Identifying and Evaluating Threshold Concepts in First Year Statistics courses at a large university in South Africa

Andre Swanepoel*
Department of Statistics, University of Pretoria, Pretoria, South Africa – Andre.Swanepoel@up.ac.za

Johann Engelbrecht
Department of Science, Mathematics and Technology Education, University of Pretoria, Pretoria, South Africa - Johann.Engelbrecht@up.ac.za

Ansie Harding
Department of Mathematics and Applied Mathematics, University of Pretoria, Pretoria, South Africa – Ansie.Harding@up.ac.za

Lizelle Fletcher
Department of Statistics, University of Pretoria, Pretoria, South Africa – Lizelle.Fletcher@up.ac.za

In the teaching of Statistics, certain central concepts/topics are experienced as more difficult to comprehend than others, especially within a group of students with diverse mathematical abilities. Misconception of such concepts/topics while studying Statistics on the 100 level, where the foundation of the discipline is laid, is problematic since it might prohibit the student from understanding and grasping the core concepts upon which the discipline is developed. These misconceptions will also influence the student`s future studies of the discipline since no proper holistic view of the inner mechanics of the different procedures and techniques nor the interrelatedness of the different procedures and techniques will be present. Failure to master these concepts/topics could also restrict the progression within a course since in all Statistics courses topics/concepts build onto one another. These concepts are referred to as *threshold concepts* where a threshold concept is a conceptual gateway that opens up a new and previously inaccessible way of thinking without which you cannot progress in the subject.

The purpose of this research is to identify the threshold concepts in 100 level Statistics at a large South African university in a three year longitudinal study, and to also determine their levels of difficulty (which describes how troublesome the concept is to master) and importance (which refers to how much follow up work is unlocked by mastering the threshold concept). A better understanding of the threshold concepts within the discipline of Statistics can give insight to educators on difficulties perceived by students which can be indicative to whether education models should be adapted and if so, how it must be adapted in order to improve the learning process, retention of knowledge and the throughput.

This study will seek to:

- verify which of the threshold concepts identified through literature are also experienced as threshold concepts among Statistics students on the 100 level,
- expand the existing list of already identified threshold concepts,
- determine the degree of importance for the observed threshold concepts as experienced by these students,
- determine the level of difficulty for the observed threshold concepts as experienced by these students.

The results discussed will be preliminary, based on data gathered in 2014 for the 100 level students of 2013 on whom the longitudinal study will be based. Data gathered additionally as part of a pilot study for the 2011 and 2012 first year students will also be analyzed along with the 2013 data.

Keywords: longitudinal study; interrelatedness; troublesome knowledge; misconceptions.

1. Introduction

Background

The study was conducted at a South African university within the Statistics Department, one of the largest in the country, comprising approximately 7000 undergraduate and approximately 150 postgraduate students. Statistics courses are presented on three different campuses and in two streams, one for science students who enroll for Mathematical Statistics (WST courses) and one for commerce students who enroll for Statistics (STK courses).

Problem Statement

In the teaching of Statistics, certain central concepts/topics are experienced as more difficult to comprehend than others, especially within a group of students with diverse mathematical abilities. Misconception of such concepts/topics while studying Statistics on the 100 level, where the foundation of the discipline is laid, is problematic since it might prohibit the student from understanding and grasping the core concepts upon which the discipline is developed. These misconceptions will also influence the student`s future studies of the discipline since no proper holistic view of the inner mechanics of the different procedures and techniques nor the interrelatedness of the different procedures and techniques will be present. Failure to master these concepts/topics could also restrict the progression within a course since in all Statistics courses topics/concepts build onto one another. This problem led to the research on threshold concepts.

According to Meyer and Land (2003) a threshold concept is similar to a portal, opening up a new and previously inaccessible way of thinking and it represents a transformed way of understanding, or interpreting, or viewing a concept without which the student cannot progress. A threshold concept can of itself represent “troublesome knowledge” – knowledge that is conceptually difficult to understand according to Perkins (1999). A threshold concept can also be characterised by the following characteristics: Transforming - *Shift in perception achieved*, Irreversible - *unlikely to be forgotten once understood*, Integrative - *exposes previously hidden interrelatedness*, Bounded - *conceptual space have terminal frontiers* and Troublesome - *conceptually difficult knowledge* Meyer and Land (2003). For the purpose of this study we define a threshold concept as a conceptual gateway that opens up a new and previously inaccessible way of thinking without which you cannot progress in the subject.

Pinpointing the different threshold concepts as well as the level of misconception at which these threshold concepts are present is complicated by the fact that the student population enrolled for Statistics on the 100 level at the three campuses of the university is diverse in terms of mathematical ability and background.

Purpose of Study

The purpose of this research is to identify threshold concepts in 100 level Statistics at a large South African university in a three year longitudinal study, and to also determine their levels of difficulty (which describes how troublesome the concept is to master) and importance (which refers to how much follow up work is unlocked by mastering the threshold concept). A better understanding of the threshold concepts within the discipline of Statistics can give insight to educators on difficulties perceived by students which can be indicative to whether education models should be adapted and if so, how it must be adapted in order to improve the learning process, retention of knowledge and the throughput.

This study will seek to:

- verify whether the data from the questionnaires support the literature regarding the information on already identified threshold concepts among Statistics students on the 100 level,
- check whether additional threshold concepts can be identified and added to the list,
- determine the importance of the observed threshold concepts,

- determine the level of difficulty for the observed threshold concepts.

2. Identifying the threshold concepts

In order to acquire information regarding threshold concepts a questionnaire was developed and distributed to different groups of students. The questionnaire covered 28 threshold concepts that were identified through literature (Cope & Byrne, 2006), (Dunne, 2003), (Kiley, 2009) and (Meyer & Land, 2006). It also gave students the opportunity to identify additional threshold concepts based on the definition of a threshold concept. Students were asked to indicate the level of difficulty (which describes how troublesome the concept is to master) on a Likert scale (Very easy, Easy, Difficult or Very difficult). They were also asked to determine the importance (which refers to how much follow up work is unlocked by mastering the threshold concept) on a Likert scale (Not important at all, Not so important, Fairly important or Very important) for each threshold concept. It was also decided to conduct a three year longitudinal study running from 2014 to 2016. The groups targeted were the 2014 second year commerce students and the second year science students. These two groups will again be accessed in 2015 (when they are third year students) and in 2016 (when they are graduate students). A pilot study was also undertaken by gathering additional data from the 2014 third year students (first year students of 2012) and honours students of 2014 (first year students of 2011). The results discussed will be preliminary, based on data gathered in 2014.

In total 640 students completed the questionnaires and their course enrollment is summarized in Table 1. From Table 1 it can also be seen that 379 (59.22%) of the 640 students will be part of the longitudinal study.

Table 1: Composition of students

Course Level	Course	Number	Percentage
200 (longitudinal study group)	WST221 (science students)	214	33.44
	STK220 (commerce students)	126	19.69
	STK281 (commerce students)	39	6.09
300	WST321 (science students)	158	24.69
	STK320 (commerce students)	64	10.00
Honours	Honours	39	6.09
Total		640	100

The 28 threshold concepts identified through literature are all verified as being threshold concepts for first year students at the university with varying importance and difficulty. Furthermore an additional 14 threshold concepts also with varying importance and difficulty are identified by the students. The threshold concepts can all be analysed in terms of both importance and difficulty in order to gain more insight on how they are experienced by students. Apart from generating frequencies and percentages or calculating descriptive statistics various graphical displays are used to describe the data. Given below are a few such graphical displays. In Figures 1 and 2 the *Centre of Gravity of Scale* is calculated along with 95% error bars for the importance and difficulty of the threshold concepts: *Spread and Variation* as well as *Nonparametric Methods*. In the case of the *Spread and Variation* it can be seen that, although it is regarded as quite important, the level of difficulty is low across all the different modules while in the case of *Nonparametric Methods* the magnitude of the difference between importance and difficulty is considerably less across all the modules, and in the cases of WST221 (science students) and STK220 (commerce students) the level of difficulty is more than the importance.

Figure 1

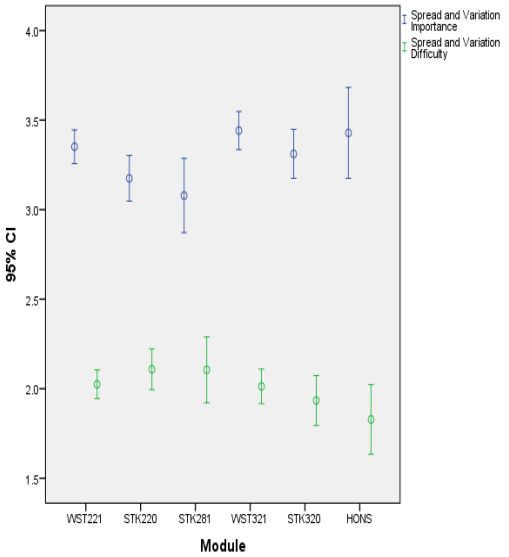
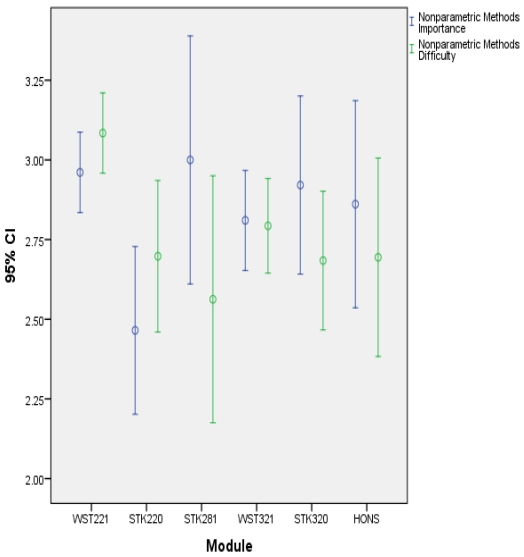


Figure 2



From Figure 3 below it can be seen that *Hypothesis Testing* is regarded as an important threshold concept while Figure 4 indicates that the students have differing opinions on it`s difficulty.

Figure 3

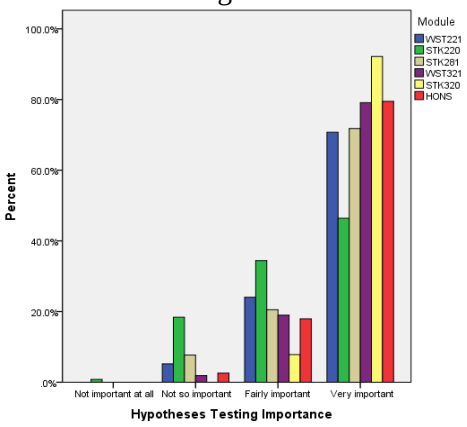
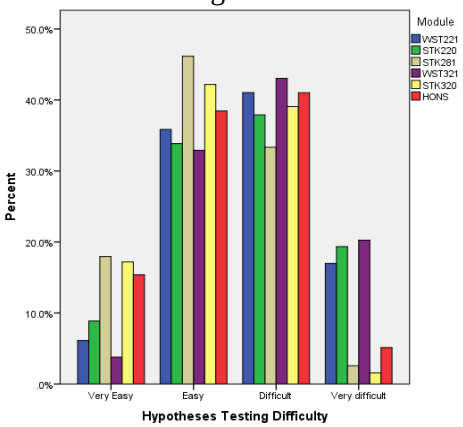


Figure 4



From Figure 5 below it can be seen that the *p-value* is regarded as an important threshold concept while Figure 6 indicates that the students experienced it as relatively easy.

Figure 5

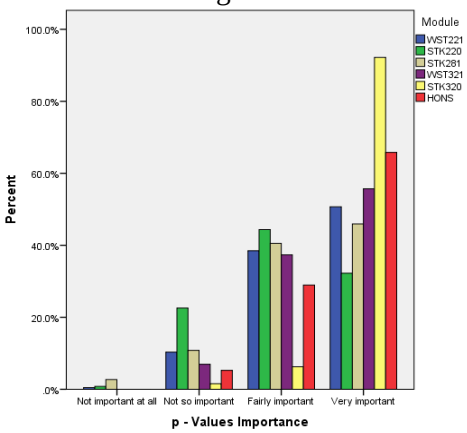
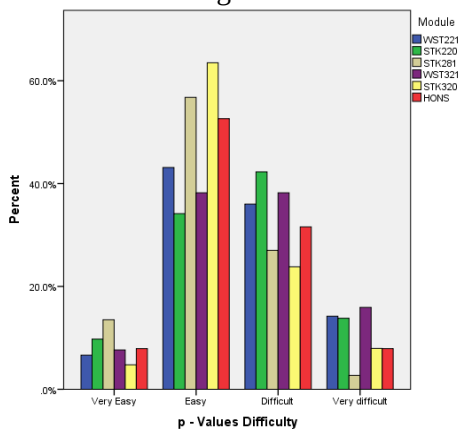


Figure 6



3. Evaluating the threshold concepts

In determining the degree of importance for the observed threshold concepts as experienced by these 640 Statistics students (i.e. the total group) they were asked to list the most important and the least important threshold concepts and overall the most important threshold concepts are *Hypothesis Testing* (19.1%) and *Probability Distributions* (19.1%) as indicated by Figure 7 while the least important threshold concepts are *Bayes' Theorem* (11.6%), *Index Numbers* (11.0%) and *Randomnes*s (10.8%) as indicated by Figure 8.

Figure 7

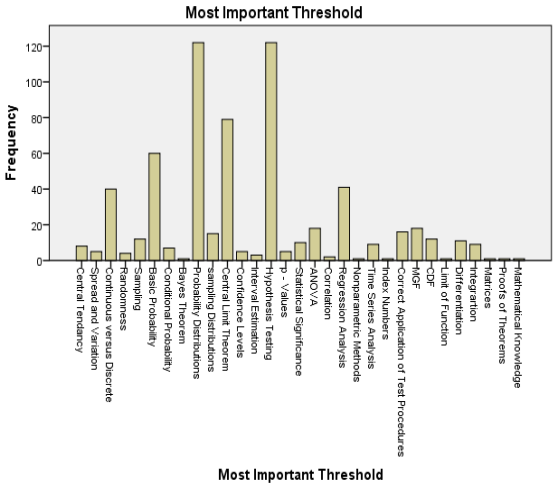
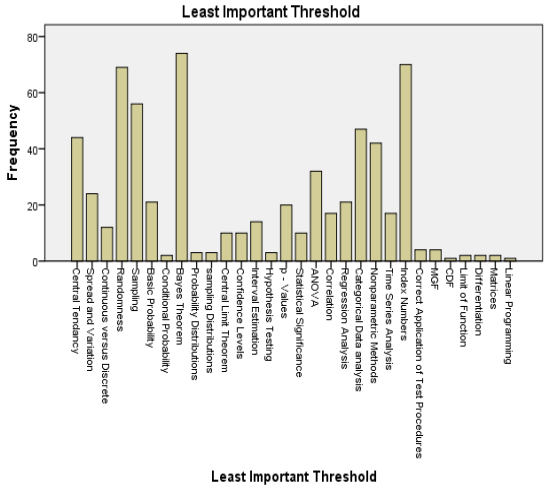


Figure 8



In determining the level of difficulty for the observed threshold concepts as experienced by these 640 Statistics students they were asked to list the most difficult and the easiest threshold concepts and overall the most difficult threshold concepts are *Hypothesis Testing* (14.4%), *ANOVA* (13.6%) and *Regression Analysis* (11.0%) as indicated by Figure 9 while the easiest threshold concept is *Basic Probability* (27.3%) as indicated by Figure 10.

Figure 9

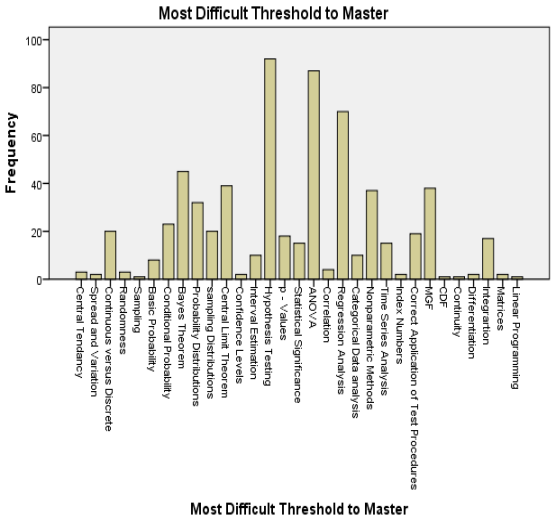
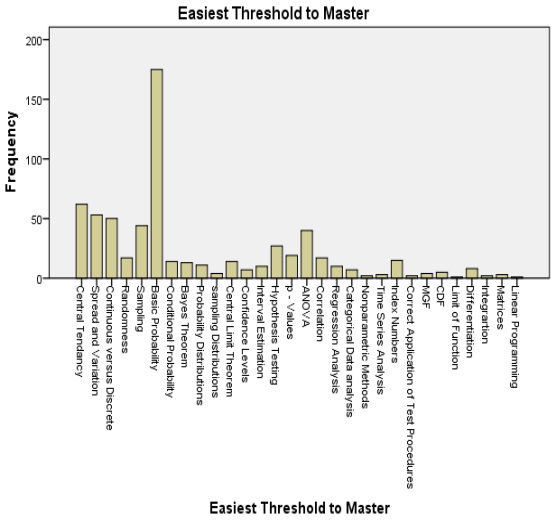


Figure 10



Similar analyses were performed within each of the six observed courses but due to space constraints the results are not discussed in this document.

4. Conclusions

In studying the topic of threshold concepts it became clear that there are differing views on what the most suitable definition for a threshold concept should be. In the study it also became clear that commerce students studying Statistics and science students studying Mathematical Statistics did not always share a common interest in the same threshold concepts and they also rated them differently in terms of importance and difficulty (sometimes marginal and sometimes significantly). In analyzing the data it became evident that students regard *Hypothesis Testing* as the most important and the most difficult threshold concept. This should alert educators at the university who teach this concept on the 100 level that enough time must be allocated in their study program to sufficiently deal with this concept since students regard it as a troublesome concept to master which in turn, when mastered, unlocks a fast amount of follow up work. Failure to master this concept could also restrict the progression within the rest of the course since in all Statistics courses topics/concepts build onto one another. Further analysis will be conducted in future to optimally explore the information contained in this dataset.

References

Cope, C.J., & Byrne, G. (2006). Improving teaching and learning about threshold concepts : The example of confidence intervals, *Threshold Concepts Symposium,* University of Strathclyde, Glasgow, August 30th – September 1st.

Dunne, T., Low, T., & Ardington, C. (2003). Exploring threshold concepts in basic statistics, using the internet, *In Proceedings of The IASE Satelite Conference on Statistics Education and the Internet*, Berlin.

Kiley, M. (2009). Identifying threshold concepts and proposing strategies to support doctoral candidates, *Innovations in Education and Teaching International*, 46(3), 293-304.

Meyer, J.H.F., & Land, R. (2003). Threshold concepts and troublesome knowledge: linkages to ways of thinking and practising within the disciplines, *Improving student learning – Ten Years On*, C. Rust(Ed), Oxford: OCSLD, 412 – 424.

Meyer, J.H.F., & Land, R. (2006). *Overcoming Barriers to Student Understanding: Threshold Concepts and Troublesome Knowledge,* London and New York: Routledge.

Perkins, D. (1999). The many faces of constructivism, *Educational Leadership*, 57(3).