprehension monitoring. *Journal of Educational Psychology*, 88, 689.
- Kleitman, S. (2003). Bias score - An instance of systematic irrationality? Published Proceedings of the ICCS/ASCS-2003, Joint International Conference on Cognitive Science. Sydney, Australia 13 - 17 July 2003.
- Kleitman, S., & Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. *Journal of Applied Cognitive Psychology*, 15, 321.
- Koch, A. (2001). Training in metacognition and comprehension of physics texts. *Science Education*, 85, 758.
- Kost, L. E., Pollock, S. J., & Finkelstein, N. D. (2009). Characterizing the gender gap in introductory physics. *Phys. Rev. ST Phys. Educ. Res.* 5, 010101.
- Laidlaw, E. N., Skok, R. L., & McLaughlin, T. F. (1993). The effects of notetaking and self-questioning on quiz performance. *Science Education*, 77, 75.
- Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology*, 42(2), 236.
- McCullough, L. (2004). Gender, context, and physics assessment. *Journal of International Women's Studies*, 5(4), 20.
- Muller, D. A., Sharma, M. D., & Reimann, P. (2008). Raising cognitive load with linear multimedia to promote conceptual change. *Science Education*, 92(2), 278.
- Mullis, I. V. S., Martin, M. O., Fierros, E. G., Goldberg, A. L., & Stemler, S. E. (2000). Gender Differences in Achievement. IEA'S Third International Mathematics and Science Study (TIMSS) Boston College, TIMSS International Study Center, Chestnut Hill, MA.
- Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles*, 48, 265.
- Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., & Stankov, L. (2002). The role of individual differences in the accuracy of confidence judgments. *Journal of General Psychology*, 129, 257.
- Seymour, E. (1995). The loss of women from science, mathematics and engineering undergraduate majors: An explanatory account. *Science Education*, 79, 437.
- Sharma, M. D., Mendez, A., & O'Byrne, J. W. (2005). The relationship between attendance in student-centred physics tutorials and performance on university exams. *International Journal of Science Education*, 27(11), 1375.
- Sharma, M. D., Sefton, I. M., Cole, M., Whymark, A., Millar, R. M., & Smith, A. (2005). Effects of re-using a conceptual exam question in physics. *Research in Science Education*, 35, 447.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613.
- Thornton, R. K., Kuhl, D., Cummings, K., & Marx, J. (2009). Comparing the force and motion conceptual evaluation and the force concept inventory. *Phys. Rev. ST Phys. Educ. Res.* 5, 010105.
- Thornton, R. K., & Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: the force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics*, 66, 338.
- Wittmann, W. W. (2005). Group differences in intelligence and related measures. In *Handbook of Understanding and Measuring Intelligence* Eds. O. Wilhelm and R. W. Eagle, Ch 13, 223-239. Thousand Oaks, California: Sage Publications.

Acknowledgements

The authors acknowledge support from Dr Derek Muller and the Sydney University Physics Education Research (SUPER) group. They also appreciate the sincere efforts of the School of Physics staff who facilitated the project, and the participating students.

Copyright © 2011 Manjula Devi Sharma