THE USE OF A HIERARCHICAL CONSTRUCT TO INVESTIGATE STUDENTS' LEARNING OF INFERENTIAL STATISTICS

KRİSHNAN SARAS[1] & IDRIS Noraini[2]

[1]Institute of Graduate Studies, University Malaya, Malaysia

[2]Faculty of Education, University Malaya, Malaysia

Contact email: saraskrishnan@yahoo.com

ABSTRACT
At present, there is still a need for more research in the teaching and learning of inferential statistics because of the limitedness of literature in this area of statistics education. Moreover, there is continuing evidence of students' partial or unsuccessful learning of many aspects of inferential statistics. This is one of the concerns brought to attention in my postgraduate research whereby part of my work involved the development of a hierarchical construct to identify the different levels of students' learning of inferential statistics. This paper particularly discusses the use of this hierarchical construct to investigate the learning of inferential statistics among students.

Keywords: *hierarchical construct; students' learning; inferential statistics*

1. BACKGROUND OF STUDY

Statistics curriculum is formed based on its subject matter which is contextual data Schield (2004). The goals for statistics curriculum include developing and evaluating inferences based on data and using correct statistical methods to analyze data and make valuable interpretations (NCTM, 2000). Knowledge and skills of this sort which involves statistical inference has become an influential factor in the decision-makings of individuals, businesses, researchers, policy makers and governments. Inferential statistics is predominantly concerned with making inferences about the population by collecting, analyzing and interpreting sample data. Inferential statistics is a statistical content of utmost importance and is taught in almost all the mathematics courses.

I believe successful and logical inference requires students to know the required formulae to be able to perform the calculations and also the concepts of inferential statistics as suggested by Gal's (2002) model of statistical literacy whereby the knowledge elements constitute mathematical knowledge, statistical knowledge and contextual knowledge. Students must also be able to tell when and how a particular inferential procedure can be used. This compels the students to critically evaluate a statistical situation, decide on a suitable inferential method and make connections between the procedural steps involved in that test. More importantly, students must be competent enough to make informed decisions and valid conclusions with relation to the context of the statistical problem. These demands a deep and connected understanding of the many concepts of statistical inference in which students are lacking as found by Glaser (2003), and Kadijevich, Kokol-Voljic and Lavicza (2008).

The learning of inferential statistics is an issue that obligates our concern especially because research continues to show that students have difficulty in understanding many aspects and concepts of inferential statistics although to a large extent they can execute the rudimentary calculations. Students have difficulties in grasping the logic behind the different but related concepts which are a must for intelligent use of inferential tools (Francis, Kokonis & Lipson, 2007) and so are prone to commit misconceptions when dealing with inferential problems (Sotos, Vanhoof, Noortgate & Onghena, 2007). Students also have little sense of the rationale of statistical inference or how the concepts of inferential statistics can be applied to different real-life situations (Weinberg, Wiesner & Pfaff, 2010).

The issue of students' learning of inferential statistics in introductory statistics courses is addressed in my postgraduate research. In particular, I develop a hierarchical construct to assess students' learning of inferential statistics. This paper presents the hierarchical construct and discusses the reliability and the validity of the instrument used to develop this construct. A more detailed explanation on the development of the construct is provided in my postgraduate thesis. Second, this paper discusses students' learning of inferential statistics with particular focus on the mistakes they make and their lack of understanding of the concepts of inferential statistics.

2. A REVIEW ON RESEARCH IN INFERENTIAL STATISTICS

According to Sotos et al. (2007), research interest in statistical inference is fueled by three important factors. First, inferential statistics is pertinent for research development of empirical sciences. Second, it is an important topic in statistics courses for a majority of disciplines and third, students appear to be prone to misconceptions of different sorts in learning inferential statistics. Past studies on inferential statistics include hypothesis testing (e.g., Kaplan, 2009; Sotos et al., 2009; Weinberg et al., 2010), confidence intervals (e.g., Rossman, Chance & Obispo, 2004; Fidler, 2006; Canal & Gutierrez, 2010) and sampling distributions (e.g., Francis et al., 2007; Pfannkuch, 2008; Kadijevich et al., 2008). It appears that a number of research work in inferential statistics involve introductory statistics students. This is perhaps because inferential statistics is taught in almost all introductory statistics course.

Studies have dealt with problem representations in inferential statistics (e.g., Glaser, 2003), misconceptions in different areas of statistical inference (e.g., Bower, 2003; Kadijevich et al., 2008), unconventional teaching methods such as cooperative learning technique (e.g., Evangelista & Hemenway, 2002) and using computer programs to aid learning of inferential statistics (e.g., delMas & Liu, 2005). In addition, there has been discussion on sampling techniques (e.g., Evangelista & Hemenway, 2002), students' level of confidence (e.g., Rossman, Chance & Obispo, 2004), students' communication of inferential understanding (e.g., Evangelista & Hemenway, 2002) and informal inferential reasoning (e.g., Zieffler, Garfield, delMas & Reading, 2008).

An area of interest that has been gaining popularity in recent years especially with the reformation in statistics education is students' conceptual understanding of statistical inference (e.g., Alacaci, 2004; delMas & Liu, 2005; Francis et al., 2007; Pfannkuch, 2008). In fact, students' conceptual understanding has been a major concern in statistics education and research in statistics education. Most students are found to have incomplete understanding of statistical concepts (Pfannkuch, 2008) whether it is the procedural understanding (delMas & Liu, 2005) or the conceptual understanding (Evangelista & Hemenway, 2002). In addition, researchers found that students have difficulty in making connections among different concepts and in knowledge integration (e.g., Kadijevich et al., 2008). I find that research in inferential statistics mainly involve students' conceptual understanding (e.g., Alacaci, 2004; delMas & Liu, 2005; Francis et al., 2007; Pfannkuch, 2008) and on the misconceptions associated with them (e.g., Bower, 2003; Fidler & Cumming, 2005; Kadijevich et al., 2008; Sotos et al., 2009).

3. METHODOLOGY

3.1 OBJECTIVE AND SCOPE

The objective of this study is to use a hierarchical construct to identify students' learning of inferential statistics. Inferential statistics being one of the two branches of statistics is a broad area encompassing the central limit theorem, hypothesis testing and confidence intervals among others. In this study, the scope of inferential statistics is restricted to hypothesis test of one population mean. Meanwhile, the sample of 150 students for this exploratory study was taken from two universities in two different states in Malaysia. The students are also from four different programs that are two pre-university programs and two degree programs.

3.2 INSTRUMENTATION

The instrument used to develop the hierarchical construct is a task-based questionnaire with 10 main items and 21 items altogether. The questions are open ended to gather as much information as possible and to obtain diverse answers from the respondents. Analysis of the instrument particularly in terms of the reliability and validity of the instrument was carried out using the Partial Credit Model. The methodology adopted here is from the work of Watson and Callingham (2003) who developed the six levels statistical literacy construct. The statistical literacy construct is based upon two frameworks which are Biggs and Collis' (1982, 1991) SOLO Taxonomy and Watson's (1997) three-tiered statistical literacy. The work of Watson and Callingham (2003) and others who developed similar constructs (e.g., Watson, Kelly & Izard, 2005; Kaplan & Thorpe, 2010) also employed the Partial Credit Model.

The reliability of the instrument is ascertained by checking if the item reliability, the person reliability and the Cronbach's alpha values are greater than 0.70. For this study, the item reliability is 0.98, the person reliability is 0.79 and the Cronbach's alpha is 0.75. Hence, the reliability conditions have been met. In Rasch analysis, the validity is determined by examining two types of fit statistics that is the infit mean square statistics and the infit standard deviations. There are no set rules when it comes to the acceptable values of the fit statistics although there have been suggestions for acceptable ranges as discussed by Green and Frantom (2002). In this study, the values used by Watson and Callingham (2003) are adopted that is 0.77 to 1.30 for the infit mean square statistics and -2 to +2 for the infit standard deviations. The infit mean squares for the items in this instrument are between 0.86 and 1.15 while the infit standard deviations are between -1.6 and 1.4. Therefore, the validity of the instrument in this study has been established too.

Table 1. Reliability and fit indices

Item Separation Reliability	6.48
Item Infit Mean Square	1.00 (s.d. 0.08)
Person Separation Reliability	1.77
Person Infit Mean Square	1.03 (s.d. 0.33)
Cronbach Alpha	0.75

3.3 THE HIERARCHICAL CONSTRUCT

The Rasch analysis generated item separation reliability value of 6.48 as shown in Table 1 suggesting six hierarchical levels. The item-person map generated by the Rasch analysis was divided into six strata as suggested by the value of the item separation reliability. The items were grouped together according to their attributes and their difficulties levels. The descriptors for each level was then formed and described as shown in Figure 1. The levels are in ascending order whereby the lowest level (Level 1) deals with knowledge of inferential terminologies and symbols while the highest level (Level 6) involves complete knowledge of hypothesis decision making, communicating understanding of hypothesis testing procedure and process, and knowledge of underlying inferential and hypothesis testing concepts.

Levels	Descriptors
Level 1	ability to identify inferential terminologies and symbols when presented in contextual form.
Level 2	demonstrates procedural knowledge of the hypothesis testing procedure in decision making.
Level 3	ability to understand and use inferential terminologies and symbols when presented in contextual form, demonstrates knowledge of basic steps of the hypothesis testing procedure in decision making.
Level 4	demonstrates knowledge of different steps of the hypothesis testing procedure in decision making, ability to infer in simple context, able to communicate understanding of hypothesis testing procedure and process, demonstrates superficial knowledge of underlying inferential and hypothesis testing concepts.
Level 5	demonstrates understanding of random sampling, ability to infer in simple and more complex contexts, able to communicate understanding of hypothesis testing procedure and process in contextual form, demonstrates some knowledge of underlying inferential and hypothesis testing concepts.
Level 6	demonstrates complete knowledge of hypothesis decision making, successfully communicates complete understanding of hypothesis testing procedure and process, demonstrates knowledge of underlying inferential and hypothesis testing concepts.

Figure 1. The hierarchical construct

Table 2 shows the number and percentages of students in each level of the hierarchical construct. The highest percentages for Level 3 and Level 4, and the lower percentages for the other levels correspond to a normal distribution shape for the respondents on the item-person map. This is interpreted as most students will have average ability whereas a small proportion of students have only fundamental understanding. Likewise, few students will have a higher ability when dealing with inferential problems.

Table 2. Number and percentage of respondents

Level	Number of respondents	Percentage
Level 1	12	8%
Level 2	22	14.67%
Level 3	50	33.33%
Level 4	50	33.33%
Level 5	8	5.33%
Level 6	8	5.33%

More specifically, from the sample of 150 respondents, 75.33% successfully and completely answered the items in Level 1. The percentages get smaller as the level gets higher as expected. In detail, 68% of the students were successful in Level 2, 38% in Level 3, 1.33% in Level 4, 0.67% in Level 5 and another 0.67% in Level 6. Further, 73.45% of the students who gave complete answers for the items in Level 1 succeeded in giving complete answers for the items in Level 2. Meanwhile, 37.35% successfully proceeded from Level 2 to Level 3 and 3.23% proceeded from Level 3 to Level 4. Of the 1.33% of the sample of students who managed to answer the items in Level 4 completely, none was able to successfully answer the items in the subsequent levels that are Level 5 and Level 6.

4. STUDENTS' LEARNING OF INFERENTIAL STATISTICS

With reference to the descriptors of the levels in Figure 1, the higher levels involve:

(a) understanding of random sampling.
(b) ability to infer in contexts.
(c) communication of hypothesis testing procedure and process.
(d) complete knowledge of hypothesis decision making.
(e) complete knowledge of underlying inferential and hypothesis testing concepts.

In the following paragraphs, I discuss these components in more detail with respect to the responses gathered in this study.

(a) Understanding of random sampling.

Some of the respondents not only do not understand random sampling but cannot differentiate a sample from a population as well especially when presented in contextual form. I suspect that in most situations, students remember the term but do not know exactly what it means.

For instance, consider the following item in the questionnaire:

The domestic department reports that the mean monthly expenditure for families of four in Malaysia is RM4500. A group of consumers is interested to know if this claim is true. The consumer group took two hundred families of four in one of the states as their sample for this hypothesis testing. Give reason why this may not be a good sample.

One of the students' response was "*because the sample will not have an accurate amount of salary*". Clearly, the student does not even have the basic idea of sampling and that different samples can result in different sample means. Meanwhile, another student's response was "*the sample is not random because only involve one state whereas the claim is about whole Malaysia*". In my opinion, although the idea of random sample is suggested but the student's response in general fail to convince that the student has a clear understanding of random sampling.

(b) Ability to infer in contexts.

Students basically make a surface level inference that is they make a statement about the population mean in the hypothesis question but do not relate the mean value to the problem situation. In other words, they are not inferring in context. Sotos et al. (2007) in citing earlier researchers stated that one of the two aspects in most misconceptions about the learning of hypothesis test is the interpretation of the results.

For instance, consider the following item in the questionnaire:

In a winery in Spain, the actual volume filled into the champagne bottles varies slightly from bottle to bottle. Pepito found that the mean volume of champagne from a random sample of two hundred bottles is 371 ml. Consider the hypothesis statements $H_0: \mu = 375, H_a: \mu \neq 375$. Say that Pepito's hypothesis testing results in the decision "*do not reject H_0*". What conclusion does Pepito make?.

One of the students' response was "*there is insufficient evidence to indicate the population mean is not 375ml*". Although the interpretation is correct but this student is not interpreting in the context of the hypothesis problems.

(c) Communication of hypothesis testing procedure and process.

According to Batanero (2000), understanding the concept of a hypothesis test is one of the three most difficult aspects of statistical hypothesis testing. In fact, there are elements in every stage of hypothesis testing that students misunderstand. For the same item mentioned in part (b) above, when the students were asked to explain the hypothesis test, one of the responses was "*to prove that the population sample of 200 bottles is 375ml is wrong*". Firstly, the student seems to have no idea of population and sample. Second, the student has the wrong understanding that hypothesis test is conducted to make a decision about the sample mean. Also, the student seems to think that a hypothesis test is conducted with the purpose of rejecting the null hypothesis.

(d) Complete knowledge of hypothesis decision making.

There are two problems in the respondents' hypothesis decision making:

(i) when the test statistic is in the region of rejection, respondents make the decision to reject the null hypothesis but do not further say that the alternate hypothesis is accepted.

For instance, for the item:

Consider the hypothesis statements $H_0: \mu = 32, H_a: \mu \neq 32$. What hypothesis decision will you make based on a z-score of 2.05?, a number of students just said "*Reject* H_0".

(ii) when the test statistic is not in the region of rejection, respondents make the decision to accept the null hypothesis. In this case, the students believe that null hypothesis is proved to be true or false that is they view hypothesis test as a mathematical proof instead of a probabilistic proof (Sotos et al., 2007).

For instance, for the question mentioned in part (c), one of the responses given was "*to find out if the null hypothesis is accepted*". This response clearly shows that student has the incorrect understanding that the null hypothesis can be accepted.

(e) Complete knowledge of underlying inferential and hypothesis testing concepts.

The respondents in this study showed lack of knowledge of fundamental concepts of hypothesis testing. Moreover, they also have limited understanding of the regions of rejection, the P-value and the significance level. For instance, one of the students wrote "*the P-value is in the region of rejection*". It could be a problem with language but it could also be that this student does not know how P-values and region of rejections are related. This observation is supported by earlier research work (e.g., Haller & Krauss, 2002; Sotos et al., 2007; Schneiter, 2008) who found that students have difficulty understanding the meaning of related concepts in hypothesis test such as P-values. In fact, the P-value and the significance level are found to be the most complicated concepts associated to hypothesis testing results (Haller and Krauss, 2002). I agree with Schneiter (2008) that students remember "reject H_0 if $p < 0.05$" but do not attempt to understand it. Some students are only intent on getting the correct answers and getting good grades but are not much concerned with the concepts underlying the calculations. On the other hand, in some situations, they teachers do not go beyond simply teaching the students how to use the P-value perhaps due to the constraint of the course curriculum or time.

It is crucial that instead of focusing on the procedural skills in hypothesis test, students are made to understand the underlying concepts related to the procedural steps. For instance, when teachers tell the students that null hypothesis is rejected when P-value is smaller than the significance value, then they must explain to the students why this is so and how it is related to the regions of rejection. A good way to promote students' understanding as well as communication skills is to get them to present their results. For example, in a class that is relatively small say twenty students, give each student one real-life hypothesis testing situation and get them to analyze it and present the results to the class. This method can be used in larger classes as well but it may take a few lessons to complete it. Evangelista and Hemenway (2002) found that group activities can also enhance learning of statistics. One such technique is the use of jigsaw where the students are divided into groups and given different tasks. Then, at least one member of each group is put together in a new group and they put together the pieces of their task to form a complete picture. I think this technique is especially workable for teaching hypothesis test because the different steps of the hypothesis testing procedure are independent from each other. Besides that, the use of technology for investigation, simulation and illustration can increase students' interest and understanding of statistics (Schneiter, 2008). For example, the P-value applet and the Chi-square applet created by Schneiter (2008) are designed to be simple and easy-to-use in teaching hypothesis testing.

ACKNOWLEDGEMENT

I would like to thank the authors and researchers whose work I have referred to in writing this paper.

REFERENCES

Alacaci, C. (2004). Inferential statistics : Understanding expert knowledge and its implications for statistics education. *Journal of Statistics Education*, 12(2).

Batanero, C. (2000). Controversies around the role of statistical tests in experimental research. *Mathematical thinking and learning*, 2(1&2), 75-97.

Biggs, J.B. & Collis, K.F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. Academic Press: New York.

Biggs, J.B. & Collis, K.F. (1991). Multimodal learning and the quality of intelligent behaviour. *Intelligence: Reconceptualization and measurement*, 57-76.

Bower, M.D. (2003). Some Misconceptions about Confidence Intervals. *Six Sigma Forum-American Society for Quality*.

Canal, G.Y. & Gutierrez, R.B. (2010). The Confidence Intervals : A Difficult Matter, even for Experts. *Proceedings of the 8th International Conference on Teaching Statistics, ICOTS-8.* The Netherlands : ISI.

delMas, R.C. & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, 4(1), 55-82.

Evangelista, F. & Hemenway, C. (2002). The Use of the Jigsaw in Hypothesis Testing. *Proceedings of the 2nd International Conference on the Teaching of Mathematics at the Graduate Level*. Hersonissos, Crete, Greece.

Fidler, F. (2006). Should psychology abandon p-values and teach CIs instead? Evidence-based reforms in statistics education. *Proceedings of the 7th International Conference on Teaching Statistics*. IASE.

Fidler, F. & Cumming, G. (2005). Teaching confidence intervals : Problems and potential. *Proceedings of the 55th International Statistics Institute Session*.

Francis, G., Kokonis, S. & Lipson, K. (2007). Enhancing Student Understanding in Statistical Inference-Assessing the Effectiveness of a Computer Interaction. *IASE*.

Gal, I. (2002). Adults' statistical literacy : Meanings, components, responsibilities. *International Statistical Review*, 70(1), p.1-51.

Glaser, R. (2003). Assessing Expert Knowledge Representations of Introductory Statistics. *CSE Tech. Rep. No. 600*. Los Angeles. University of Calirfornia, Center for Research on Evaluation, Standards, and Student Testing.

Green, K. E., & Frantom, C. G. (2002). Survey Development and validation with the Rasch Model. In *International Conference on Questionnaire Development, Evaluation, and Testing*. Charleston, SC.

Haller, H. & Krauss, S. (2002). Misinterpretations of significance : A problem students share with their teachers?. *Methods of Psychological Research*. 7(1), 1-20.

Kadijevich, D., Kokol-Voljic, V. & Lavicza, Z. (2008). Towards a Suitably Designed Instruction on Statistical Reasoning : Understanding Sampling Distribution with Technology. *Proceedings of the ICMI Study 18 Conference and IASE 2008 Round Table Conference*. International Commision on Mathematical Instruction & International Association for Statistics Education.

Kaplan, J.J. (2009). Effect of Belief Bias on the Development of Undergraduate Students' Reasoning about Inference. *Journal of Statistics Education*, 17(1).

Kaplan, J.J. & Thorpe, J. (2010). Post Secondary and Adult Statistical Literacy: Assessing Beyond the Classroom. *Proceedings of the 8th International Conference on Teaching Statistics, ICOTS-8.* The Netherlands : ISI.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA : NCTM

Pfannkuch, M. (2008). Building sampling concepts for statistical inference : A case study. *Proceedings of the 11th International Congress of Mathematics Education Proceedings* (pp. 6-13). Monterrey, Mexico.

Rossman, A., Chance, B. & Obispo, C. P. S. L. (2004). Anticipating and addressing student misconceptions. In *ARTIST Conference on assessment in Statistics, Lawrence University* (pp.1-4).

Schneiter, K. (2008). Two Applets for Teaching Hypothesis Testing. *Journal of Statistics Education*, 16(3).

Schield, M. (2004). Statistical literacy and liberal education at Augsburg College. P*eer Review*, 6(4), 16-18.

Sotos, C., Vanhoof, S., Noortgate, W. & Onghena, P. (2007). Students' misconceptions of statistical inference : A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2(2), 98-113.

Sotos, C., Vanhoof, S., Noortgate, W. & Onghena, P. (2009). How confident are students in their Misconceptions about Hypothesis Tests?. *Journal of Statistics Education*, 17(2).

Watson, J.M. (1997). Assessing statistical literacy using the media. *The Assessment Challenge in Statistics Education*, 107-121.

Watson, J.M. & Callingham, R.A. (2003). Statistical literacy : A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3-46.

Watson, J.M., Kelly, B.A. & Izard, J.F. (2005). Statistical Literacy over a Decade. *Building connections: Theory, research and practice*, 775-782.

Weinberg, A., Wiesner, E. & Pfaff, T.J. (2010). Using Informal Inferential Reasoning to Develop Formal Concepts : Analyzing an Activity. *Journal of Statistics Education*, 18(2).

Zieffler, A., Garfield, J., delMas, R. & Reading, C. (2008). A Framework to support research on Informal Inferential Reasoning. *Statistics Education Research Journal*, 7(2), 40-58.