A multiple regression model of statistical reasoning: A Malaysian context

Foo Kien Kheng [a], Noraini Idris [b], Ibrahim Mohamed [c] , Foo Su Lyn [d]
[a] Institute of Graduate Studies, University of Malaya, Kuala Lumpur, Malaysia.
[b] Universiti Pendidikan Sultan Idris 35900 Tanjong Malim Perak , Malaysia.
[c] Institute of Mathematical Sciences, Faculty of Science, University of Malaya, Kuala Lumpur.
[d] Nan Chiau High School 20, Anchorvale Link, Singapore.
Corresponding authour: kkfoo@siswa.um.edu.my

OIDA International Journal of Sustainable Development, Ontario International Development Agency, Canada
ISSN 1923-6654 (print) ISSN 1923-6662 (online) www.oidaijsd.com
Also available at http://www.ssrn.com/link/OIDA-Intl-Journal-Sustainable-Dev.html

Abstract: There is a noticeable slide in Malaysian mathematics achievement reported in the last three TIMSS studies and in particular in the area of Data and Chance. This is actually not only unique to Malaysia but also in many other countries in Asia and Africa. Recent studies have shown the influence of higher order thinking skills like reasoning and decision making on statistics achievement. Chan, Ismail and Sumintono (2014) found that statistical reasoning among Malaysian secondary school students to be poor. The purpose of this study is to determine the influence of language, and misconception on statistical reasoning using a sample size of 374 Diploma of Science students from a campus of a large Malaysian public university. A quantitative research design was employed as the objective of this study was to measure the strength and direction of the effect. The flexibility and power to analyze complex multivariate relationships concurrently are possible using a multivariate linear regression approach. The research procedure included a pilot study to determine the feasibility of the procedure and suitability of the adapted Statistical Reasoning Assessment (SRA) to the population of interest. A survey form was used to collect both primary and secondary data. The form comprised of items to collect respondent profile information, grades from relevant courses they took previously and self-reported grades of their mathematical achievement and language proficiency in the public examinations. The findings showed that students did not do well in statistical reasoning (SR) and had a substantially high level of misconception (MC) about statistics. SR (M = 38.17, SD = 13.83) and MC (M = 34.44, SD = 11.56). Language mastery (ENG) was found to be above average, (M = 3.26, SD = .73). The regression coefficients indicated that Language mastery (ENG), and Misconception (MC) significantly predicted Statistical Reasoning (SR). The best model generated was $SR = 47.07 - 0.59(MC) + 3.50(ENG)$. The coefficient of determination for the regression model was R^2 = 0.309 indicating that ENG and MC alone explained 30.90% of the total variance. Squared semi-partial correlation (sr^2) informs us of the unique variance explained by each of the variable. sr^2 for ENG is given by (.186 X .186 = .035) while MC is calculated by using (-.493 X -.493= .243). These indices showed that ENG and MC accounted for 3.5% and 24.3% respectively of the variances. This paper concludes with a discussion on the pertinent issues related to the administration of the SRA instrument and recommendation for further research in the field of language, statistical reasoning and misconception.

Keywords: Language, misconception, regression model, statistical reasoning

Introduction

Since its inception in 1995, Trends in International Mathematics and Science Study (TIMSS) - a four yearly studies, had shown that mathematical achievement for the 9 year-old and 13 year-old students in Malaysia was mediocre in comparison to other countries (Chan, Ismail & Sumintono (2014); International Association for the Evaluation of Educational Achievement (IEA), 2009). The 2011 TIMSS report (IEA, 2013) showed Malaysia's Eighth Grade mathematics result dropped 34 points from 474 to 440 in 2011 as compared to Singapore which recorded an increase of … In addition, Malaysia recorded a drop in Data and Chance component in the TIMSS study as compared to their performance in the other 3 components, i.e., Number, Algebra and Geometry. This worrying trend has been noted since 1999 (Gonzales et al., 2008). Hence, this phenomenon is a real cause for concern especially for the teaching and learning of Statistics. Furthermore, an analysis of the achievement in introductory Statistics for Diploma students in a Malaysian university showed a similar weakness in both statistical performance and reasoning (Zuraida, Foo, Rosemawati & Haslinda, 2012). The teaching and learning of statistical reasoning is crucial as there are studies showing that statistical reasoning has influence on students' achievement (Zuraida et al., 2012; Tempelaar, van der Loeff & Gijselaers, 2007)

Research have indicated that achievement and reasoning in statistics were directly predicted by a variety of cognitive and non-cognitive factors (Tremblay, Gardner & Heipel, 2000; Nasser, 2004; Chiesi & Primi, 2010). This raised the next question, 'What are the cognitive determinants that predict performance and reasoning?' This paper specifically looked at statistical reasoning and the determinants that influence it.

From a psychological angle, cognitive theories provide the framework to study these factors centering on the mental processing of information. The Information Processing Theory (IPT) is chosen to analyze, describe and elucidate these mental processes. Problem-solving and reasoning are skills that one develops so that one can act independently as adults. How students acquire reasoning and problem solving skills and how they acquire misconception are critical areas of study and a good and logical theory to explain the origin and acquisition of these skills has important educational and practical implications (Kalat, 2011). Many cognitive psychologists also believe that the Schema Theory too plays an important role in assisting them to understand the thinking and mental processes that go on in the brain (Anderson, 1997; Axelrod, 1973). Rumelhart believes that: '. . . schemata truly are the building blocks of cognition. They are the fundamental elements upon which all information processing depends. Schemata are employed in the process of interpreting sensory data (both linguistic and non-linguistic) in retrieving information from memory, in organizing actions, in determining goals and sub-goals, in guiding the flow of processing in the system.' (Rumelhart, 1980, pp 33).

Reasoning, noted Galotti (2008) involves cognitive processes that turn bits and bytes of data into useful information so that one can come to a conclusion. From a psychological perspective, reasoning is thought to be a mental process to derive inferences or conclusion from information known as premises. Garfield and Chance (2000) defined statistical reasoning as the way people reason with statistical ideas and make sense of the information. Statistical reasoning is a crucial cognitive skill to master and it is related to the content knowledge of the students.

Misconception is a construct that is closely related to reasoning. Misconceptions are systemic conceptual errors caused by underlying contrary beliefs and principles deeply ingrained in the students' cognitive structures (Olivier, 1989). Lèonard and Sackur-Grisvard (1987) provided a succinct explanation of the persistency of misconceptions among novices and even experts that is very difficult to correct. They went on to say that misconceptions are so stable that the carry over effect on reasoning is great. From an IPT point of view, reasoning rely very much upon the thought process and thereby causing the internal information to run into problems that sometimes give rise to misconceptions (Levitin, 2002).

The product of thinking is known as thoughts. Language is a medium for a person to communicate one's thoughts through the use of complicated rules that helps to form and string together symbols thus generating meaningful sentences or utterances. Thoughts and language are two closely related cognitive processes that are dynamic and complex. Language facilitates and expresses those thoughts through sound and symbols (Bransford, Brown & Cocking, 1999). Girotto (2004) asserted that much of the difficulty of reasoning lies with understanding the language. Reed (2011) noted that organization of the text in an item or the story structure has an effect on performance. Shaughnessy (1992) added if the context of the test item is abstract, the achievement on this item is much lower but if put into familiar context the success rate increased significantly. The mathematical language that is employed in test items also influence the success rate in solving reasoning tasks.

Methodology

Sample and Data Collection

A sample of 374 second-year Diploma of Science students in two branch campuses of a large Malaysian public university. The respondents took a course in Statistics for Technology 1. This course covers basic topics like descriptive and inferential statistics as well as elementary probability theory. Before the actual study began, a trial version of the SRA was distributed to a small sample of diploma students. The piloting of the instrument was carried out twice. After administrating the instrument, a focus group of 10 students who took the test, was formed. The students were selected based on a set of criteria to ensure maximum output from the group discussions. A set of guided questions were used in the group discussions. The two meetings lasted for about 45 minutes each. The transcriptions were then analysed by comparing similarities and differences of their responses. The group discussions were carried out using both English and Bahasa Melayu (Malay language), their first language. After some improvements to Garfield's (2003) instrument, it was rerun again. The data collected from the second pilot testing was used to further improve some of the items. Based on these initial findings, a few of the items in the SRA were further modified to suit local needs.

There are three variables used in this study. They are Statistical Reasoning (SR), Misconception (MC), and language (ENG).The instrument to measure Statistical Reasoning (SR) and Misconception (MC) was the adapted SRA (Garfield, 2003) and finalized after two rounds of pilot testing (see Appendix 1).

Measures and Instrument

The original SRA consisting of 20 multiple choice items is used to evaluate students' comprehension in basic concepts with a focus on their statistical reasoning skills. Each item in SRA describes a statistics or probability problem, both correct and incorrect. Students were instructed to select the response that best matches their own thinking about each problem.

The adapted version of the SRA investigated 6 correct reasoning subscales (CC1-CC6) and 5 misconceptions subscales (MC1-MC5). The scoring for each subscale was calculated by dividing the total number of correct/incorrect responses by the number of items in that subscale. The score ranges between 0 and 1.

Data Analysis Procedure: Multiple Linear Regressions

The research methodology is based on an analysis using multiple regression modelling. The model attempts to describe the relations between an outcome variable and some selected response variables. In this study the outcome variable is Statistical Reasoning (SR), while the response variables are: Misconception (MC) and Language mastery (ENG)

Many multivariate methods are based on the assumption that the data has a multivariate normal distribution. Shapiro-Wilks test and chi-square plot were used to check the assumption of normality. The probability value for Shapiro-Wilks must be more than 0.05 and the skewness value ±1. In order to see which hypothesis can be accepted the test for significance of regression (ANOVA) was carried out. If the observed value of F is large, then at least one variable differs. Statistical tests on individual regression coefficients were assessed. If *p*-value is less than 0.05, the correlation is considered significant.

Finding

This analysis used 374 samples from Diploma of Science students who took Statistics for Technology 1 course. Of particular interest is how such factors as Misconception (MC) and language mastery (ENG) have impact on Statistical Reasoning (SR) and the interactions that occur among them. Some summary statistics are given in Table 1 for each variable involved.

The students showed good mastery of the English language (ENG) at the time of the study (M = 3.26, SD = .73). However, the respondents achieved a moderately low level of mastery in Statistical Reasoning (SR) (M= 38.17, SD = 13.83) with a significantly high level of Misconception (MC) about statistics (M = 34.44, SD = 11.56). The low scores for both SR and MC are not surprising as the trend is almost similar in other studies in Malaysia or other parts of the world (Garfield, 2003; Tempelaar, 2004, 2006; Zuraida et al, 2012).

Table 1: Descriptive Statistics

	Mean	Std. Deviation	*N*
Statistical Reasoning	38.17	13.83	374
English Language	3.26	.73	374
Misconception	34.44	11.56	374

A correlation matrix (Table 2) was obtained to show the correlations between the selected variables. The correlation between SR and MC is moderately low with a negative sign $(r = -.525,\ p < .01)$. This implied an inverse relationship between the variables meaning that students with high misconception level do not do well in statistical reasoning. Similarly the correlation between MC and ENG is weak but significant with an inverse relationship $(r = -.170,\ p < .01)$. This indicates that those students with weak mastery of the English language suffers from high level of misconception. The association between MC and ENG on the other hand, was found to be positive $(r = .270, p < .01)$.

Table 2: Correlation Matrix

		Statistical Reasoning	English Language	Misconception
Pearson Correlation	Statistical Reasoning	1.000	.270**	-.525**
	English Language	.270**	1.000	-.170**
	Misconception	-.525**	-.170**	1.000
Sig. (1-tailed)	Statistical Reasoning	.	.000	.000
	English Language	.000	.	.000
	Misconception	.000	.000	.

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Assumption Checks

Figure 1 shows a random distribution of data points. Thus it can be concluded there exist linearity, homoscedasticity and normality of residuals. Furthermore, the histogram in Figure 2 provide further evidence that the normality assumption is complied with. Table 2 shows that all the correlation values are less than 0.7 among the independent variables, indicating that multicollinearity does not exist. Furthermore Durbin-Watson statistic of 1.80 suggests that multicollinearity is not a problem (Hair, Anderson, Tatham, & Black, 1998) (see Table 3).

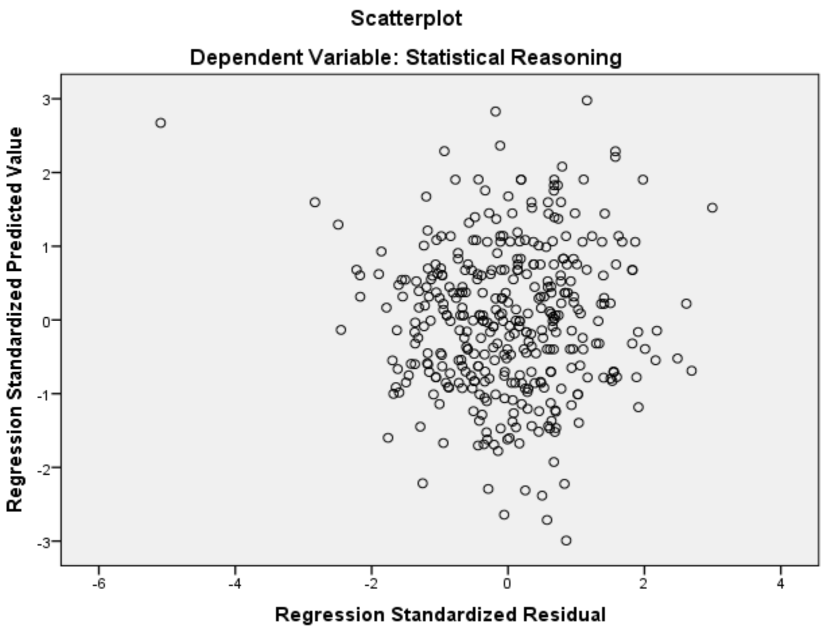

Figure 1: Scatterplot on zpred versus zresid to Check for Normality, Linearity, Homoscedasticity and Independence.

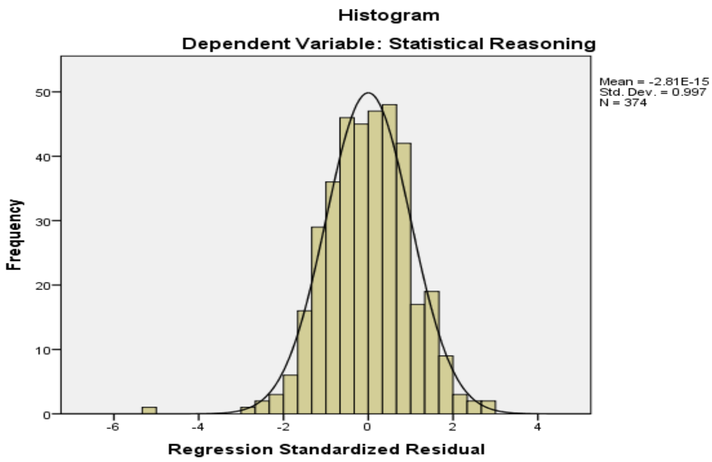

Figure 2: Histogram on Statistical Reasoning.

Table 3: Summary statistics on R square and Adjusted R square

Model Summary[b]

					Change Statistics					
Model	***R***	***R* Square**	**Adjusted *R* Square**	**Std. Error of the Estimate**	*R* Square Change	*F* Change	*df1*	*df2*	Sig. *F* Change	**Durbin-Watson**
1	.556[a]	.309	.305	11.526	.309	82.957	2	371	.000	1.800

a. Predictors: (Constant), Misconception, English Language

b. Dependent Variable: Statistical Reasoning

The output from Table 3 indicated that approximately 30.90% of the variance of Statistical Reasoning $\left(R^2 = 0.309, \text{Adjusted } R^2 = 0.305\right)$ could be attributed to MC and Language factors and the ANOVA table (Table 4) showed that the Model 1 was statistically significant $\left(F_{2,371} = 82.957, p < 0.01\right)$ and comparing the R square and the Adjusted R square, there is a shrinkage of $0.309 - 0.305 = 0.004$ or .4% which is very small. This is taken to mean that the model is generalizable using this sample (Hair et al., 1998). The effect size (*ES*) for multiple regression is given by $f^2 = \frac{R^2}{1 - R^2}$ (Cohen, 1992). This gives an *ES* = 0.447 which is a moderately strong effect.

Table 4: Analysis of Variance

ANOVA[a]

Model		**Sum of Squares**	***df***	**Mean Square**	***F***	**Sig.**
1	Regression	22042.338	2	11021.169	82.957	.000[b]
	Residual	49288.594	371	132.853		
	Total	71330.932	373			

a. Dependent Variable: Statistical Reasoning

b. Predictors: (Constant), Misconception, English Language

Best model for the regression analysis

In conclusion, the general model takes the form of:

$$Y = B_0 + B_1 x_1 + B_2 x_2$$

where Y= Statistical Reasoning (SR)
x_1= Misconception (MC)
x_2= Language (ENG)

The regression model is: $SR = 47.07 - 0.59(MC) + 3.50(ENG)$ with only MC $(b = -.590, SE_b = .052, \beta = -.493, p < .001)$ and ENG $(b = 3.497, SE_b = .825, \beta = .186, p < .001)$ being significant contributors to SR (Model 1 in Table 5). Observe the standardized coefficients (β) for the two factors above, MC contributed over two and a half times more to the total variance as compared to ENG implying greater importance of MC as compared to ENG in influencing the outcome of SR. Squared semi-partial correlation (sr^2) informs us of the unique variance explained by each of the variable (Hair et al., 1998). This index is calculated using the Part column under Correlations list of Table 5 for the variables concerned. sr^2 for ENG is given by (.186 x .186 = .035) while MC is calculated by using (-.493 x -.493 = .243). This is interpreted as ENG and MC uniquely accounted for roughly 3.50% and 24.30% respectively to the variance of SR.

Table 5: Summary statistics on unstandardized and standardized coefficients

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	*t*	**Sig.**	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound
1	(Constant)	47.072	3.537		13.308	.000	40.117	54.028
	English Language	3.497	.825	.186	4.241	.000	1.876	5.119
	Misconception	-.590	.052	-.493	-11.263	.000	-.693	-.487

a. Dependent Variable: Statistical Reasoning

Conclusion and recommendation

The main objective of this paper was to determine the impact of cognitive factors like language and misconception on statistical reasoning in a Malaysian context. Findings indicated that misconception (MC) and Language (ENG) are significant predictors of Statistical Reasoning (SR) (see Table 3). This result concurred with previous studies by Chiesi & Primi, (2010); Lalonde & Gardner, (1993); Nasser, (2004); Tempelaar, (2006).

The coefficient of determination for the regression model was R^2 = 0.309 indicating that MC and ENG can explain 30.90% of the total variance. The standardized coefficient for MC is over two and a half times stronger than ENG indicating the contributions of these two variables to SR. In conclusion, a moderately low statistic of R^2 = 0.309 showed that is a construct like SR is influenced by not only cognitive factors but non-cognitive variables as the results clearly shown that 69.1% of the variance is unaccounted for.

According to Bransford et al. (1999), language does impact on reasoning to a certain extent. The results of this study showed that there was a significant effect of language on reasoning though not strong (r=0.156, p<.01). These findings are preliminary and need more research to explore this relationship using controlled experiments.

As for the misconception variable, it was found to have a strong effect on reasoning. The study reiterates that misconceptions of the students must not be taken lightly for it is generally high among the respondents and ignoring its role in statistics would have consequences on the outcome of their examination results. This study also showed that Statistical Reasoning Assessment (SRA) can be an effective tool to capture statistical misconceptions of students. Hence SRA can be given to students on the first day of any statistics course and the misconception scores calculated. The scores would provide a good indication of the misconception problem of the learners.

This study provides some evidence that students' statistical reasoning in class is a complex construct that has many dimensions to it. Studies have shown many cognitive and non-cognitive determinants like student previous course of study, their grade point average, language skills, self-efficacy, student's attitude towards statistics or student perception of statistics as a tough subject are responsible (Lalonde & Gardner, 1993; Chang & Cheo, 2012). Further research is also recommended to look into new methodological approaches to study the variables here using additional instruments like Comprehensive Assessment of Outcomes in a first Statistics course (CAOS) or the Quantitative Reasoning Quotient (QRQ). Future research may want to study them using a different paradigm like qualitative research methodologies where in-depth examination of these few determinants across cultures and creed using the diversity in this country to the best of its advantage.

References

[1] Anderson, R.C. (1997). The notion of schema and the educational enterprise: General discussion of the conference. In R.C. Anderson, J. Spiro, & W.E. Montague (Eds.), *Schooling and the Acquisition of knowledge*. Hillsdale: Erlbaum.

[2] Axelrod, R. (1973). Schema theory: An information processing model of perception and cognition.*American political science review, 67*(4), p. 1248-1266

[3] Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (1999). *How people learn: brain, mind, experience, and school*. Washington, DC: National Academy Press.

[4] Chang D. W. & Cheo, R. K. (2012). Determinants of Malaysian and Singaporean economics undergraduates' academic performance. International Review of Economics Education, 11(2), 7-27.

[5] Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students' achievement. *Statistics Education Research Journal*, 9(1), 6-26.

[6] Cohen, J. (1992) Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (1), 155-159.

[7] Galotti, K. M. (2008). *Cognitive psychology: In and out of the laboratory* (4th ed.). Singapore: Thomson Wadsworth.

[8] Garfield, J. B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1), 22-38, Retrieved from http://www.stat.auckland.ac.nz/%7Eiase/serj/SERJ2(1).pdf.

[9] Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematics Thinking and Learning*, *2*(1-2), 99-125.

[10] Girotto, V. (2004). Task understanding. In J. P. Leighton & R. J.Sternberg (Eds.), *The nature of reasoning* (pp. 103–125). New York: Cambridge University Press.

[11] Gonzales, P., Guzmán, J.C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., and Williams, T. (2004). Highlights from the trends in international mathematics and science study (TIMSS) 2003 (NCES 2005-005). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

[12] Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C. (1999) *Multivariate data analysis* (5th ed.). Upper Saddle River, New Jersey: Prentice Hall.

[13] Kalat, J.W. (2011) *Introduction to Psychology* (9^{th} Ed.). Wadsworth Publishing.

[14] International Association for the Evaluation of Educational Achievement (IEA). (2009). *Trends in International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: IEA TIMSS & PIRLS International Study Center.

[15] International Association for the Evaluation of Educational Achievement (IEA). (2013). *Trends in International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: IEA TIMSS & PIRLS International Study Center. Retrieved from http://timssandpirls.bc.edu/timss2011/international-results-mathematics.html.

[16] Lalonde, R. N., and Gardner, R. C. (1993). Statistics as a second language? A model for predicting performance in psychology students. *Canadian Journal of Behavioural Science, 251, 108-125.*

[17] Lèonard and Sackur-Grisvard (1987). Necessity of a triple approach of erroneous conceptions of students, example of the teaching of relative numbers: Theoretical analysis. In J.C. Bergeron (Ed.). *Proceedings of the Eleventh International Conference for the Psychology of Mathematics Education* (Vol. 2, pp. 444-448). Montreal.

[18] Levitin, D.J. (2002). *Foundations of cognitive psychology: Core readings.* Bradford Book.

[19] Nasser, F. M. (2004). Structural model of the effects of cognitive and affective factors on the achievement of Arabic-speaking pre-service teachers in introductory statistics. *Journal of Statistics Education*, *12*(1), 1-28.
[20] Olivier, A. (1989). Handling pupils' misconception. *Pythagoras*. 21, 10-19.
[21] Plotnik, R. and Kouyoumdjian, H. (2011) *Introduction to Psychology*, (9th Ed). Wadsworth Publishing.
[22] Reed, D. K. (2011). A review of the psychometric properties of retell instruments. *Educational Assessment*, 16, 123–144.
[23] Rumelhart, D. E. (1980). Schemata: The building blocks of cognition, in R. T. Spiro, B. C. Bruce and W. F. Brewer (eds.), *Theoretical Issues in Reading Comprehension*, Erlbaum, Hillsdale, NJ, pp. 33-34.
[24] Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In A. Grouws (Ed.), *Handbook of research on mathematics teaching and* learning (pp. 465-494). New York: MacMillan Publishing Company.
[25] Tempelaar, D. (2004). Statistical reasoning assessment: An Analysis of the SRA instrument. In Proceedings of the ARTIST Roundtable Conference, Lawrence University. Retrieved from http://www.rossmanchance.com/artist/Proctoc.html.
[26] Tempelaar, D. (2006). A structural equation model analyzing the relationship students' statistical reasoning abilities, their attitudes toward statistics, and learning approaches. In Proceedings of the 7th International Conference on the Teaching of Statistics. Retrieved from http://iase-web.org/documents/papers/icots7/2G3_TEMP.pdf
[27] Tempelaar, D. T., van der Loeff, S., and 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*, Retrieved from http://www.stat.auckland.ac.nz/serj
[28] Tremblay, P. F., Gardner, R. C., & Heipel, G. (2000). A model of the relationships among measures of affect, aptitude, and performance in introductory statistics. *Canadian Journal of Behavioral Science,* 32, 40-48.
[29] Zuraida, Jaafar, Foo, Kien-Kheng., Rosemawati Ali. & Haslinda, Abdul Malek. (2012) Cognitive factors influencing statistical performance of diploma science students: A structural equation model approach. In Proceedings of Langkawi Conference-ICSSBE. (pp 562-566)

Appendix 1: Statistical Reasoning Assessment

Direction: Please do not answer in this question paper. Choose the best option

1) A small object was weighed on the same scale separately by nine students in a science class. The weights (in grams) recorded by each student are shown below.
   6.2 6.0 6.0 15.3 6.1 6.3 6.2 6.15 6.2
   The students want to determine as ACCURATELY as they can the actual weight of this object. Of the following methods, which would you recommend they use?
   a) Use the most common number, which is 6.2.
   b) Add up the 9 numbers and divide by 9.
   c) Throw out the 15.3, add up the other 8 numbers and divide by 8.
   d) Sort the 9 numbers from smallest to largest and choose the middle number.

2) The following message is printed on a bottle of prescription medication:
   WARNING: For applications to skin areas there is a 15% chance of developing a rash. If a rash develops, consult your physician.
   Which of the following is the best interpretation of this warning?
   a) Don't use the medication on your skin, there's a good chance of developing a rash.
   b) For application to the skin, apply only 15% of the recommended dose.
   c) If a rash develops, it will probably involve only 15% of the skin.
   d) About 15 of 100 people who use this medication develop a rash.
   e) There is a very low chance of getting a rash using this medication.

3) Mr Ahmad, a meteorologist wanted to determine the accuracy of his weather forecasts. Past records showed that forecast of 70% chance of rain have been reported. He compared these forecasts to records of actual rainy days for a particular period.
   The forecast of 70% chance of rain can be considered very accurate if it rained on:
   a) 95% - 100% of those days.
   b) 85% - 94% of those days.
   c) 65% - 74% of those days.
   d) 55% - 64% of those days

4) A teacher wants to know how many questions her students ask per day. A record of the number ofquestions asked by her 8 students on a particular day is shown below.

Student	Ali	Rashid	Shila	Jeya	Chia	Nik	Jamal	Nurul
Number of comments	0	5	2	22	3	2	1	2

   She wants to summarize this data by computing the TYPICAL number of questions made that day. Of the following methods, which would you recommend she use?
   a) Use the most common number, which is 2.
   b) Add up the 8 numbers and divide by 8.
   c) Throw out the 22, add up the other 7 numbers and divide by 7.
   d) Throw out the 0, add up the other 7 numbers and divide by 7.
   e) Sort the 8 numbers from smallest to largest and choose the middle number.

5) Two containers, labelled A and B are filled with red and blue marbles in the following quantities:

Container	Red	Blue
A	6	4
B	60	40

   Each container is shaken vigorously. After choosing one of the containers, you will reach in and, without looking, draw out a marble. If the marble is blue, you win RM50. Which container gives you the best chance of drawing a blue marble?
   a) Container A (with 6 red and 4 blue).

b) Container B (with 60 red and 40 blue).
c) Equal chances from each container.

6) Which of the following sequences is most likely to result from flipping a fair coin 5 times?
a) H H H T T
b) T H H T H
c) T H T TT
d) H T H T H
e) All four sequences are equally likely.

7) Select one explanation for the answer you gave for question 6 above.
a) Since the coin is fair, you ought to get roughly equal numbers of heads and tails.
b) Since coin flipping is random, the coin ought to alternate frequently between landing heads and tails.
c) Any of the sequences could occur.
d) If you repeatedly flipped a coin five times, each of these sequences would occur ABOUT as often as any other sequence.
e) If you get a couple of heads in a row, the probability of tails on the next flip increases. Every sequence of five flips has exactly the same probability of occurring.

8) Listed below are the same sequences of H's and T's that were listed in question 6. Which of the sequences is least likely to result from flipping a fair coin 5 times?
a) H H H T T
b) T H H T H
c) T H T T T
d) H T H T H
e) All four sequences are equally unlikely.

9) Miss Sally wants to buy a new car, and she has narrowed her choices to model A or model B. She talked to three friends, one model A owner and two model B owners. She also consulted a Consumer Report about model A and B. The table below summarizes her findings.

Recommendations	**Model A**	**Model B**
Three friends	This friend was very dissatisfied and swore never to buy this model again	These two friends reported few mechanical problems and nothing major
Consumer report	Using a sample of 400, the report found that model A had fewer mechanical problems complaints	Model B was reported to have more mechanical problems as compared to model A

Which model would you recommend that she buy?
a) I recommend that she buy model B.
b) I recommend that she buy model A.
c) I would tell her that it didn't matter which car she buys. She may as well toss a coin to decide.

10) One face of a fair coin is painted black, and one face is painted white. The coin is tossed two times. Which of the following results is more likely?
a) Black side up in one of the tosses; white side up on the other toss.
b) Black side up on both tosses.
c) a and b are equally likely.

11) Half of all newborns are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 20% or more female births?
a) Hospital A (with 50 births a day).
b) Hospital B (with 10 births a day).
c) The two hospitals are equally likely to record such an event.

12) The town mayor of a small town wanted to determine the average number of children per household in his town. He divided the total number of children in the town by 50, the total number of households. Which of the following statements must be true if the average number of children per household is 2.2?
    a) Half the households in the town have more than 2 children.
    b) More households in the town have 3 children than have 2 children.
    c) There are a total of 110 children in the town.
    d) There are 2.2 children in the town for every adult.
    e) None of the above.

13) When two dice are simultaneously thrown it is possible that one of the following two results occurs: Result 1: a '5' and a '6' are obtained. Result 2: a '5' is obtained twice.
    Choose the option that correctly represents the possible outcomes for Result 1 and Result 2

Option	**Result 1**	**Result 2**
a)	{(5,6)}	{(5,5)}
b)	{(5,6), (6,5)}	{(5,5)}
c)	{(5,6), (5,5)}	{(5,6), (5,5)}
d)	{(5,6), (6,5)}	{(5,5), (6,6)}

14) Select the response that you agree with the most for question 13:
    a) The chance of obtaining each of these results is equal.
    b) There is more chance of obtaining result 1.
    c) There is more chance of obtaining result 2.
    d) It is impossible to give an answer.

15) When three dice are simultaneously thrown, which of the following results is MOST LIKELY to be obtained?
    a) Result 1: a '5', a '3' and a '6'.
    b) Result 2: a '5' three times.
    c) Result 3: a '5' twice and a '3'.
    d) All three results are equally likely.