The Journal of Social Sciences Research
ISSN(e): 2411-9458, ISSN(p): 2413-6670
Special Issue. 6, pp: 1056-1061, 2018
URL: https://arpgweb.com/journal/journal/7/special_issue
DOI: https://doi.org/10.32861/jssr.spi6.1056.1061
Academic Research Publishing Group
Original Research
Open Access

Probabilistic Outcome of Rasch Polytomous Model Upon Measurement of Statistical Learning

Zamalia Mahmud*
Centre of Statistical and Decision Science Studies Faculty of Computer and Mathematical Sciences Universiti Teknologi MARA (UiTM) 40450 Shah Alam Selangor, Malaysia

Wan Syahira Wan Ramli
Centre of Statistical and Decision Science Studies Faculty of Computer and Mathematical Sciences Universiti Teknologi MARA (UiTM) 40450 Shah Alam Selangor, Malaysia

Shamsiah Sapri
Centre of Statistical and Decision Science Studies Faculty of Computer and Mathematical Sciences Universiti Teknologi MARA (UiTM) 40450 Shah Alam Selangor, Malaysia

Balkish Mohd Osman
Centre of Statistical and Decision Science Studies Faculty of Computer and Mathematical Sciences Universiti Teknologi MARA (UiTM) 40450 Shah Alam Selangor, Malaysia

Haliza Hasan
Centre of Statistical and Decision Science Studies Faculty of Computer and Mathematical Sciences Universiti Teknologi MARA (UiTM) 40450 Shah Alam Selangor, Malaysia

Abstract

Assessing students' ability in learning statistics has been a challenge for many statistical educators. To be able to know exactly how well students have learned and mastered the statistical contents can be difficult without the use of proper measurement tools. Even though types of assessment of students learning differ from one instructor to another, however common method of assessment based on the overall raw scores of student performance is still being employed. However, the drawback of the conventional assessment based on summated raw scores is that it does not tell us specifically which part of the statistical concepts that student understand the most or the least. Hence, the motivation of this study is to find an alternative approach to measure students' ability in learning statistical contents based on the Rasch Polytomous measurement model. The main objective is to assess individual student ability in learning statistics based on the computation of probability values of the measurement model. The learning assessment are measured based on the students' actual learning ability on their final examination scores. Students' ability in answering correctly their final examination questions given the ability of the students and difficulty of the questions varies at different levels of probability values.

Keywords: Learning statistics; Rasch polytomous; Students' ability.

1. Introduction

Traditionally, assessment has been used primarily to assign grades and give periodic feedback on student learning. More recently, assessment has come to include practices that better inform the instructor of students' understanding and reasoning processes, develop students' learning skills, and improve instructional practices (Ben-Zvi, 1999). Common assessment method is normally based on assigning of grades that can be used to help students learn how to improve their performance, either on the current task or on future ones (Garfield and Chance, 2000).

Data gathered from the assessments also can help instructor about what the students are learning, about their competencies, areas of weakness and reactions to the course. There are various methods of assessment used to evaluate students' learning. In a study on assessing students' learning ability in an undergraduate statistical course, Zamalia *et al.* (2013) has employed a method based on the entrance and exit survey in order to gauge students' perceived understanding of the subject of interest using Rasch analysis. Based on a case study of sixty undergraduate students who took a Statistics and Probability course, items or constructs were developed based on the respective course learning outcomes. Entrance survey was administered at the beginning of their first class where students were required to respond to the entrance survey items through UiTM i-learn portal. At the end of 14 weeks, exit survey was administered to the students after they went through the learning process. This study showed that Rasch measurement has provided an alternative method to measure the actual ability of the students in learning statistical concepts (Nor Lide *et al.*, 2010). This is in contrast with the conventional assessment method where summated raw scores could not specifically measure which part of the statistical concepts that students understand the most or the least. Hence, the motivation of this study is to find an alternative approach to measure students' ability in learning statistical contents based on a specific Rasch Polytomous model.

*Corresponding Author

2. Literature Review

2.1. Students' Learning of Statistical Concepts

Statistical course has become increasingly important as part of the mainstream curriculum through all levels of education (Mills, 2004). In a study of assessing students' conceptual understanding after a first course in statistics, Delmas *et al.* (2007) found that many students do not have a good understanding on contents that represent important learning outcome for an introductory statistics course. When students were asked whether a sample of 10 tosses or 100 tosses of a fair coin was more likely to have exactly 70% heads, students tended to correctly choose the small sample, which seemed to indicate that they understood that small sample are more likely to deviate from the population than are large samples. In other words, it can be shown that although students may be able to answer some test items correctly or perform calculations correctly, they may still misunderstand basic ideas and concepts (Garfield and Delmas, 1991). Measuring students' understanding of statistical concept revealed that it is more difficult to get a better conceptual understanding of basic statistical concepts than statistical competence (Zamalia *et al.*, 2013). The results also suggest that students more likely overestimated their understanding of basic statistical concepts, particularly those requiring conceptual understanding of the concepts.

2.2. Measurement of Learning Based on Rasch Measurement Model

Rasch model was used to measure students' performance in an undergraduate engineering subject (Saidfudin *et al.*, 2010). They found that Rasch measurement model can classify examination grades into learning outcomes more accurately as illustrated by the item measure, person measure and variable map. Rasch Model can precisely classify the students according to their achievements and reveals the true degree of learning abilities of the students (Osman *et al.*, 2011). Variable map was also established in positioning students' ability and item difficulty against the course outcomes of a Civil Engineering II subject and point measure correlation was reported to indicate suitability of items. Based on the same study, they discovered that Rasch Model is appropriate in evaluating students' performance in IPTA due to its ability to analyze student achievement more accurately (Osman *et al.*, 2012). This is supported by another study that measure students' performance in the Intelligent System Program by generating logit values which led to the computation of probability of CLOs achievement for each student (Ahmad *et al.*, 2010). This model is also able to produce the association pattern between students and their performance level for each CLO. Generally, most measurement of learning studies stressed on the use of the Rasch measurement tools and produce the logit values which is the transformation of raw score into interval scales. However, there is a lack of presentation of a probability distribution of the Rasch measurement model. Although a few studies mentioned about the probability outcomes but its distribution is seldom explored. Hence, this study will measure students' learning of statistics based on the distribution of Rasch probabilistic outcomes.

3. Methodology

3.1. Probabilistic Outcome based on Rasch Polytomous Model

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

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

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

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 δ_i to indicate the difficulty of Threshold 1 in item i (Bond and Fox, 2007).

In this study, Rasch polytomous model is used to assess students' ability in learning Probability and Statistics course based on their final examination scores. Students have to answer five questions and each question carries an equal full mark of 20 points. Scores based on correct answers from each question are then transformed into rating scale. In this study, a typical order rating scale of 1 to 5 is used: ≥ 75 = (5), 60-74 = (4), 47-59 = (3), 40-46 = (2), <39 = (1). This numerical coding is necessary for further evaluation of students' learning ability using Winsteps 3.90.2. Example of probability values of correct item response calculated from equation (1) are calculated for a sample of students as shown in Table 1.

Table-1. Probability Outcome of Correct Response given Certain Level of Person's Ability and Item Difficulty

	Item	Question 1	Question 2	Question 3	Question 4	Question 5
Student	Logit	-1.67	-1.23	1.57	2.61	3.44
1	1.79	96.95 (0.9695)	95.35 (0.9535)	55.48 (0.5548)	30.58 (0.3058)	16.11 (0.1611)
2	1.20	94.63 (0.9463)	91.91 (0.9191)	40.85 (0.4085)	19.62 (0.1962)	9.62 (0.0962)
3	0.65	91.05 (0.9105)	86.76 (0.8676)	28.50 (0.2850)	12.35 (0.1235)	5.79 (0.0579)
4	-0.50	76.31 (0.7631)	67.48 (0.6748)	11.20 (0.1120)	4.27 (0.0427)	1.91 (0.0191)
5	-1.24	60.59 (0.6059)	49.75 (0.4975)	5.68 (0.0568)	2.08 (0.0208)	0.92 (0.0092)

Probability outcome of correct response can be defined as follow:
For Student 1 answering Question 1: P11(Y = 1 | β_1 = 1.79, δ_1 = -1.67) = 96.95% ~ 0.9695
For Student 4 answering Question 3: P43(Y = 1 | β_4 = -0.50, δ_3 = 1.57) = 11.20% ~ 0.1120

3.2. Secondary Data Sources

Data used in this study are raw scores from students' final examination of a Probability and Statistics course. Data collected from a sample of 60 students were transformed into logit values using Winstep 3.90.2 software. The analysis involved checking for the reliability and separation index and data validation based on mean square fit statistics for persons and items. This process has enabled us to detect some problematic, misfit items or unusual responses from the data.

4. Results and Discussion

4.1. Probabilistic Outcome of Students' Performance Based on Final Examination

Table 2 shows part of the half of the listed items, items logit and probability outcomes of students answering each question correctly. Large negative logit for person measure indicates that the students are less able while large positive logit indicates that students are more able. Items with low negative logit are considered easy for students to answer while large positive logit indicates that items are more difficult for students to answer. Table 2 shows the probability distribution of students with high and low ability of answering each final examination question correctly. For example, the probability of student 127G3F able to answer Question 4 is 81.91% compared to Question 2 at 68.78%. This indicates that his ability to answer Question 4 is higher than his ability to answer Question 2. Question 2 relates to the topic on "identifying various counting rules and concept of probability" while Question 4 is on "discrete and continuous random variable." Basically students find that Question 4 is slightly easier to attempt than Question 2.

Table-2. Probability of Students' Answering Examination Questions Correctly based on Item Difficulty

Student ID	**Item Measure**	Q4 **-0.31**	Q5 **-0.19**	Q 3 **-0.1**	Q1 **0.19**	Q2 **0.41**
127G3F	**1.2**	**81.91**	**80.06**	**78.58**	**73.30**	**68.78**
128G3F	**1.2**	**81.91**	**80.06**	**78.58**	**73.30**	**68.78**
6G1F	**0.9**	**77.03**	**74.84**	**73.11**	**67.04**	**62.01**
69G2M	**0.9**	**77.03**	**74.84**	**73.11**	**67.04**	**62.01**
74G3F	**0.9**	**77.03**	**74.84**	**73.11**	**67.04**	**62.01**
86G3M	**0.9**	**77.03**	**74.84**	**73.11**	**67.04**	**62.01**
90G3F	**0.9**	**77.03**	**74.84**	**73.11**	**67.04**	**62.01**
95G3F	**0.9**	**77.03**	**74.84**	**73.11**	**67.04**	**62.01**
96G3M	**0.9**	**77.03**	**74.84**	**73.11**	**67.04**	**62.01**
2G1F	**0.66**	**72.51**	**70.06**	**68.14**	**61.54**	**56.22**
3G1M	**0.66**	**72.51**	**70.06**	**68.14**	**61.54**	**56.22**
4G1M	**0.66**	**72.51**	**70.06**	**68.14**	**61.54**	**56.22**
40G1F	**0.66**	**72.51**	**70.06**	**68.14**	**61.54**	**56.22**

Figure 1 below shows the probability distribution chart for Q2 and Q4. The upper blue line represents the probability distribution of Q4 that relates to discrete and random continuous variable topic. From the chart, 88% of students are able to answer Q4 with a probability value of at least 0.5. This indicates that majority of the students are able to answer Q4 correctly. The lower red line represents the probability distribution of Q2 that relates to the probability and counting rules topic. Only 59% of the students are able to answer Q2 correctly with a probability of 0.5 or greater while 41% are able to answer Q2 correctly with a probability of less than 0.5.

Fig-1. Probability Distribution of Students' Performance

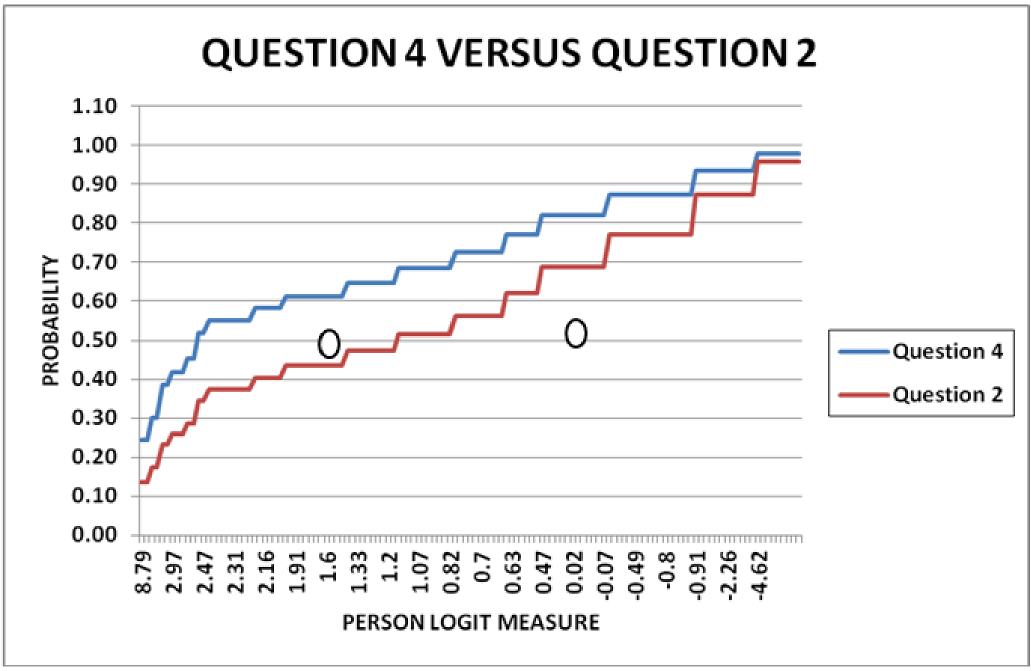
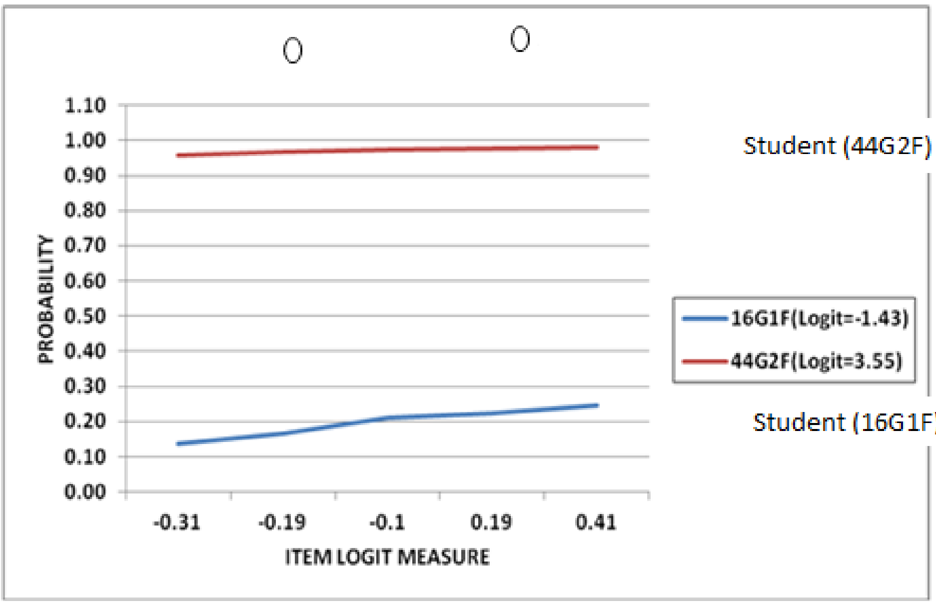


Figure 2 shows the probability distribution of two students answering Question 4 (Q4) in the final examination. The upper red line corresponds to the most able student 44G2F at logit 3.55. It can be seen that this student had shown a good and highly consistent performance in the question where the probability of answering Q4 correctly is close to 1.00. The lower blue line represents the least able student 16G1F at a logit of -1.43. From the graph, her probability of answering Q4 correctly lies between 0.14 to 0.25. This indicates that student 16G1F is a low performer with a minimum and maximum probability of answering Q4 correctly is 0.14 and 0.24, respectively.

Fig-2. Probability Distribution of Students' Performance



4.1. Fit Statistics

An item is considered as misfit when all three measures namely, point measure correlation (PTMEA), mean square (MNSQ) infit and outfit and standardized normal scores (ZSTD) are not located within the specified acceptable range (Linacre, 2006). In Table 3, the PTMEA for all questions is within the acceptable range of between 0.4 and 0.8 logit. For MNSQ, the acceptable range of fit is between 0.4 and 1.6. Overfit is indicated by a MNSQ less than 0.4 and underfit (noise) is indicated by a MNSQ more than 1.6. From the table, it can be seen that MNSQ infit and outfit for all questions lies within the acceptable range between 0.4 and 1.6. Final check based on ZSTD shows that all items fall within the acceptable range of -2.0 and 2.0. This indicates that all items (or questions) have specifically met the criteria of a quality data and no further review of items is required.

Table-3. Misfit Item Order for Final Examination Questions

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
4	488	120	-.31	.10	1.18	1.2	1.01	.1	A .59	49.2	47.9	Question 4
5	477	120	-.19	.10	1.17	1.2	1.06	.4	B .56	49.2	47.8	Question 5
3	467	120	-.10	.10	1.05	.4	.93	-.4	C .62	41.7	43.9	Question 3
2	405	120	.41	.09	1.01	.1	.94	-.4	b .64	36.7	35.9	Question 2
1	434	120	.19	.09	.82	-1.4	.92	-.5	a .65	38.3	41.9	Question 1
MEAN	454.2	120.0	.00	.09	1.05	.3	.97	-.2		43.0	43.5	
S.D.	30.5	.0	.26	.01	.13	1.0	.05	.3		5.3	4.5	

In minimizing response errors, analysis on person misfits is performed. Table 4 shows that students who are considered misfit as their logit values fall outside the acceptable range of fit statistics. For instance in the case of student 7G1F, it is a misfit based on the existence of negative point measure correlation -0.01 (< 0.4), infit and outfit MNSQ at 2.89 and 2.88, respectively (greater than 1.6) and ZSTD at more than two standard deviation. Misfits are due to several reasons which include guessing for answers and lack of knowledge in respective topics.

Table-4. Person Misfit Order for Final Examination Questions

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEA CORR.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
63	19	5	.47	.42	2.49	1.8	2.98	2.2	A-.64	40.0	35.7	63G2F
7	17	5	.15	.38	2.89	2.5	2.88	2.4	B-.01	.0	33.5	7G1F
10	17	5	.15	.38	2.31	1.9	2.47	2.0	C-.36	20.0	33.5	10G1M
19	8	5	-1.15	.48	1.95	1.2	2.44	1.5	D-.36	20.0	40.4	19G1F
96	21	5	.90	.51	2.13	1.3	2.38	1.5	E-.44	20.0	36.1	96G3M
49	18	5	.30	.40	1.85	1.3	2.25	1.7	F-.46	20.0	34.1	49G2M
29	16	5	.02	.37	1.94	1.6	2.08	1.8	G-.48	.0	31.1	29G1M
9	16	5	.02	.37	2.02	1.7	2.04	1.7	H-.20	20.0	31.1	9G1M
22	17	5	.15	.38	1.75	1.3	2.01	1.6	I-.46	.0	33.5	22G1F
30	11	5	-.64	.38	1.93	1.7	1.79	1.4	J .01	.0	18.3	30G1F

5. Discussion and Conclusion

This study had shown that true measurement of learning statistics can be measured effectively based on the Rasch Polytomous model. A logit scale ruler was developed based on the transformation of ordinal raw scores into interval scales through this model. The logit ruler was also used to measure students' ability in answering the final examination questions based on the calibration between person ability and item difficulty logits. Compared to the conventional way of assessing students' performances using raw scores, Rasch measurement model is able to provide details and information about students' ability and their level of understanding in certain topics where these abilities vary at different levels of probability values given the difficulty of the questions at certain levels (Nor Lide *et al.*, 2010). Probability analysis shows that at least 67% of the students have a good learning ability in probability and statistics since they are able to answer well in the final examination where their probability values of getting correct answers in the final examination questions is greater than 0.5. In contrast, about 33% of the students have poor learning ability where their probability of getting correct answers in the final examination is less than 0.5. The interpretation of these probabilistic outcomes is consistent with several other studies that used Rasch model as a tool to measure survey responses (Zamalia *et al.*, 2013).

Acknowledgment

We wish to extend our gratitude to Universiti Teknologi MARA for funding this research and our participation at the conference through the Lestari Grant iniatives (Grant No. 600-IRMI/DANA 5/3/LESTARI (0132/2016)). Our appreciation also goes to the respondents for their undivided support, cooperation and their willingness to participate in the study.

References

Ahmad, K., Mohamad, A. N. and Zainudin, S. (2010). An improved course assessment measurement for analyzing learning outcomes performance using Rasch model. *Procedia Social and Behavioral Sciences*, 18(2010): 442–49.

Ben-Zvi, D. (1999). Alternative assessment in Statistics education, 52nd Session of the International Statistical Institute. 3(1999): 175–76.

Bond, T. G. and Fox, C. M. (2007). *Applying the Rasch model, Fundamental measurement in the human sciences.* 2nd ednRoutledge Taylor & Francis Group.

Delmas, R., Garfield, J., Ooms, A. and Chance, B. (2007). Assessing students conceptual understanding after a first course in Statistics. *Statistics Education Research Journal*, 2(6): 28-58.

Garfield, J. and Delmas, R., 1991. "Students' conceptions of probability." In *3rd International Conference on Teaching Statistics, D. Vere-Jones (Ed.), Voorburg, The Netherlands: International Statistical Institute.* pp. 340-49.

Garfield, J. and Chance, B. (2000). Assessment in Statistics education, Issues and challenges. *Mathematical Thinking and Learning*, 2(2): 99–125.

Linacre, J. M. (2006). *WINSTEPS Rasch measurement computer program.* WINSTEPS. Com: Chicago.

Mills, J. D. (2004). Learning statistics concepts using simulation. *Educational Research Quarterly*, 4(28): 18-33.

Nor Lide, A. K., Nor-Zatul, I. I., Zamalia, M. and Mohamad Said, Z. (2010). Measuring students understanding of statistical concepts using Rasch measurement. *International Journal of Innovation, Management and Technology*, 1(2010): 13-19.

Osman, S. A., Naam, S. I., Jaafar, O., Wan Badaruzzaman, W. H. and Abdullah, R. A., 2012. "Application of Rasch model in measuring students' performance in civil engineering design II course." In *International Conference on Teaching and Learning in Higher Education, Procedia - Social and Behavioral Sciences.* pp. 59 – 66.

Osman, S. A., Wan Badaruzzaman, W. H., Roszilah, H., Kamarudin, A. B., Abdul Khalim, R. and Othman, J. (2011). Assessment on students performance using Rasch model in reinforced concrete design course examination. *Recent Researches in Education*, (2011): 193-98.

Saidfudin, M., Azrilah, A. A., Ridzwan, N. A., Omar, M. Z., Azami, Z. and Basri, H. (2010). Use of Rasch analysis to measure students performance in engineering education. *Latest Trends on Engineering Education*, (2010): 435-41.

Zamalia, M., Nor Azura, M. G. and Rosli, R., 2013. "Assessing students learning ability in a postgraduate statistical course, A Rasch analysis." In *2nd Cyprus International Conference on Educational Research, Social and Behavioral Sciences.* pp. 890 – 94.