USING RASCH ANALYSIS TO EXPLORE WHAT STUDENTS LEARN ABOUT PROBABILITY CONCEPTS

Zamalia Mahmud[1], Anne Porter[2]

[1]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia
[2]School of Mathematics and Applied Statistics, University of Wollongong, NSW Australia
e-mail: zamalia@tmsk.uitm.edu.my

Abstract

Students' understanding of probability concepts have been investigated from various different perspectives. This study was set out to investigate perceived understanding of probability concepts of forty-four students from the STAT131 Understanding Uncertainty and Variation course at the University of Wollongong, NSW. Rasch measurement which is based on a probabilistic model was used to identify concepts that students find easy, moderate and difficult to understand. Data were captured from the e-learning Moodle platform where students provided their responses through an on-line quiz. As illustrated in the Rasch map, 96% of the students could understand about sample space, simple events, mutually exclusive events and tree diagram while 67% of the students found concepts of conditional and independent events rather easy to understand.

Keywords: Perceived Understanding, Probability Concepts, Rasch Measurement Model

Abstrak.

Pemahaman siswa terhadap konsep peluang telah diteliti dari berbagai perspektif yang berbeda. Penelitian ini dilaksanakan untuk menyelidiki pemahaman yang dipersepsi oleh empat puluh empat siswa tentang konsep peluang dari perkuliahan STAT131 Memahami Ketidakpastian dan Variasi di University of Wollongong, NSW. Pengukuran Rasch yang didasarkan pada model probabilistik digunakan untuk mengidentifikasi konsep yang mudah, sedang dan sulit dimengerti oleh siswa. Data diambil dari platform Moodle e-learning dimana siswa memberikan tanggapan mereka melalui kuis on-line. Seperti digambarkan dalam peta Rasch, 96% siswa dapat memahami tentang ruang sampel, kejadian sederhana, kejadian saling eksklusif dan diagram pohon sementara 67% siswa mudah memahami konsep kejadian bersyarat dan independen.

Kata Kunci: Pemahaman yang Dipersepsi, Konsep Peluang, Model Pengukuran Rasch

Statistics is an important element of the curriculum for students in a variety of majors. Increasingly elements of data analysis and probability are also being emphasized in industry in a variety of disciplines including engineering and computer science. It is becoming increasingly prevalent as students are required to learn the skills of statistical reasoning and develop the ability to translate information (Jensen & Kellogg, 2010).

Students' difficulties in learning and understanding probability have been known from several research studies and have been well documented (Garfield, 2003; Shaughnessy 1992; Konold, 1989; Garfield & Ahlgren, 1988). According to Garfield and Ahlgren (1988), students have an underlying difficulty with fundamental ideas of probability. Apart from their weakness with rational number concepts and proportional reasoning (Matthews & Silver, 1983), probability ideas appear to conflict with students' experience about how they view the world. In a recent study, Zamalia *et. al.* (2013)

discovered that about 38% of the students perceived little understanding on certain basic probability concepts such as conditional probability and independent events. Thus the main purpose of this study is to investigate the level of students' perceived understanding of probability concepts and identify which concepts were found most difficult by the students to understand.

Over the years, research into how students learn has evolved in many different directions. A large number of studies has been carried out in areas such as cognitive aspects of learning (Kolb, 1984; Sadler-Smith, 1996; Garfield, 1995; Garfield and Chance, 2000). Students enter learning processes with different background characteristics such as a preference for deep learning versus surface learning, and specific subject attitudes, and different perceptions of the learning context. Most of these contexts allow all students to achieve satisfactory learning outcomes, with different learning paths (Tempelaar, 2006).

Statistical concepts are the basis of learning statistics and therefore should be given extra attention by every educational institution. Much research in the different types of statistical reasoning such as reasoning about variation, distribution, and sampling distributions, has created important insights into the developmental process of a student's learning of statistical reasoning skills (Tempelaar, 2006). Studies have also shown that students have difficulty with reasoning about distributions and graphical representations of distributions (Garfield and Ben-Zvi, 2004), understanding concepts related to statistical variation such as measures of variability (delMas, Garfield & Chance, 1999) and sampling distributions (Saldanha & Thomson, 2001). Contemporary research in statistics education distinguishes an array of different but related cognitive processes in learning statistics: statistical literacy, statistical reasoning, and statistical thinking. Literacy, reasoning, and thinking are to some extent achieved even before formal schooling in statistics takes place. Those naïve conceptions learned outside school can be correct or incorrect in nature (Tempelaar, Schim & Gijselaers, 2007).

Garfield (2003) made the attempt to assess student's reasoning through the Statistical Reasoning Assessment (SRA) but the items in the SRA are focused more on the probability topics instead of basic statistical concepts. The SCI (Statistics Concept Inventory) too was developed to assess statistical understanding but it was specifically designed for the engineering students (Reed-Rhoads, Murphy, & Terry, 2006). After three years of research on their Assessment Resource Tools for Improving Statistical Thinking (ARTIST) project, funded by the NSF (National Science Foundation), delMas, Garfield, Ooms and Chance (2007) produced an online test, Comprehensive Assessment of Outcomes in Statistics (CAOS). The objective of CAOS is to measure students' understanding on the topics contained in most introductory statistics courses.

METHOD

Study Design

A survey was administered on 44 undergraduate students representing the mathematics and computer sciences. They enrolled in the STAT131 Understanding Variation and Uncertainty as part of the requirement for their various programmes of study. The students responding had volunteered to participate by providing brief information about their profile. They were given a set of questionnaire to answer. The questionnaire asked how they perceived their understanding in probability concepts. The items constructed are related to the probability concepts where students would need to read through and understand the term, definition or examples. A sample of the items is shown in Table 1.

The students had responded to the items based on the perceived level of understanding scales of between (1) and (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.

Table 1. Items Representing Perceived Understanding of Probability Concepts

B. Relationships Among Events	
B1_i Complementary Event Let E = Event E occurs Let E' = Event E does not occur. then P(E') = 1- P(E)	(1) (2) (3) (4) (5)
B1_ii Example: A die is toss once. The sample space S={1,2,3,4,5,6}, so n(S) = 6 Let A = Event obtaining a 3 on the uppermost face Let B = Event not obtaining a 3 on the uppermost face P(A) = 1/6 P(B) =1-1/6 = 5/6	(1) (2) (3) (4) (5)
B2_i General Addition Rule Given two events, A and B, the probability of their union, $A \cup B$ is equal to $P(A \cup B) = P(A) + P(B) – P(A \cap B)$	(1) (2) (3) (4) (5)

In order for the calibration to hold between person and test items, students' responses to the questions were captured and raw scores obtained which are then converted to interval logit values using the Polytomous Rasch measurement model. Students' responses to the questionnaires were captured in Moodle site and later exported as an Excel file. Data were analyzed using Winsteps 3.74.0 software to produce the relevant Rasch output (Linacre, 2007).

Polytomous Rasch Model

Also known as a probabilistic model, Rasch measurement takes into account two parameters – test item difficulty and person ability.

The polytomous (rating scale) Rasch model establishes the relative difficulty of each item from the lowest to the highest levels the instrument is able to record. It is more complex than the dichotomous Rasch model as it is possible to endorse one of the many response categories on a scale. The items indicate a rather more complicated representation than the one for dichotomous data. For dichotomous data, each item is represented as having a single item estimate, with an associated error estimate. For rating-scale data, not only does each item have a difficulty estimate, but the scale also has a series of thresholds (i.e., the level at which the likelihood of failure at a given response category [below the threshold] turns to the likelihood of success at that category [above the threshold]).

Response categories in Likert instruments may include ordered ratings, such as "Strongly Disagree/ Disagree/ Agree/ Strongly Agree", to represent a respondent's increasing inclination towards the concept questioned. The response rating scale, when it works, yields ordinal data which need to be transformed to an interval scale to be useful. This is achieved by the Rasch rating scale model (Andrich, 1978).

The polytomous "Rasch Rating Scale" model is a mathematical probability model, which incorporates an algorithm that expresses the probabilistic expectations of item and person responses, which estimates the probability that a person will choose a particular response category or an item as:

$$\ln\left(P_{nij} \,/\, P_{ni(j-1)}\right) = B_n - D_i - F_j$$

where,

ln	=	a natural logarithm
P_{nij}	=	the probability of respondent *n* scoring in category *j* for item *i*
$P_{ni(j-1)}$	=	the probability of scoring in category (*j*-1)
B_n	=	the person measure/ability of respondent *n*
D_i	=	the difficulty of item *i*
F_j	=	the difficulty of category step *j*

(the threshold at which there is a 50-50 chance of scoring in category *j* and category *j* – 1)

Table 2. Thresholds and Category Fit

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	120	7	-1.40	-1.46	1.08	1.04	NONE	(-2.93)	1
2	2	241	14	-.39	-.48	.99	.94	-1.65	-1.19	2
3	3	365	21	.36	.42	1.08	1.17	-.44	.06	3
4	4	438	25	1.16	1.28	1.12	1.04	.67	1.22	4
5	5	579	33	2.19	2.11	.84	.94	1.42	(2.78)	5
MISSING		5	0	1.17						

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

Information in Table 2 helps the investigation of the rating scale quality as to whether the categories fit the model sufficiently well and whether the thresholds indicate a hierarchical pattern to the rating scale. Basic examination of rating scale used in the Table 2 indicates that each category has provided enough observations for an estimation of stable threshold values. The recommended minimal number of responses per category is 10 (Linacre, 1999a). Based on step calibrations of Andrich threshold, all categories are ordered and increases monotonically. For example, Category 1 was recorded as -2.93 which can be interpreted as the average ability estimate, or logit score, for persons who chose Category 1 on any item in the questionnaire. Similarly for Category 2 until Category 5. To further support this, observation based on outfit mean squares for each category shows the fit of each rating scale category to the unidimensional Rasch model meet the criterion of mean square statistics less than 2.0 (Linacre, 1999a).

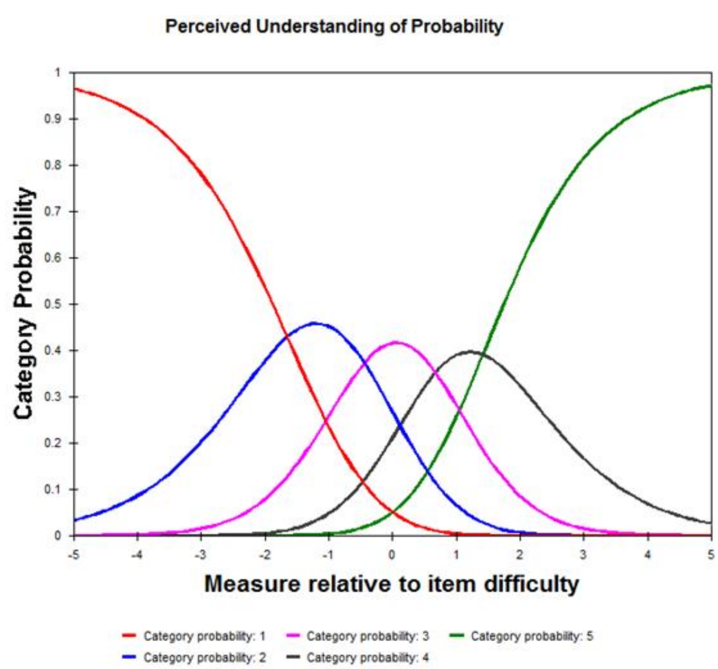

Figure 1. Probability curves for a well-functioning five category rating scale

The Rasch analysis places persons (B_n) and items (D_i) on the same measurement scale where the unit of measurement is the logit (logarithm of odds unit). The person's likely score is defined by the interaction between the person's measure, the item's difficulty, and the score's category threshold.

These parameters are assumed to be interdependent. However, separation between the two parameters is also assumed. For example, the items (questions) within a test are hierarchically ordered in terms of their difficulty and concurrently, persons are hierarchically ordered in terms of their ability. 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 a logarithmic odds that the person will correctly answer the items (Bond & Fox, 2007). This is represented in a logit scale. When this is estimated for all persons, the logits can be plotted on one scale.

RESULTS AND DISCUSSION

Perceived Understanding in Probability Concepts

Table 3 presents the summary statistics for perceived understanding in probability concepts based on the analysis of data using Rasch measurement tools. The statistics show the mean infit and outfit for person and item mean squares are close to 1.0 which indicate that in general the data had shown acceptable fit to the model. The mean standardized infit and outfit for person is between -0.3 and -0.2. The standardized outfit is within acceptable range of rasch measurement (± 1.0). The mean standardized infit and outfit for items is located at 0. This indicates the items measure are slightly overfit and that the data fit the model somewhat better than expected. (Bond & Fox, 2007).

Table 3. Summary Measures of Perceived Understanding in Probability Concepts

SUMMARY OF 46 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	137.9	37.9	.94	.21	.99	-.3	1.03	-.2
S.D.	22.9	.4	.96	.03	.49	2.0	.65	2.2
MAX.	181.0	38.0	3.39	.36	2.56	5.1	3.78	7.6
MIN.	64.0	36.0	-2.02	.19	.30	-4.6	.30	-4.6

REAL RMSE .23 TRUE SD .93 SEPARATION 4.05 Person RELIABILITY .94
MODEL RMSE .21 TRUE SD .94 SEPARATION 4.42 Person RELIABILITY .95
S.E. OF Person MEAN = .14

Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .95

SUMMARY OF 38 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	166.9	45.9	.00	.19	1.01	.0	1.03	.0
S.D.	34.7	.3	1.08	.02	.30	1.3	.39	1.5
MAX.	209.0	46.0	2.42	.24	1.83	3.1	2.40	4.7
MIN.	89.0	45.0	-1.52	.17	.61	-2.2	.59	-2.1

REAL RMSE .20 TRUE SD 1.06 SEPARATION 5.23 Item RELIABILITY .96
MODEL RMSE .19 TRUE SD 1.06 SEPARATION 5.55 Item RELIABILITY .97
S.E. OF Item MEAN = .18

UMEAN=.0000 USCALE=1.0000
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00

Table 3 shows the standard deviation of the standardized infit as an index of overall misfit for persons and items. Using 2.0 as a cut-off criterion, standardized infit/outfit standard deviation for persons is between 2.0 and 2.2 and standardized infit/outfit standard deviation for items is between 1.3 and 1.5. All show an overall acceptable fit.

Separation is the index of spread of the person positions or item positions. Separation of 1.0 or below indicates the items may not have sufficient breadth in position. For persons, separation is 4.05 for the data at hand (*real*) indicating approximately four levels of person ability. The item on the other hand has a separation index of 5.23 which indicates item difficulty can be separated into 5 levels. Person and item separation and reliability of separation assess instrument spread across the trait continuum. Separation also determines reliability. Higher separation in concert with variance in person or item position yields higher reliability. The person separation reliability estimate for this data is 0.94 which indicate a wide range of students' ability. The item separation reliability estimate is 0.96 which indicates items are replicable for measuring similar traits.

The mean of the item logit position is always arbitrarily set at 0.0, similar to standardized z-score. The person mean is 0.94 suggesting that a small group of students had perceived their understanding of probability concepts quite well. From the perspective of Rasch measurement, this indicates some items were easily endorsed or easy to agree with.

***Person-Item Distribution Map* for Perceived Understanding**

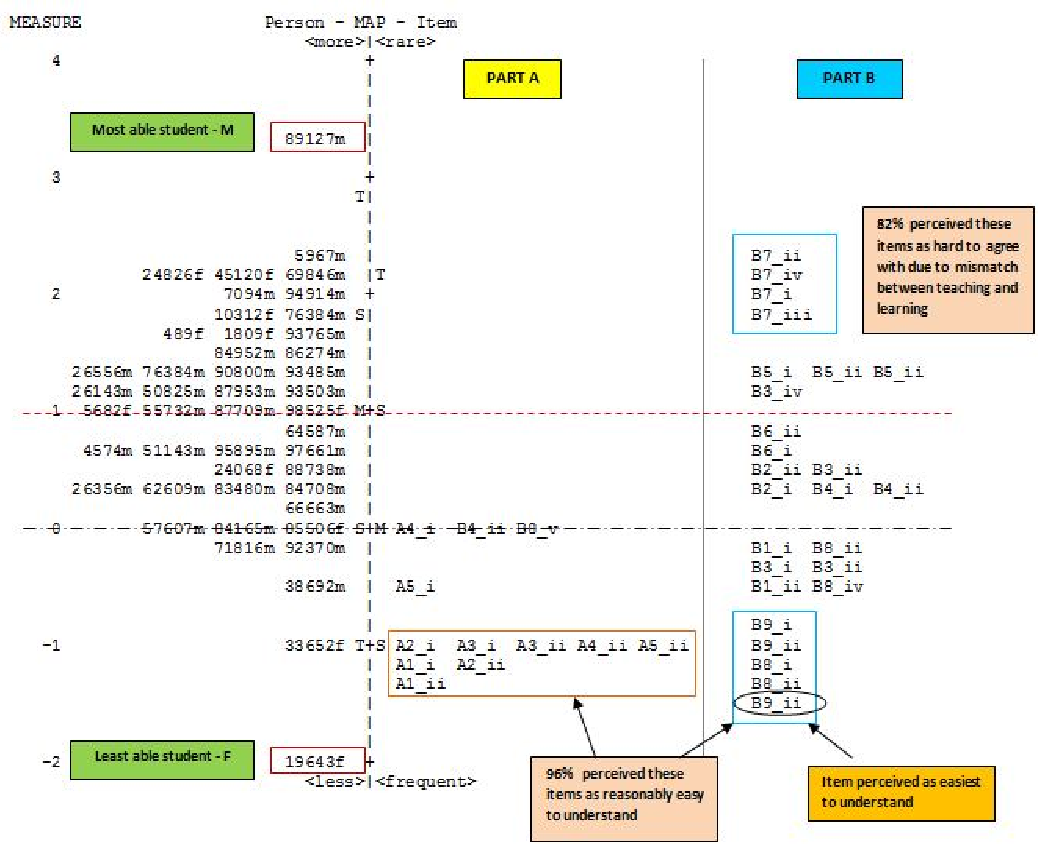

Figure 2. Person-Item Distribution Map of Perceived Understanding of Probability Concepts

Figure 2 shows the person-item distribution map of perceived understanding of probability concepts. The map display 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. The map also displays the items according to the difficulty levels.

It is expected that many students will have little or no understanding about Bayes' theorem and conditional probability concepts. At the time when this instrument was administered, conditional probability was exposed using few practical examples while the illustration of the Bayes' Theorem formula was not emphasized. Hence, there is a slight mismatch between how the concept was taught and the development of the items. This explains why majority of the students could not endorse items B7i, B7ii, B7iii and B7iv (logit values between 2.0 and 2.5), items which are related to the Bayes' Theorem concept. On the other hand, about 97% of the students found concepts A1ii, B8iii and B9iii (at logit value of -1.0) which are directed to simple definitions of event, probability and tree diagrams are the easiest to endorse. Only about 33% of the students found concepts of conditional and independent events as difficult to understand. Generally students have perceived the items as quite easy to understand as the item mean logit is lower than the person mean logit.

In the investigation of data fitting the model, the distribution of empirical data are plotted across the expected values for the perceived understanding Likert scale items (Group L) as shown in Figure 3. 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 of probability items. This also signals the data fit the model better than expected.

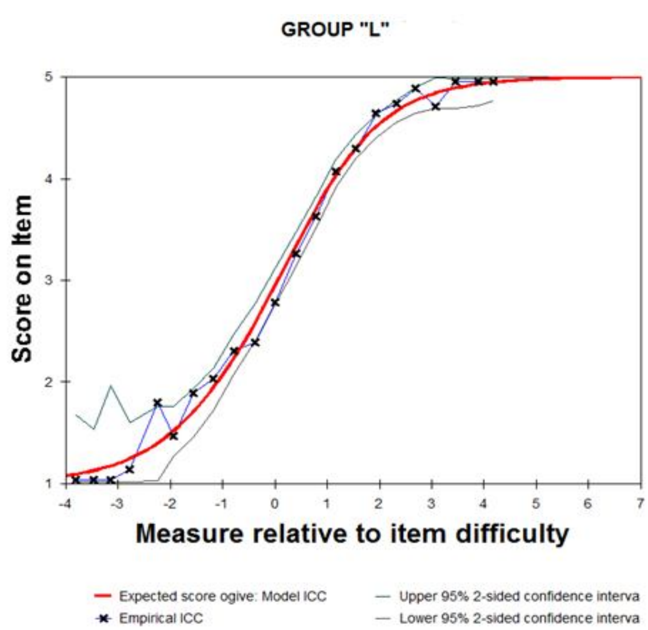

Figure 3. Empirical-Expected Item Characteristic Curves For Perceived Understanding Items

CONCLUSION AND SUGGESTION

This study has shown that students' level of perceived understanding of probability concepts can be identified using the Rasch polytomous measurement tools. Generally a large number of students (96%) perceived a good understanding about sample space, simple events, complementary events, and mutually exclusive events. About 96% of the students could understand about sample space, simple events, mutually exclusive events and tree diagram while 67% of the students found concepts of conditional and independent events rather easy to understand. A brief interview with several students confirmed that they have difficulties learning these concepts due to lack of exposure to these concepts at schools. However, current teaching in the STAT131 class has helped them to deal with prior misunderstandings of probability concepts. Students who initially have little understanding of the probability concepts wish to demonstrate a greater understanding of the concepts after two weeks of exposure to the topics.

REFERENCES

Andrich, D. (1978). A Rating Formulation for Ordered Response Categories. *Psychometrics, 43,* 561-73.

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

delMas, R., Garfield, J., & Chance, B. (1999). A Model of Classroom Research In Action: Developing Simulation Activities To Improve Students' Statistical Reasoning. *Journal of Statistics Education, 7 (3)*.

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing Students' Conceptual Understanding After A First Course In Statistics. *Statistics Education Research Journal, 6* (2), 28-58.

Garfield, J. & Ahlgren. (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. & Ben-Zvi, D. (2004). *Developing Students Statistical Reasoning: Connecting Research And Teaching Practice*. Springer.

Garfield, J. (1995). How Students Learn Statistics. *International Statistical Review, 63 (1),* 25-34.

Garfield, J. (2003). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education, 10,* 3.

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

Jensen, D. & Kellogg, S. (2010). Improving Conceptual Understanding in Probability and Statistics. *American Society for Engineering Education,* 2-15.

Kolb, D.A. (1984). *Experiential Learning: Experience As The Source of Learning And Development.* Englewood Cliffs, NJ: Prentice Hall

Konold, C. (1989). Informal Conceptions of Probability. *Cognition and Instruction, 6,* 59-98.

Linacre, J.M. (2002). Understanding Rasch Measurement: Optimal Rating Scale Category Effectiveness. *Journal of Applied Measurement, 3,* 85-106.

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

Reed-Rhoads, T., Murphy, T. J., & Terry, R. (2006). *The Statistics Concept Inventory (SCI)*. Retrieved from http://coecs.ou.edu/sci/

Sadler-Smith, E. (1996). Approaches to Studying: Age, Gender, and Academic Performance. *Educational Studies, 22 (3)*, 367-379.

Saldanha, L.A. & Thompson, F.W. (2001). Students' Reasoning about Sampling Distributions and Statistical Inference. In R. Speiser & C. Maher (Eds.), *Proceedings of The Twenty-Third Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* Snowbird, Utah. Columbus, Ohio: ERIC Clearinghouse, 449-454.

Shaughnessy, J.M. (1992). Research in probability and statistics: Reflections and directions. In *Handbook of Research on Mathematics Teaching and Learning*. D.A. Grouws (Ed.). pp. 465-494. New York: Macmillan

Tempelaar, D. (2006). A Structural Equation Model Analyzing The Relationship Students' Statistical Reasoning Abilities, Their Attitudes Toward Statistics And Learning Approaches. *Proceedings of The 7th International Conference on the Teaching of Statistics*. ICOTS-7.

Tempelaar, D., Schim van der Loeff, S., & Gijselaers, W.H. (2007). A Structural Equation Model Analyzing the Relationship of Students' Attitudes Toward Statistics, Prior Reasoning Abilities, and Course Performance. *Statistics Education Research Journal, 6 (2)*, 78-102.

Zamalia, M., Porter, A., Nor Izzati, O., Masniyati, S., Az'lina, A. H., & Shamsiah, S. (2013). Understanding Probability Concepts and Learning Styles Among Students of the Malay Culture: A Rasch Analysis. *Wulfenia Journal, 20 (1)*, 98-115.