Pertanika J. Soc. Sci. & Hum. 24 (S): 73 – 82 (2016)

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Students' Perceived Understanding and Competency in Probability Concepts in an e-Learning Environment: An Australian Experience

Zamalia, M. [1]* **and Porter, A.** [2]

[1]*Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*
[2] *School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522, Australia*

ABSTRACT

Student understanding and competency in probability have been investigated from different perspectives. Competency is often measured in the form of tests. The purpose of this study was to investigate whether perceived understanding and competency can be calibrated and assessed together using Rasch measurement tools. The study comprised 44 students who enrolled in the STAT131 Understanding Uncertainty and Variation course at University of Wollongong, Australia. Their voluntary participation in the study was through the e-learning Moodle platform where tests and assessment were administered online. Data were analysed using the Rasch measurement models. The study revealed majority of the students had little understanding about conditional and independent events prior to learning them but tended to demonstrate a slightly higher competency level afterward. Based on the Rasch map, there is an indication of some increase in learning and knowledge about probability concepts at the end of the two weeks lesson.

Keywords: Perceived understanding, competency, probability concepts, e-learning, Rasch measurement models

ARTICLE INFO

Article history:
Received: 24 February 2016
Accepted: 30 September 2016

E-mail addresses:
zamalia@tmsk.uitm.edu.my/zamal669@salam.uitm.edu.my (Zamalia, M.),
alp@uow.edu.au (Porter, A.)
* Corresponding author

ISSN: 0128-7702

INTRODUCTION

Many studies have examined difficulties faced by students in learning probability concepts (Garfield and Ahlgren, 1988; Shaughnessy, 1992; Garfield, 2003). According to Garfield and Ahlgren (1988) students have an underlying difficulty understanding the fundamentals of probability. Probability ideas also seem to conflict with students' real time experience in solving problems. Zamalia, Masniyati and Nor Azura (2013) showed that students perceived little understanding of certain basic probability concepts such as conditional probability and independent events.

No matter how the concepts of probability are taught, the question always arises as to how students study and understand the concepts. Educators acknowledge the complexity of this process. Conventional method of assessment takes raw scores as benchmark for student learning. How students perceive their learning and how well they perform in the tests are always treated separately. Therefore, this study will attempt to calibrate these two measures on a single interval scale so students' learning of probability can be gauged accurately.

The following are the research questions:

i. How do students perceive their level of understanding in probability concepts?
ii. What are students' competency in probability concepts?
iii. Do students who profess to having good understanding of probability concepts demonstrate a good competency in probability concepts?

LITERATURE REVIEW

Learning Statistics

Statistics courses are challenging for students in the social sciences (Forte, 1995; Yilmaz, 1996; Townsend, Moore, Tuck, & Wilton, 1998). Research into how students study statistics and probability have also been carried out from the cognitive aspects of learning (Kapadia, 1985; Garfield & Chance, 2000; Kassim, Ismail, Mahmud, Zainol, 2010). This is an important area of study because students with different backgrounds and characteristics undergo the learning processes in many different ways. In spite of the various methods of teaching and learning, many are still facing difficulty in learning statistics because of insufficient computation skills and negative attitudes towards the subject (Garfield, Hogg, Schau & Whittinghill, 2002).

Perceived Ability and Competency in Statistics

Students' perceived ability is an important indicator in predicting the level of performance or motivation among them. Perceived ability or perceived self-efficacy refers to one's belief about one's capabilities to achieve certain level of performance or ability in specific situations (Bandura, 1994). This core belief is the foundation of human motivation, accomplishments, and emotional well-being (Bandura, 1997, 2006). Harter (1982) on the other hand, considers perceived competence as a more global construct than self-efficacy which is consistent with Roberts, Klieber and Duda (1981) that the terms self-efficacy, perceived ability, perceived and physical competence are interchangeable.

Rumsey (2002) states that statistical competence includes data awareness, an understanding of certain basic statistical concepts and terminology, knowledge of the basics of collecting data and generating descriptive statistics, the ability to describe what the results mean in the context of the problem and being able to explain the results to someone else. Thus, every time the students go through the process, they will reinforce their understanding of the terms and concepts, reasoning and thinking skills.

METHODOLOGY

A survey was conducted among 44 mathematics and computer sciences undergraduates in the in the e-learning Moodle platform to. They had enrolled in the STAT131 Understanding Variation and Uncertainty as part of their programme requirement. They were given two sets of questionnaires to answer. The first set of questionnaire asked how they perceived their understanding of probability concepts. The items were related to the probability concepts requiring students to read through and understand the terms, definitions or/ and examples (see Figure 1).

The students were required to respond based on rating scales (from 1-5) as follows:

1) I have **NO UNDERSTANDING** of the term, definition or example.
2) I have **LITTLE UNDERSTANDING** of the term, definition or example.
3) I have **SOME UNDERSTANDING** of the term, definition or example.
4) I have **GOOD UNDERSTANDING** of the term, definition or example.
5) I have **FULL AND COMPLETE UNDERSTANDING** of the term, definition and example.

B5_i Conditional probability The conditional probability of B, given that A has occurred, is $P(B/A) = \frac{P(A \cap B)}{P(A)}$ if $P(A) \neq 0$	(1) (2) (3) (4) (5)
B5_ii The conditional probability of A, given that B has occurred, is $P(A/B) = \frac{P(A \cap B)}{P(B)}$ if $P(B) \neq 0$	(1) (2) (3) (4) (5)
B5_iii Example: A and B are two independent events such that P(A) = 0.2 and P(B) = 0.15. Then, P (A ∩ B) = P(A) = 0.2	(1) (2) (3) (4) (5)
B6_i Independent Events Event B is said to be independent of event A or event A does not affect the probability of the occurrence of event B. So: P(B/A) = P(B) P(A/B) = P(A) P (A ∩ B) = P(A). P(B)	(1) (2) (3) (4) (5)

Figure 1. Perceived Understanding of Probability Concepts Items

The second set of questionnaire tests student knowledge and competency in probability concepts. The items are constructed based on how they should solve probability problems. Students are required to state whether the solutions for each question is either true or false (see Figure 2).

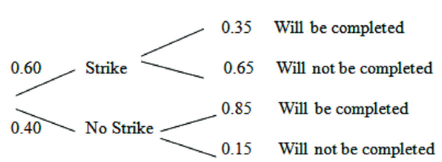
1. The completion of construction job may be delayed because of a strike. The probabilities are 0.60 that there will be a strike, 0.85 that the construction job will be completed on time if there is no strike, and 0.35 that the construction job will be completed on time if there is a strike.	
a. Draw a tree diagram.	
	
b. What is the probability given that the construction job will be completed if there is a strike?	T F

Figure 2. Items Representing Students' Competency in Probability Concepts

In order for the calibration to hold in both instruments, students were matched in both samples and their responses to the questionnaires were captured in Moodle site and later exported as an Excel file. Data were analysed using Winsteps 3.81.0 software to produce the relevant Rasch output (Linacre, 2007; Linacre, 2011).

Rasch measurement models

Two Rasch measurement models, namely dichotomous and polytomous rating scale, are used for calibrating the instruments. Also known as a probabilistic model, Rasch measurement takes into account two parameters – test item difficulty and the person's ability.

Dichotomous Rasch Model

This is a mathematical probability model that incorporates an algorithm that expresses the probabilistic expectations of item and the person's performances:

$$P_{ni}\{x_{ni}=1\,|\,B_{n,}D_i\}=\frac{\exp(B_n-D_i)}{[1+\exp(B_n-D_i)}\qquad(1)$$

Eq. (1) represents the conditional probability of person B_n on item D_i responding with a correct response (x = 1) or receiving a rating of 1. This Rasch model is a useful way to conceptualise the relationship of responses with person and item locations on the latent variable.

The equation shows that the probability of success is a function of the difference between a person's ability and the item difficulty. Thus, when $B_n = D_i$ and $B_n - D_i = 0$, the probability of a correct answer, $P\{X_{ni} = 1\} = 0.5$ (equal to half or 50%). When $B_n > D_i$ and $B_n - D_i > 0$, the probability of a correct answer, $P\{X_{ni} = 1\} > 0.5$ (more than half or 50%). When $B_n < D_i$ and the difference between $B_n - D_i < 0$, the probability of a correct answer, $P\{X_{ni} = 1\} < 0.5$ (less than half or 50%) (Bond & Fox, 2007).

Rasch Rating Scale Model

Rasch Polytomous / Rating Scale model is an extension of Rasch Dichotomous model where the items have more than two response categories or rating scale such as (1=strongly disagree, 2=disagree, 3=agree, 4=strongly agree) and it is modelled as having three thresholds. Each item threshold (k) has its own difficulty estimate (F), and this estimate is modelled as threshold at which a person has 50/50 chance of choosing one category over another.

The first threshold, for example, is modelled as the probability of choosing a response of 2 (disagree) instead of response 1 (strongly disagree), and is expressed using the following formula:

$$P_{ni1}\{x_{ni} = 1 \mid \beta_n, \delta_i, F_1\} = \frac{\exp(\beta - [\delta_i + F_1])}{1 + \exp(\beta - [\delta_i + F_1])} \quad (2)$$

where P_{ni1} is the probability of student n choosing "disagree" (Category 2) over "strongly disagree" (Category 1) on any item (i). In this equation, F_1 is the difficulty of the first threshold, and this difficulty calibration is estimated only once for this threshold across the entire set of items in the rating scale. The threshold difficulty F_1 is added to the item difficulty δi to indicate the difficulty of Threshold 1 in item i.

The Rating Scale model decomposes the category parameter, δ_{ij}, into two parameters: a location parameter δ_i that reflects item difficulty and a category parameter τ_j. The separation is achieved by using a probabilistic approach in which a person's raw score in a test is converted into a success-to-failure ratio and then into logarithmic odds that the person will correctly answer the items (Linacre, 2011). This is represented in a logit scale. When this is estimated for all persons, the logits can be plotted on one scale.

Assessing Data Fit

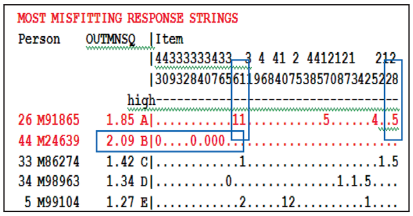
A Rasch analysis is a procedure for assessing the quality of raw score data using fit statistics, z-standard residuals, and point measure correlations (Bond & Fox, 2007). A Rasch analysis involves checking the degree to which the data match a unidimensional measurement model, identifying and diagnosing sources of discrepancy, removing items or persons if they are degrading the overall quality of measurement.

Infit and outfit mean square fit statistics are used in assessing quality of data. They provide summaries of the Rasch residuals, responses that differ from what is predicted by the Rasch model for each item and person. High mean square fit statistics indicate a large number of unexpected responses. High person mean square values indicate persons who filled in responses randomly and have unusual gaps in their knowledge. Item infit mean square values between 1.5 and 2.0 are considered to be unproductive for measurement, and values higher than 2.0 are actually degrading (Linacre, 2011).

ANALYSIS AND RESULTS

Perceived understanding and competency in probability concepts

The summary statistics shows the results of the perceived understanding and competency in probability concepts based on the analysis of data using Rasch measurement tools. The mean infit and outfit for person and item mean squares are 0.95, 1.09, 1,0 and 1.09 respectively. This indicates that the data had shown acceptable fit to the model. The mean standardised infit and outfit for person is between -0.4 and 0.1 which is within Rasch measurement acceptable range. The mean standardised infit and outfit for items is between 0.1 and 0.2. This indicates the items measure are slightly overfit and that the data fit the model somewhat better than expected [30]. The standard deviation of the standardised infit is an index of overall misfit for persons and items. Using 2.0 as a cut-off criterion, standardised infit/outfit standard deviation for persons is between 1.5 and 1.8 and standardised infit/outfit standard deviation for items is between 1.2 and 1.3. All show an overall acceptable fit. Separation is the index of spread of the person positions or item positions. Separation of 2.0 and above indicates the items have sufficient breadth in position. For persons, separation is 3.80 for the data at hand (*real*) indicating an approximately four levels of ability. The item on the other hand has separation index of 2.91 which indicates item difficulty can be separated into 3 difficulty levels. The person separation reliability estimate for this data is 0.94 (Cronbach's Alpha) which indicates a wide range of students' ability. The item separation reliability estimate is 0.89 which indicates items are replicable for measuring similar traits. The mean of the item logit position is arbitrarily set at 0.0, similar to standardised z-score. The person mean is 1.09 which suggests that a small group of students had a slightly good perception of understanding of probability concepts. For quality check, the data had gone through two stages of data cleanup where misfit responses on some items based on outfit mean square values of above 1.6 were identified and removed. Figure 3 shows the most misfitting response came from two male students (corresponding to ID number 26 and 44) with outfit mean square values of above 1.60. The table shows that the students did not respond appropriately according to the Rasch model. For example, student 26 was expected to disagree with a scale of 1 or 2 to the most difficult item 28 and agree with a scale of 3 or 4 to the fairly difficult items 6 and 1. Similarly, student 44 was expected to agree with items which are fairly simple for his ability but the reverse happened.



```
MOST MISFITTING RESPONSE STRINGS
Person      OUTMNSQ  |Item
                     |44333333433 3 4 41 2 4412121   212
                     |30932840765611968407538570873425228
               high---------------------------------------
26 M91865     1.85 A|..........11..........5......4..5
44 M24639     2.09 B|0....0.000........................
33 M86274     1.42 C|...........1..................1.5
34 M98963     1.34 D|.........0..............1.1.5....
 5 M99104     1.27 E|...........2.....12.........1....
```

Figure 3. Misfitting Response Strings of Students' Competency in Probability Concepts

Figure 4 shows the Wright map of perceived understanding and competency in probability concepts. The map displays the distribution of students (on the left side of the map) according to their ability from most able to least able in endorsing items as agree or correct. It also displays the items according to the difficulty levels. Four concepts from the perceived understanding instrument (i.e., B7i, B7ii, B7iii, B7iv) at logit values between 2.0 and 2.5 were found to be difficult to understand by 97% of the students while concepts A1ii, B8iii and B9iii at -1.0 logit value were found easiest to understand by 98% of the students. It was observed that majority of the students perceived little or no understanding about Bayes' theorem and conditional probability prior to the teaching of the concepts. At the competency level, there is a slight increase in the learning of conditional probability in between 1.0 and 1.5 logit. It was discovered that students found it hard to understand the concepts through the Bayes' formula (as in B7i, B7ii, B7iii and B7iv) but they understood more when the concepts were demonstrated in the form of solutions (as in Q1b).

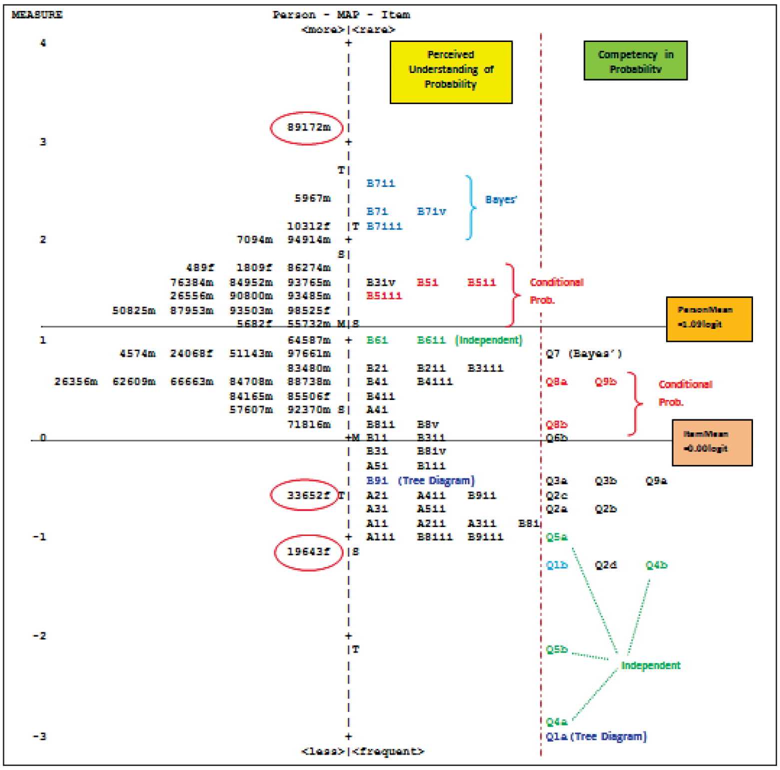


Figure 4. Wright Map of Perceived Understanding and Competency in Probability Concepts

About 40% of the students perceived some understanding to no understanding about the concept of independent events (as in B6i and B6ii) prior to studying it. However, most students found the concepts of independent events (as in Q4a, Q4b, Q5a, Q5b) easy to understand as shown in the location of logit values (between -1.0 and -3.0) on the map. In exploring students' perceived understanding of the probability concepts, about 45% perceived as having moderate to good understanding of about 60% of the concepts. The map shows a wide spread of competency items ranging between +1.0 and -3.0 logit. Small gaps were seen in between the competency items and the range of item difficulty did not match quite well with the ability of 70% of the students. About 70% of the probability test items were considered easy by the students. As the person (mean logit = +1.00) was greater than the item (mean logit =0.00), generally the test was considered easy by majority of the students. In investigating if data fit the model, the distribution of empirical data was plotted across the expected values for the perceived understanding of items in the Likert scale (Group L) and competency in probability concepts dichotomy items (Group D). This is shown in Figure 5. The characteristic curve for all empirical values in Group L falls along the expected ogive curve and within the upper and lower bound of the 95% confidence interval. This indicates a good item person targeting for the perceived understanding towards probability items. On the other hand, the characteristic curve for all empirical values in Group D mostly falls along the upper 50% of expected ogive curve and within the upper and lower bound of the 95% confidence interval. A wide confidence interval is seen around the middle section of the curve compared to the upper section. Two empirical observations did not behave according to the Rasch model. However, these points are considered negligible as most of the other empirical points were closer to the upper section of the expected Rasch model. This also signals the data fit the model better than expected.

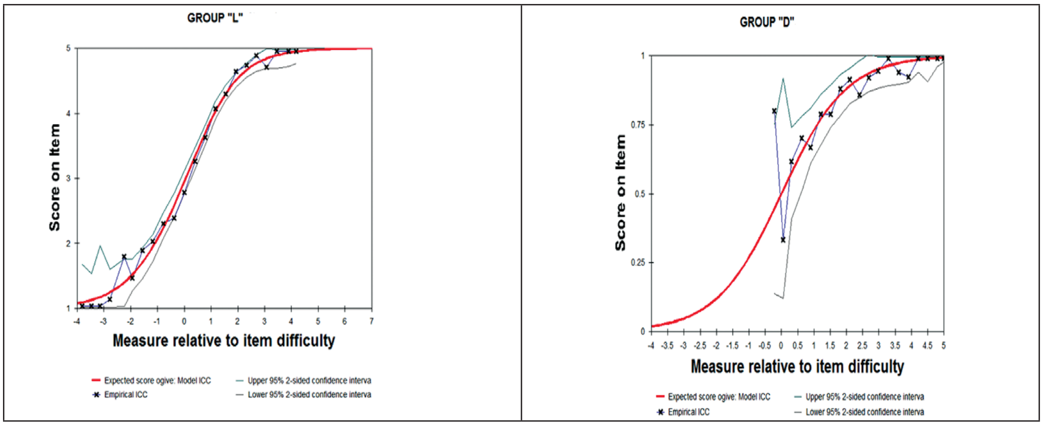


Figure 5. Empirical-Expected Item Characteristic Curves for Likert Scale and Dichotomous Items

DISCUSSION AND CONCLUSION

This study has attempted to show that perceived understanding and competency can be calibrated and assessed together using the Rasch measurement tools. Rasch measurement which is based on the Rasch probabilistic models were used to calibrate the responses from two survey instruments and investigate the interactions between them. The study showed that majority of the students perceived little understanding about conditional and independent events prior to studying about them but tended to demonstrate a slightly higher competency level afterward. Based on the Rasch map, there was indication of some increase in learning and knowledge about probability concepts at the end of the two weeks lesson. The study discovered that students had perceived a greater understanding of probability concepts after two weeks of exposure to them. However, when perceived understanding was calibrated against their competency in probability concepts, the students performed better than expected. Many students who initially perceived they had little understanding of probability concepts had shown a much higher understanding of the concepts after two weeks of study.

ACKNOWLEDGEMENTS

The authors would like to extend their utmost appreciation to Universiti Teknologi MARA and University of Wollongong, Australia for funding and providing them the opportunity to conduct this study. They also record their appreciation to PETANIKA for publishing this paper.

REFERENCES

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71-81). New York: Academic Press. (Reprinted in H. Friedman [Ed.], Encyclopedia of Mental Health. San Diego: Academic Press, 1998.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (2006). Toward a Psychology of Human Agency. *Perspectives on Psychological Science*, *1*(2), 164–180.

Bond, T. G., & Fox, C. M. (2007). *Applying The Rasch Model: Fundamental Measurement In The Human Sciences*. New Jersey: Lawrence Erlbaum Associates.

Forte, J. A. (1995). Teaching statistics without sadistic. *Journal of Social Work Education*, *31*(2), 204–218.

Garfield, J. (2003). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education, 10*(3), 58-69. Retrieved from http://www.amstat.org/publications/jse/v10n3/garfield.html.

Garfield, J., Hogg, B., Schau, C., & Whittinghill, D. (2002). First Courses in Statistical Science: The Status of Educational Reform Efforts. *Journal of Statistics Education*, *10*(2), 456-467.

Garfield, J., & Ahlgren, D. (1988). Assessing Students' Conceptual Understanding After A First Course In Statistics; Difficulties in Learning Basic Concepts in Probability and Statistics: An Implication for Research. *Journal for Research in Mathematics Education, 19*(1), 44-63.

Garfield, J., & Chance, B. (2000). Assessment in Statistics Education: Issues and Challenges. *Mathematics Thinking and Learning, 2*(1-2), 99-125.

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development*, *53*(1), 87–97.

Kapadia, R. (1985). A Brief Survey of Research on Probabilistic Notions. In A. Bell, B. Low, & J. Kilpatrick (Eds.), *Theory, Research and Practice in Mathematical Education* (pp. 261-265). Nottingham, UK: Shell Centre for Mathematical Education.

Kassim, N. L. A., Ismail, N. Z. I., Mahmud, Z., & Zainol, M. S. (2010). Measuring Students' Understanding of Statistical Concepts using Rasch Measurement. *International Journal of Innovation, Management and Technology, 1*(1), 13-20.

Linacre, J. M. (2007). *A User's Guide to WINSTEPS*. Chicago: Winsteps.com.

Linacre, J. M. (2011). Rasch Measures and Unidimensionality. *Rasch Measurement Transactions, 24*(4), 1310.

Roberts, G., Klieber, D., & Duda, J. (1981). An analysis of motivation in children's sport: The role of perceived competence in participation. *Journal of Sport Psychology*, *3*(3), 206–216.

Rumsey, D. J. (2002). Statistical Literacy as a Goal for Introductory Statistics Courses. *Journal of Statistics Education, 10*(3), 6-13. Retrieved from http://www.amstat.org/publications/jse/v10n3/rumsey2.html]

Shaughnessy, J. M. (1992). Research In Probability And Statistics: Reflections And Directions. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 465-494). New York: Macmillan.

Townsend, M., Moore, D., Tuck, B., & Wilton, K. (1998). Self-concept and anxiety in university students studying social science statistics within a co-operative learning structure. *Educational Psychology*, *18*(1), 41–53.

Yilmaz, M. R. (1996). The Challenge of teaching statistics to non-specialists. *Journal of Statistics Education*, *4*(1), 1-9. Retrieved from www.amstat.org/publications/jse/v4n1/yilmaz.html

Zamalia, M., Masniyati, S., & Nor Azura, M. G. (2013). Diagnosis of Engineering Students' Learning Ability in Selected Statistical Topics Using Rasch Measurement Tools. In *Proceedings of 9th WSEAS International Conference on Educational Technologies*.