Teaching Analytics and Teacher Dashboards to visualise SET data: Implication to Theory and Practice

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

by

Ifeanyi Glory Ndukwe

UNIVERSITY of OTAGO

SAPERE AUDE

Te Whare Wānanga o Otāgo

NEW ZEALAND

August, 2020

ACKNOWLEDGEMENT

First and foremost, I would like to express my profound gratitude to God almighty for his unfailing love throughout my research journey and for granting me the strength, knowledge and ability to undertake this research to completion.

With immense pleasure and a deep sense of gratitude, I wish to express my sincere thanks to my primary supervisor Associate Professor Ben Daniel, Higher Education Development Centre, University of Otago. Without his motivation, continuous encouragement and contributions, this research would not have been successfully completed. I would also like to express my sincere appreciation to my co-supervisor Dr Russell Butson for the numerous discussions that have shaped this work. I have been so fortunate to have these supervisors with a wealth of knowledgeable who ensured I was always on the right track.

I am grateful to the Head of Department, Higher Education Centre, initially headed by Processor Tony Harland, and currently headed by Associate Professor Ben Daniel, for motivating me to carry out research at the University of Otago and for providing me with infrastructural facilities and many other resources needed for my research. I would also like to appreciate Prof. Vijay Mallan for his support, and the postgraduate coordinators, Higher Education Centre, Dr. Vivienne Anderson and Professor Joyce Koh, for always encouraging and providing additional support throughout the journey.

I would like to express my thanks to our administration team: Jacqueline Fraser, Candi Young, Trilby Ashworth and Joanne Smith, for their support. I still remember all the happy morning teas and celebrations we all shared together.

I would like to thank my fellow postgraduates at the Higher Education Development Centre for the friendship, support and encouragement we provided each other. Special thanks to my office mates Larian, Nick, and Farah, and the gang (Charmaine and Agnes).

I express my sincere thanks to my friends Olusinas, Magbagbeola, Logans, Don for thier kind words of support, encouragement and prayers. I would like to acknowledge the support rendered by my colleagues in several ways throughout my research work.

I wish to extend my profound sense of gratitude to my family: Mr Chukwuemeka Godwin Ndukwe of blessed memory, my loving mother Mrs Ebere Grace Ndukwe, Dr. Henry and family, Ogechi Victoria Chiwuzoh and family, Uche, Favour and Faith, and my wonderful in-laws, Mr Anthoney and Mrs Justina Abraham, for all the sacrifices they made during my research and also providing me with moral support and encouragement whenever required.

Last, but not the least, I would like to thank my lovely wife Harriet, my daughter Mmachi Pearl and son Chiemezuo Othniel, for their constant love, encouragement and moral support along with patience and understanding. Harriet, you have been so sacrificial to supporting me through this entire process.

Ifeanyi Glory Ndukwe

ABSTRACT

Teaching Analytics (TA) is an emergent theoretical approach that combines teaching expertise, visual analytics, and design-based research to support teachers' diagnostic pedagogical ability to use data as evidence to improve teaching quality. The thesis is focused on designing dashboards to help teachers visualise Student Evaluation of Teaching (SET) data as a form of TA for improving the quality of teaching. The research examined the role of TA by deploying customisable dashboards to support teachers in using data to design and facilitate learning. The researcher carried out an integrated literature review to explore the notion of TA and SET data. Moreover, a Data Science Life Cycle model was proposed to guide teachers and researchers using SET data to improve learning and teaching quality. The research comprised several phases. In phase I, a simulated data technique was used to generate SET scores that informed the development of a preliminary teacher dashboard. Phase II surveyed teachers' use of SET data. The survey results indicated that more than half of the participants used SET for improving teaching practice. The research also showed that participants valued the free-text qualitative comments in SET data. Hence, phase III collected real free-text qualitative comments in SET data on students' perceptions of a previously tutored course. The survey results further indicated that although teachers were unaware of a dashboard's value in presenting data, they wanted to visualise SET data using dashboards. Phase IV redesigned the preliminary dashboards to present the real SET data and the simulated SET scores. Finally, phase V carried out usability testing to evaluate teachers' perceptions of usability and usefulness of the teacher's dashboards. Overall, the result of the usability study indicated the perceived value of the teacher's dashboards.

TABLE OF CONTENTS

Appendices

LIST OF FIGURES

LIST OF TABLES

LIST OF TERMS AND ABBREVIATIONS

CHAPTER 1

Introduction

1.1 Overview

More than 16,000 universities worldwide teach many students in different fields and topics per semester (IAU, 2006). A significant component of the teaching process is to request some form of student feedback on teaching and to constructively utilise the feedback to improve teaching practice. It is part of the teacher's responsibilities to manage and resolve student concerns and expectations, some of which could be addressed with feedback from Student Evaluation of Teaching (SET) (Cheng and Tam, 1997). It is also possible to use feedback from SET to influence teaching methodologies and identify problem areas.

Collecting feedback from SET data can be achieved in different ways (Elliott and Shin, 2002). It is usually composed of closed-ended questions with Likert-scale responses, accompanied by additional free-text comment areas. For example, at the University of Otago, staff can request feedback from students using the inForm system (Moskal and Cramond, 2012). The inForm survey consists of questions with Likert-scores ranging from 0 to 5 for students to rate their overall satisfaction with the course, and a more thorough free comment text response in which they can raise any questions or constructive feedback they may have. Thus, student feedback contains a combination of numerical ratings and text data.

Given the widespread availability of student data generated in learning management systems, the teaching processes can be optimised by harvesting, processing and generating insights from various analytics forms (Casey and Azcona, 2017; Cox and Ellsworth, 1997; Dietz-Uhler and Hurn, 2013). The concept of analytics involves the science of analysing raw data to make conclusions about that information (Davenport et al., 2006).

Data-driven decision-making has become the norm in the educational sector. Several institutions are leveraging various data and analytics to ensure that all the decisions that affect teaching and learning are data-driven (Mandinach, 2012). Learning Analytics (LA) is primarily concerned with the analysis of student learning and the context in which learning occurs (Siemens, 2012). It is mostly focused on tracking students and their learning (Khedher et al., 2019). However, there is limited understanding and support for all the pedagogical decision-making happening within the educational sector. Introducing the notion of Teaching Analytics (TA) helps to supplement and compensate for the various kinds of analytics associated with the teacher and teaching practice. TA can provide useful ways to perform analysis and generate insights on teaching data to assess what works for the students and courses taught (Van Harmelen and Workman, 2012).

Technology-enhanced Analytics (TEA) suggested tools that can help academics access and interpret educational data (Daniel and Butson, 2013; Drachsler and Greller, 2012; Gunn et al., 2017). McKenney and Mor (2015) also anticipated that teachers would prefer efficient tools for collecting data and explicit guidance on using the findings to inform teaching and learning. Hence, research is required to explore how teachers can engage with data generated by the teacher and the teaching environment to support reflection on teaching to improve teaching quality.

Visualisation and visual analytics are a core part of TA (Greer and Mark, 2016), and dashboards are visualisation tools that render visual displays containing a collection of vital information all on one computer screen to give users a unified view of the most critical data for insight generation and reflection (Bartlett and Tkacz, 2017). Dashboards consolidate both historical and real-time information in a simple, easy-to-understand, and dynamic format and may present information from different sources in such a way that will be informative for users (Kim et al., 2016). It can also provide teachers with an environment that aids teacher reflection on teaching and learning to improve teaching practice (Vanhoof and Schildkamp, 2014).

This research is important because it explores teachers' perceptions of SET data and the visualisation of SET to improve teaching quality. There are countless debates over SET data's validity (Benton and Cashin, 2014; MacNell et al., 2015). These debates have highlighted some shortcomings of SET data in light of instruction quality (Boring, 2015; Braga et al., 2014). However, the most common method for a teacher to get feedback from students is via SET (Benton and Cashin, 2014; MacNell et al., 2015).

Feedback from SET data can highlight different issues that students may have with the lecture. On the one hand, the teacher asks students questions in a survey about their teaching, content delivery, and learning (Casey and Azcona, 2017). On the other hand, students answers may also allow them to express their opinions, and at the same time inform the teacher of their understanding so that the teacher can revise instructional method where students are experiencing difficulties (Wyeld et al., 2021). Sadler (2010) argues that SET is important to inform teaching practice because it provides a means for the students to give teachers informative feedback and for teachers to provide appropriate interventions and improve teaching quality.

In a large class with hundreds of students, verbal feedback is not a viable means to provide teachers with feedback (Anderson et al., 2003). Several factors, such as a student's personality, or class seating arrangement can impede students' opportunities to provide proper feedback (Hessler et al., 2018). In turn, SET provides a viable means to present teachers with good feedback concerning teaching quality (Altrabsheh et al., 2014). However, reading through all the qualitative aspects or free-texts in SET data poses a potential challenge, especially in large classes where students may provide many comments that could often increase a teacher's workload and become overwhelming for a single teacher to process (Brown, 2020; Poulos and Mahony, 2008).

Despite the usefulness of institutional feedback systems in collecting SET data, they cannot be used to their full advantage without analysing SET data. Besides, other studies have raised concerns that using extra analytical tools could impose an additional burden on teachers, as they may need additional training to use them effectively (Saifee and Jay, 2013). This research addresses this problem by proposing a data science approach to automatically analyse SET data and present the teachers' visualisation dashboards. The dashboards will visualise the SET data in a meaningful way, leaving teachers with the most important information. The teachers would use this information to change teaching style, assessments and lecture materials, and improve teaching quality.

While some institutional feedback systems may support static aggregated SET scores printed in Portable Document Formats (PDFs) or sent as emails to the teachers, they cannot summarise the qualitative comments in SET (Parkin et al., 2012; Graham, 2020). Neither can they provide dynamic reports for the teachers to access and engage with to respond to the most important concerns students share in common (Sarikaya et al., 2018; Cunningham-Nelson et al., 2020). This research presents dynamic and customisable dashboards that allow teachers to drill down and focus on particular aspects of the data and monitor SET score progress over time.

Additionally, sentiment analysis research has expanded dramatically in the last decade, primarily due to the accessibility of rich text resources such as blogs and social networking sites (Agarwal et al., 2011; Saif et al., 2012; Wang et al., 2012). However, less has been reported about students' emotions and opinions in quantitative SET (Russell, 2003). Hence, managing information about students' sentiments in a teaching context could make it possible for teachers to understand students' potential needs better so as to address them (Altrabsheh et al., 2014). For example, student emotion detection using sentiment analysis is important, as previous research highlighted that positive emotions increase teacher motivation and student engagement in classrooms (Munezero et al., 2013).

Collecting and analysing SET data has many benefits, such as allowing the teacher to have an overall summary of students' opinions and promptly address their concerns (Calders and Pechenizkiy, 2012; Poulos and Mahony, 2008). The feedback in SET data can be observed over time, and trends and patterns may be identified, allowing the teachers to make appropriate changes. This research is important in the TA field, as the literature review did not reveal any research exploring the presentation of SET data using dashboards (Altrabsheh et al., 2014; Cunningham-Nelson et al., 2019; Gottipati et al., 2018a; Nitin et al., 2015; Rashid et al., 2013). This research presents a combination of number ratings and free-text comments in SET data using dashboards. This study has also applied various techniques such as data simulation, natural language processing (NLP) and machine learning (ML) to systematically process and present data using a data science life cycle to improve teaching quality.

Design of dashboards without applying interface design and collecting feedback from potential users to improve design quality may not produce a valuable outcome (Paryudi and Fenz, 2014; Shneiderman, 2004). There is little research with applied interface design and usability testing in designing teacher dashboards to improve user experience (Abel and Evans, 2013; Gibson and Martinez-Maldonado, 2017; van Leeuwen and Rummel, 2017; Verbert et al., 2014). This research has applied Shneiderman's interface design and usability testing to improve the quality of the teacher dashboards used to present SET data. Hence, system integration and evaluation are important to TEA and encourage teachers and researchers to enhance teaching and learning.

1.2 Purpose of the Study

This study explored the concept of teaching analytics and the extent to which it can improve teaching quality. The overarching question that guided the research was:

1. How can teachers use teaching analytics to optimise the quality of teaching?

    (a) How do academics perceive the value of SET data?

    (b) How can the visualisation of SET data improve teaching quality?

1.3 Scope

The scope of this research is limited to TA that can be derived from student evaluation. Hence, only SET data was used, and the other datasets were rather discussed but have not been implemented. However, in a future study, this researcher hopes to explore other forms of existing data such as physiological data and information from learning management activities to respond in real-time or almost real-time. Additionally, the dashboards used in this study are proof of concept and have not been implemented in real-life scenarios.

1.4 Thesis Structure

This thesis is structured in seven chapters. Chapter 1 provides a basic introduction to the entire study. In this chapter, the researcher has presented an overview and explained the purpose of the study. A literature review relating to the notion of Teaching Analytics and Student Evaluation of Teaching (SET) data is presented in Chapter 2. The outcome is a Data Science Life Cycle model to guide teachers and researchers in processing SET data to improve learning and teaching quality. Chapter 3 describes the research methodology and outlines the various strategies and approaches used in the research. Chapters 4, 5 and 6 present the analysis and results of the two studies in this research. Chapter 4 presents the analysis and results of an observational study on the views of university teaching staff on the value of teaching evaluation in general and that of Student Evaluation of Teaching (SET) data. The chapter also presents views on the use of teacher dashboards. Chapter 5 presents the design principles applied to the design of a Teacher's Evaluation Dashboard, aiming to improve teaching professional development. Chapter 6 presents the usability studies carried out on the Teachers Evaluation Dashboard (TED). Chapter 7 presents the discussion, conclusion and future research; the theoretical and practical implications of this study are discussed in this chapter.

CHAPTER 2

Integrated Literature Review on Teaching Analytics

2.1 Overview

This chapter presents a systematic review of teaching analytics and research on student evaluation of teaching. Discussions related to Teaching Analytics (TA) in the literature will be used to provide a generic conception of TA, and SET data's meaning and value as a form of TA. The review informs visualisation tools such as dashboards to improve teacher engagement and enhance the teaching profession. This review has led to the establishment of a Data Science framework that describes the various aspects of processing SET to enable the generation of insights from SET data and guide teachers and researchers to engage with SET data to improve teaching quality and learning outcomes.

2.2 Introduction

Learning Analytics (LA) has become a phenomenon in the educational sector due to student's massive digital footprints in this digital age (Daniel, 2017). Research into LA originated from student retention requirements, and has brought about an invention using the Course Signal System (Arnold and Pistilli, 2012). Siemens and Long (2011) defined LA as the "measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs". In other words, LA relies on the analytics of students data to gain meaningful insight capable of improving learning.

On the other hand, TA focuses on improving teaching outcomes (Bueckle and Brner, 2017; Ong, 2015). TA can be perceived as a theoretical approach that combines teaching expertise, visualisation and design-based research to support teachers' diagnostic pedagogical ability to use data as evidence to improve teaching quality (Vatrapu et al., 2011). TA is now gaining prominence because it offers teachers enormous opportunities (Wise and Jung, 2019). It also identifies optimal ways in which teaching performance can be enhanced. TA's primary outcome is to guide education researchers to develop better strategies to support the development of teachers' data literacy skills and knowledge (Gummer and Mandinach, 2015). However, for teachers to embrace data-driven approaches to learning design, there is a need to implement bottom-up approaches that include teachers as the main stakeholders of a data literacy project, rather than as end-users of data.

Pedagogical theory of student feedback from SET data describes the need for interpreting students' opinions for overall teaching evaluation and improvements (Dietz-Uhler and Hurn, 2013; Mohanan, 2005). Donovan et al. (2010) found that online student feedback increased the number of words per comment and provided more formative feedback than traditional written feedback. Also, online feedback generated longer and twice as many (54 per cent or more) comments as conventionally written comments. This results highlight the value of the collection, analysis and visualisation of SET data as a form of TA for teachers to reflect on student opinions about teaching practices and different aspects of the course that can be enhanced to improve learning outcomes. This research explores the current discussions in the literature relating to TA to contextualise the notion of SET as a form of TA.

This study reviewed articles published from 2011 to 2019. A total of 128 publications were initially identified and compiled from the Scopus database. After analysing the search results, 63 papers were selected for review (see Appendix A). This review examined research relating to TA and the analytics of SET as a form of TA.

2.3 Methods and Procedures

The review first focused on describing TA, how it is used, and how SET data can be processed as a form of TA. The current review started with searching through the Scopus database using the SciVal visualisation and analytical tool. The rationale for choosing the Scopus database is that it contains the largest abstract and citation database of peer-reviewed research literature, with diverse titles from publishers worldwide. Hence, it is possible to search for and find a meaningful balance of the published content in the areas mentioned above. Also, the review included peer-reviewed journals and conference proceedings. The researcher excluded other documents and source types, such as editorials and trade publications; these sources might lack research on TA and SET data. Also, this review excluded articles published in languages other than English.

This chapter used several keywords and combinations to search for terms related to TA and analysis of SET data. For instance: "Teaching Analytics" AND "Learning Analytics" OR "Student Feedback" OR "Student Opinion" OR "Text Mining" OR "Opinion Mining" OR "Student Opinion Mining" OR "Sentiment" OR "Visualisation" OR "Dashboards" OR "Teaching Evaluation" OR "Student Evaluation on Teaching" OR "Student Ratings". The "AND" search key was used because the term LA is often used interchangeably to refer to TA. In comparison, the "OR" search key was used to search for all the other terms used to analyse and visualise SET data.

The key search terms “Student Feedback”, “Student Opinion”, “Teaching Evaluation”, “Student Evaluation on Teaching”, “Student Ratings” are related to the first research question and were used to search for how academics perceive the value of SET data concerning improving their teaching practice. The other key search terms, such as “Text Mining”, “Opinion Mining”, “Student Opinion Mining”, “Sentiment”, “Visualisation” and “Dashboards” are related to the second research question, which will develop an understanding of the literature about the processing and visualisation of SET data to improve teaching quality.

The researcher searched for articles published between 2011 and 2019. The start year was chosen because TA’s notion of analytics for improving teaching practice began to emerge in 2011 (Dix and Leavesley, 2015; Sampson, 2017; Vatrapu et al., 2011). For instance, Vatrapu et al. (2011) introduced the concept of TA as a new theoretical approach that triangulates visual analytics, teaching expertise and design-based research to support teachers’ diagnostic pedagogical decision-making in classrooms. Subsequently, researchers proposed a conceptual framework for TA to trigger actionable pedagogy for the teacher’s professional vision (Vatrapu, 2012), and then implemented a real-time, real-place TA dashboard (Vatrapu et al., 2013c). Additionally, Dix and Leavesley (2015) conceptualised TA as LA for the teachers. Sampson (2017) went further to refer to TA as half of the inquiry cycle, while LA is the other half, and a combination of LA and TA completes the full inquiry life cycle.

The initial stage of the literature search yielded 128 papers. After the subsequent screening of previous works and removing duplicates and titles that did not relate to the research area, 101 articles remained. As such, a total of 75 studies were pursued to a full-text review. The review ensured that the articles identified for review were both empirical and conceptual papers. Each article’s relevance was affirmed by requiring that chosen papers contained various vital phrases all through the paper, title, abstract, and keywords, and afterwards, the entire essay. In essence, the articles reviewed gave specific consideration to those section(s) that expressly related to the field of TA or SET, to extract essential points of view on the processing of SET data or definitions, data sources, tools and technologies associated with TA. This review also disregarded papers that did not relate to analytics in the teachers’ context or SET data processing for teachers’ consumption. Finally, 63 articles remained for this review (see Figure 2.1).

2.4 Teaching Analytics, Tools and Value

Several studies have demonstrated that TA is an important area of inquiry (Flanders, 1970; Gorham, 1988; Pennings et al., 2014; Schempp et al., 2004), enabling researchers to systematically explore analytics associated with the teaching process. However, there

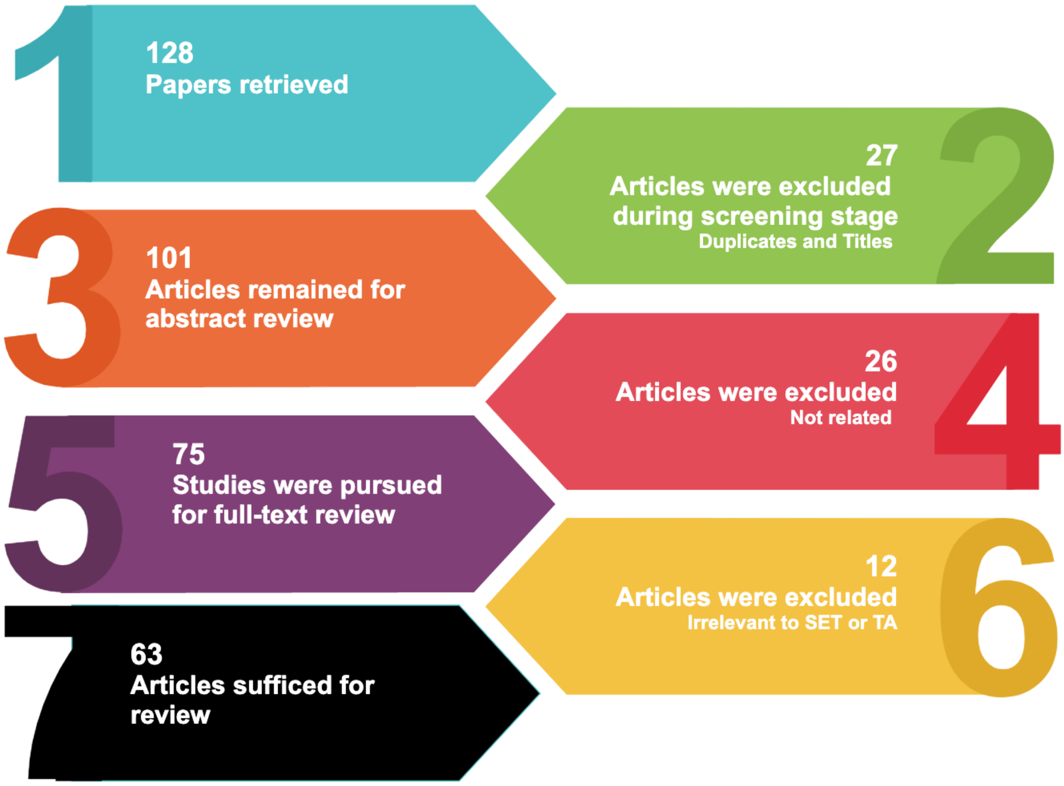


Fig. 2.1 SLR Workflow

is no consensus on what constitutes TA. Several studies suggest that TA is an approach used to analyse teaching activities (Barmaki and Hughes, 2015; Gauthier, 2013; KU et al., 2018; Saar et al., 2017), including how teachers deliver lectures to students, the student usage pattern for tools, or dialogue. In contrast, various other studies recognised TA as the ability to apply analytical methods to improve teacher awareness of student activities for appropriate intervention (Ginon et al., 2016; Michos and Hernández Leo, 2016; Pantazos et al., 2013; Taniguchi et al., 2017; Vatrapu et al., 2013b). A handful of others indicate TA as analytics that combines both teacher and student activities (Chounta et al., 2016; Pantazos and Vatrapu, 2016; Prieto et al., 2016; Suehiro et al., 2017). Hence, it is particularly problematic and challenging to systematically study analytics for the teachers to improve their teaching practice, since there is no shared understanding of what constitutes analytics and how best to approach TA.

Various studies have used custom software and online applications such as Learning Management System (LMS) and Massive Open Online Courses (MOOCs) to collect online data on classroom activities (Goggins et al., 2016; KU et al., 2018; Libbrecht et al., 2013; Mller et al., 2016; Shen et al., 2018; Suehiro et al., 2017; Vatrapu et al., 2013c; Xu and Recker, 2012). Others have used modern devices including eye-trackers, portable electroencephalograms (EEGs), gyroscopes, accelerometers and smartphones (Prieto et al., 2016, 2018; Saar et al., 2017, 2018; Vatrapu et al., 2013a), and conventional instruments such as video and voice recorders (Barmaki and Hughes, 2015; Gauthier, 2013; Thomas, 2018) to record classroom activities. However, some authors have pointed out several issues with modern devices such as expensive equipment, high human resource and ethical concerns (KU et al., 2018; Prieto et al., 2017, 2016; Suehiro et al., 2017).

Visualisation techniques can be applied to educational data to empower the teachers, increase teacher engagement, enhance teacher reflection, improve teaching professional development and promote data-driven decision-making. Visualisation techniques are a vital part of TA (Prieto et al., 2016, 2017, 2018; Saar et al., 2017; Thomas, 2018) and can be applied to SET data as an additional and cheaper alternative to generate insight to improve the teaching profession.

2.5 Visualisation of Student Evaluation of Teaching

The literature has extensively reported various data sources used for TA. This study draws attention to the visualisation of SET data as an additional form of TA. The visualisation of SET could support teacher reflection on teaching practice and add value to TA (Cunningham-Nelson et al., 2019). Nevertheless, there are countless debates over SET data's validity (Benton and Cashin, 2014; MacNell et al., 2015). These debates have highlighted some shortcomings of SET data in light of instruction quality (Boring, 2015; Braga et al., 2014). For instance, Hessler et al. (2018) found that student feedback was more favourable when teachers offered students chocolates in class immediately before student surveys were administered. Some studies have also shown that teachers from non-English-speaking backgrounds and female teachers are more likely to receive unfavourable reviews (Abrami et al., 2007; Marsh, 2007). SETs are also questioned because they assess student satisfaction instead of student learning or teaching quality (Dean and Gibbs, 2015). For some individual teachers, their perception of SET could be sufficient to undermine teachers' professional development (TPD), especially if they think they are the audit subjects (Edstrm, 2008).

However, today, SET is an integral part of the universities' evaluation process (Ducheva et al., 2013). Research has also shown that there is substantial room for using SET data for improving teaching practice, including improving the quality of instruction, learning outcomes, and teaching and learning experience (Linse, 2017; Rathke and Harmon, 2011; Stupans et al., 2016; Subramanya, 2014). SET is thought to be an effective mechanism for collecting student opinions and making their voices heard on teachers' relative performances and courses taught (Spooren et al., 2017). A recent study on SET showed that the value of measuring teaching quality using SET scores consistently emerged as the predominant theme, followed by the construction and validation of SET instruments, then the use of SET data to assess and enhance teaching (Spooren et al., 2017).

According to Brockx et al. (2012), SET data's quantitative feedback often provides more specific feedback and useful information, which suggests they could be used to improve teaching quality. Also, several studies have demonstrated that the qualitative SET data provided useful information that can be analysed and visualised to provide actionable insight for the teachers to improve teaching quality (Altrabsheh et al., 2014; Cunningham-Nelson et al., 2019; Gottipati et al., 2018a; Nitin et al., 2015; Rashid et al., 2013).

Additionally, several studies have applied statistical measures and descriptive analytics on SET data's quantitative aspect (Hajizadeh and Ahmadzadeh, 2014; Kitto et al., 2019). However, attention is shifting to SET data's qualitative aspect (Altrabsheh et al., 2014; Gottipati et al., 2018a). In response to the widespread concern about the limited use of qualitative SET data, Kitto et al. (2019) recommended the need to go beyond standard statistical measures of quantitative data when analysing and examining SET, but rather to also take into account the context of SET. Research has been done on applying automatic methods such as data mining (Rashid et al., 2013), NLP (Cunningham-Nelson et al., 2019; Gottipati et al., 2018a; Nitin et al., 2015), and ML (Altrabsheh et al., 2014; Luo et al., 2018b) techniques to analyse and visualise quantitative SET data and to present prospects for more significant decision-making. An example is a study done by Gottipati et al. (2018a), who employed Latent Dirichlet Allocation (LDA) models and sentiment mining techniques to discover topics from student comments, and also classify them into positive or negative sentiments. However, a critical caution about sentiment analysis is the need to be accurate enough to be considered reliable and useful (Ravi and Ravi, 2015). The quality assurance management processes have also recommended accuracy level is an important caveat in using sentiment analysis on SET data in institutional surveys (Cunningham-Nelson et al., 2020).

Few studies have also provided visual representations of qualitative and quantitative SET data (Shah and Pabel, 2019) to improve data-driven decision-making. However, the visualisations are usually in static form and not interactive. Other static forms of visualisations include the visual representation from Cunningham-Nelson et al. (2019) of qualitative SET data that contains both sentiment and themes. Nitin et al. (2015) used opinion summaries to present a concise and digestible summary of student opinions in a Microsoft Word static visualisation. Visualisations of SET data can be enhanced by making them more interactive and dynamic using dashboards.

2.6 Teacher Dashboards

Teacher dashboards can be useful tools that can be applied to represent and convey SET data results using visualisation techniques. Käser et al. (2017) argued that while there has been an advancement in making useful dashboards that support decision-making for teaching and learning, presentation of effective and efficient information on dashboards remains a challenge. Preceding research on educational dashboards to inform the design and development of dashboards (Greller and Drachsler, 2012; Holstein et al., 2017, 2018; Schwendimann et al., 2016; Verbert et al., 2014) can also be applied to the design of dashboards for visualising SET data.

The importance of aligning the goals of a dashboard design to the types of information presented has been shown by prior study (Greller and Drachsler, 2012). However, some information presented in previously established dashboards are problematic and may provide more information than necessary (Atterer et al., 2006). For instance, a dashboard can display information with possible data privacy issues or personal information (Holstein et al., 2018).

Several studies have shown how teachers use insight generated from dashboards to guide instruction and provide interventions to problems both in class and at the individual student level (Greller and Drachsler, 2012; Holstein et al., 2018). These studies demonstrated that teachers use data from direct teacher observations to decide on the interventions required, develop lesson plans and assess progress.

Various design models have been adopted in designing dashboards for teachers. Greller and Drachsler (2012) designed a dashboard following four dimensions: data source (collection of educational data), data users (stakeholders that will consume the information) and visualisation (visual displays, reports and dashboards) and limitations. Alternative framing was considered by Bakharia et al. (2016), exploring the following dimensions: temporal (the time interval used for results presentation); the capacity to compare results (single user view with multiple graphs); dynamic graphs that alter through the use of filters; intervention and measurement methods; and defining elements to obtain finer-grained details.

Additionally, researchers have applied user-driven design processes that may involve interviews, prototyping, and usability testing to show teacher's significance in designing teacher dashboards (Atterer et al., 2006; Greller and Drachsler, 2012; Tavares et al., 2019). Holstein et al. (2018) modelled a dashboard design as an iterative process comprising continuous interaction with teachers. This method guarantees that the end product has all the critical features required by the teachers. Dashboards are designed for simplicity, flexibility and power. However, simplicity is often neglected, making the dashboard complicated or time-consuming for teachers (Holstein et al., 2018). Dashboards should be easy for teachers to read, manipulate, and interpret.

Schwendimann et al. (2016) identified that the key data sources used in educational dashboards are: logs to monitor user activities, student-generated learning artefacts, results of activities and personal information and background. However, in our case, teacher dashboards will focus on representing qualitative and quantitative SET data.

The Shneiderman's Eight Golden Rules of User Interface Design are as follows

(Schneiderman and Plaisant, 2005):

1. **Strive for consistency.** The user interface must also behave consistently, and there should also be consistencies throughout the design in the layout, fonts, colour, capitalisation, terminology.

2. **Enable users to use shortcuts.** Since users of an application are diverse, the user interface design must be flexible with useful shortcuts to reduce time and prevent significant user errors in the system.

3. **Offer informative feedback.** System feedback is necessary for every action and important to inform and notify the users of what is happening.

4. **Design dialogue to yield closure.** A good interface design must provide users with informative guidance to perform appropriate actions. For instance, a component can provide the users with a descriptive message when they hover the mouse over an element.

5. **Permit easy reversal of actions.** A good design must allow actions to be reversed easily to encourage users to explore the application and to relieve user anxiety about making the wrong action.

6. **Support internal locus of control.** User interface design must give users the feeling they are in charge and have control to maximise the system's usability.

7. **Reduce short-term memory load.** Human short-term memory is limited, and we can only remember five items at a time. Hence, the users must not be asked to memorise information that exceeds their short-term memory capacity (Todd and Marois, 2004). For instance, the users might not remember information from one section of the user interface to apply in another portion.

8. **Offer simple error handling.** Avoiding serious errors is one way to prevent users from making mistakes. Also, systems should provide clear and informative instructions to enable users to recover from possible errors. For example, user input validation to accept numbers and reject alphabetic input prevents the users from wrongly entering alphabetic inputs.

According to Sharp (2003), interactive design involves designing an interactive product to support users' daily activities and interactions, in other words, designing an everyday user experience that enhances people's work experience. One aspect of interactive design is interface design (Kolko, 2010). Teacher dashboards can apply Shneiderman's theory to improve their interface design (Shneiderman, 2004; Shneiderman and Plaisant, 2010). Literature suggests that incorporating interface design theories can be applied to dashboards (Lempinen, 2012; Vozniuk et al., 2013). In contrast, previous studies have revealed that interface design is rarely applied to teacher dashboard design (Ali et al., 2012; Atterer et al., 2006; Greller and Drachsler, 2012; Holstein et al., 2017; van Leeuwen and Rummel, 2017; Verbert et al., 2014). Current teacher dashboards could be improved with these design considerations to increase teachers' engagement with dashboards and enhance teaching quality.

2.7 Gaps in Previous Research

TA is an important area of inquiry that could enable teachers and researchers to explore educational data systematically (Gorham, 1988; Pennings et al., 2014) and apply analytical methods (Ginon et al., 2016; Pantazos et al., 2013; Vatrapu et al., 2013b) such as visualisation techniques to improve their teaching quality. Most of the current teacher evaluation research explores the automatic extraction and analysis of free-text comments in SET data (Koufakou et al., 2016; Shah and Pabel, 2019). Despite the countless debates over the validity of SET data (Benton and Cashin, 2014; MacNell et al., 2015), there is a need to explore teachers' opinions about SET data further and visualise SET data to improve teaching performance.

Notwithstanding, visualisation of free-text comments can help teachers drill-down to investigate into more detail areas of importance. It is not a replacement for reading the full comments and context. It is evident in the findings that they can support teachers to create a story to enhance their course materials and refine teaching practice, as well as helping to provide teachers with a better understanding of the subject areas where they performed well and those that may need more attention. Nevertheless, prior studies have mostly concentrated on processing the free-text comments in SET data. Some of the studies such as those by Palmer and Campbell (2015) and Luo et al. (2018a) provided visualisations of free-text comments in SET data that allow important aspects of comments to be emphasised or summarised. However, both SET scores and free-text comments in SET data need to be visualised in dashboards to present more opportunities for the teacher to further investigate and evaluate the relationship between the numerical ratings and the free-text comments in SET data.

Shah and Pabel (2019) was the only study that visualised numerical SET scores and free-text comments in SET data. However, there was no interactivity in the visualisation. Cunningham-Nelson et al. (2020) highlighted the need to create interactive visualisations to provide the ability to click through to positive or negative comments relating to any particular theme that might encourage further interaction with the comments' consideration of possible teaching improvements. This research provides interactive visualisations of numerical SET scores that allow teachers to monitor trends and patterns over a period. In addition, interactive visualisations of free-text comments in SET data allow teachers to drill down to investigate in more detail and to focus on areas important to a particular teacher to transform the course material and improve teaching quality.

Additionally, for the visualisations themselves to be considered reliable and useful, the algorithms must be accurate. Watkins et al.'s research on Valence Aware Dictionary and Sentiment Reasoner (VADER) in 2020 showed that VADER has far outperformed expectations in accurately performing sentiment analysis on SET data. This research applied VADER to perform sentiment analysis on SET data and other forms of automatic text analytics such as words and phrases, named entity recognition, clustering and text summarisation.

Furthermore, there is also a need for appropriate design considerations to be put in place. Shneiderman's Eight Golden Rules is one interface design theory that can be applied to improve the interface design of visual teacher dashboards (Shneiderman, 2004). Also, there is a need for usability testing to determine the dashboard usability and usefulness or to measure the acceptance rate and collect possible feedback from users for further improvement on the dashboard design.

Finally, it is evident from the cited literature that there is a need to develop the proposed system, because SET data could be tedious work for the teachers to identify and validate the students' concerns. One of the obstacles is that reading and making sense of all the textual responses can be a daunting task. Several studies on SET have proposed models to understand and implement text summarisation, opinion mining and sentiment analysis on SET data (Pyasi et al., 2018; Gottipati et al., 2018a,b; Nitin et al., 2015), to enable teachers to produce statistics and generate reports on SET data. While these models are useful in processing SET data, many of these works only focus on a particular algorithm or analysis. For example, Gottipati et al. (2018a) and Pyasi et al. (2018) proposed models that apply LDA to discover aspects or topics in student comments. There is a need for a more flexible and robust model that integrates and elaborates the algorithms, tools, techniques, and processes required to handle SET data more efficiently.

2.8 Data Science Life Cycle for Processing SET Data

Data science is an interdisciplinary field aiming to turn data into real value (Donoho, 2017). According to Berman et al. (2018), Data Science Life Cycle is a new concept that extends from the "Data Life Cycle", which has a long history in information sciences and many other science domains. Additionally, "Data Science Life Cycle describes the various stages a dataset traverses as it undergoes scientific collection and investigation and is typically used to guide data management decisions and practices." (Stodden, 2020, p. 62).

This research proposed a Data Science Life Cycle to process SET data from the conception of the business understanding, through data acquisition, deployment, modelling and data understanding. This model is robust and integrates several algorithms, tools, techniques, and processes required to efficiently handle SET data and produce interactive visualisation dashboards of SET data to support individual teachers in studying their performance to identify their strengths, potential and weaknesses. This section describes a Data Science Life Cycle framework that outlines the major stages that can be followed to process SET data (see Figure 2.2).

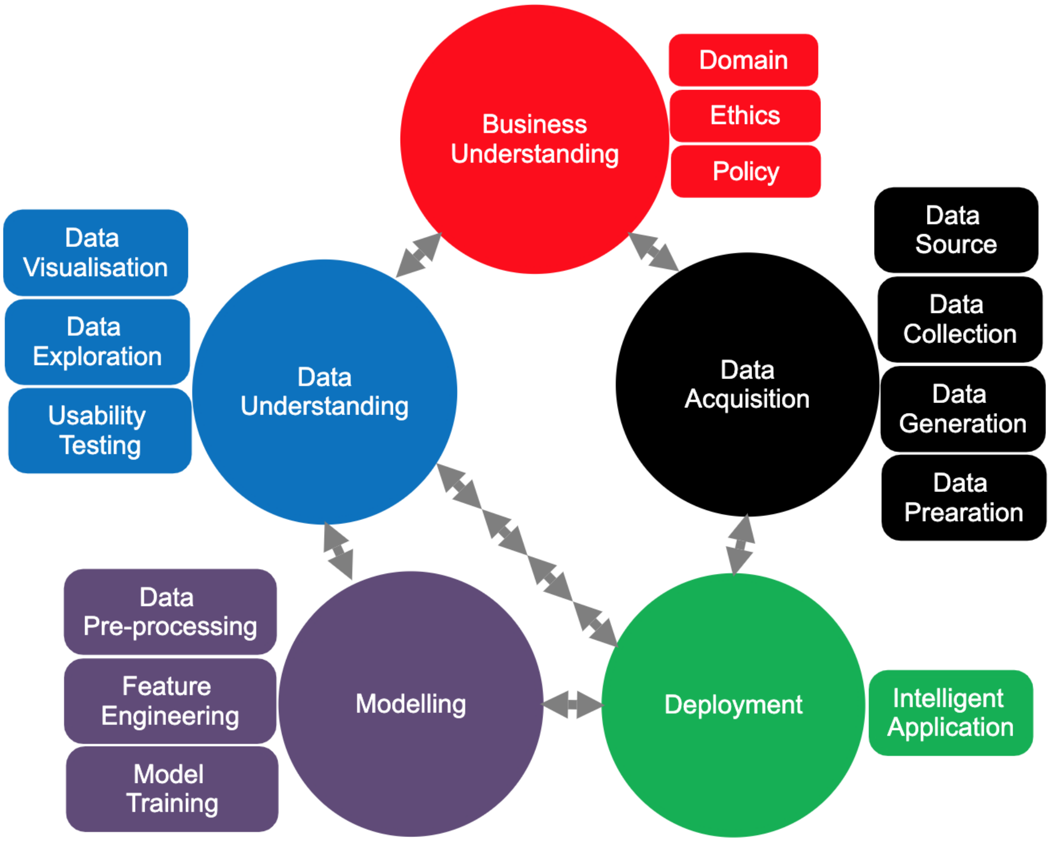


Fig. 2.2 Data Science Life Cycle Framework for processing SET data

2.8.1 Business Understanding

The Business understanding stage of SET involves identifying and specifying the key variables, such as the numbers of questions used in the SET survey; the nature of questions to be analysed (quantitative and qualitative SET data); the period for the SET data to be collected and used for the analysis; and the articulation of clear ethics, regulatory and governance policies around the collecting and the use of SET data (McCormack, 2005). It also involves understanding the domain or literature about SET data.

According to Van Der Aalst (2016), business understanding involves determining the relevant questions to be asked and how answering them can achieve the overall business goal. Concerning the business understanding of SET data, the literature has extensively revealed two key categories of questions researchers seek answers to, and they are: what happened? (Reporting); and why did it happen? (Anomaly detection).

Descriptive analytics is a statistical method that focuses on gathering and summarising raw historical data to be easily interpreted to provide vital context for understanding information and numbers (Olson and Lauhoff, 2019). Descriptive analytics is used to answer "what happened?". Data aggregation and data mining are two techniques used in descriptive analytics to discover historical data (Assuno et al., 2015; ur Rehman et al., 2016). Data is first gathered and sorted by data aggregation to make the datasets more manageable. Then the next step involves searching and mining the data to identify meaningful patterns.

Several researchers have applied descriptive analytics to SET data (Cunningham-Nelson et al., 2019; de Oliveira et al., 2019; Luo et al., 2018b; Nitin et al., 2015; Palmer and Campbell, 2015; Shah and Pabel, 2019). For instance, de Oliveira et al. (2019) used descriptive analytics to generate a word cloud describing the most used positive and negative words in sentences and to generate trends of average scores of the positives, negatives and neutral scores over six years. Tseng et al. (2018) used descriptive analytics to show positive, negative and neutral opinion data.

While descriptive analytics is the initial step of the analytical process, diagnostic analytics takes it a step further to uncover the reason behind specific results (Olson and Lauhoff, 2019). Diagnostic analytics describes the techniques used to ask data: "why did it happen?" It is doing a deep dive into data or focusing on particular aspects of the data to search or discover valuable insights (Banerjee et al., 2013).

To understand the "why" behind what happened, researchers have used several techniques such as data discovery, drilling down, data mining, and correlations to discover hidden insight in SET data (Chathuranga et al., 2018; Gottipati et al., 2018b; Lalata et al., 2019; Luo et al., 2018a; Nitin et al., 2015; Sindhu et al., 2019). For example, Gottipati et al. (2018b) proposed a solution for drilling down into student feedback to extract various aspects or to produce summarised bullet points. Lalata et al. (2019) performed diagnostic analytics to discover whether there was any correlation between the students' numerical ratings and their written comments.

2.8.2 Data Acquisition

This stage aims to produce clean, high-quality and transformed data before deployment into the appropriate analytics environment for modelling. SET data may be collected from a database (Cunningham-Nelson et al., 2019; Luo et al., 2018b; Rashid et al., 2013) or a simple flat file (e.g. Excel spreadsheets, Word documents, PDF documents or text files) (Altrabsheh et al., 2014; Gottipati et al., 2018a), originating from either a single source or multiple sources.

Data can also be generated using methods such as data-driven simulations. Data-driven simulations are wholly parameterised simulations that provide data through a set of data inputs, allowing users to create and run simulation models with generated data (Wang et al., 2011). They are applied in areas where analytics on actual data could be costly, time-consuming, dangerous, or critical (Jung, 2018). It can also be used as a proof of concept before findings can be applied to the real data (Vanbrabant et al., 2019).

According to Price et al. (2019), teachers use simulations for content learning, student motivation, and engagement in science practices. Data-driven simulation can serve as an informative decision support tool for a teacher's quality assessment (Goodall et al., 2019). The assessment of teaching quality is a crucial piece of teaching management, and the quality of teaching is associated with the development of high-quality students (Martinez et al., 2016; Warman, 2015). Benner and McArthur (2019), developed a pedagogical approach to advance data-driven simulations in the education sector. Hence, the assessment of a teacher's teaching quality encourages the teacher to enhance their self-development, promote the quality of teaching, and, lastly, accomplish the goal of improving their teaching quality. Additionally, Zhang et al. (2019) proposed using the simulation technique for simulating a teaching quality index. Babik et al. (2019) also used simulation to separate the structural component of peer assessment from the cognitive effects.

Data sources may be structured (e.g. student number rating or quantitative Likert-scale numbers) or unstructured (e.g. students' free-text or qualitative comments). Unfortunately, researchers are increasingly exploring the unstructured SET data (Altrabsheh et al., 2014; Cunningham-Nelson et al., 2019; Gottipati et al., 2018a; Nitin et al., 2015; Santhanam et al., 2018), which requires data to be prepared. Data preparation is cleaning the data from noise to ensure data quality (Altrabsheh et al., 2013).

The phrase "Extract, Transform and Load (ETL)" is used to describe the process that involves: extracting data from single or multiple data sources, transforming data (data cleansing, manipulation or wrangling) to fit operational needs, and loading transformed data into the target system (Vassiliadis et al., 2002). For example, Gottipati et al. (2018a) recommended using named entity recognition tools to clean sensitive information such as anonymising the names in the qualitative part of the raw SET data before loading it into the database.

2.8.3 Data Deployment

The data deployment stage includes intelligent application systems and web services that deploy the model and the data pipeline to production or production-like environment for real-time or batch-basis user consumption. Deployed models can be exposed to open Application Programming Interfaces (APIs) to be easily consumed by various applications, such as dashboards, custom applications and back-end applications.

Santhanam et al. (2018) highlighted some drawbacks and advantages of using selected applications. Moreover, some researchers have concluded that a particular application to handle all tasks does not exist and each task may need a unique application (Ordenes et al., 2014; Puschmann and Powell, 2018), whereas others have claimed that different applications should be used for a single data source (Nelson et al., 2018). Downer et al. (2019), recommended using intelligent applications that can be loosely described as "big data" to perform analysis on SET data rather than relying on server response systems.

2.8.4 Modelling

Data pre-processing is a normalisation and transformation process carried out on data to prepare for computational operations (Altrabsheh et al., 2013). Some of the pre-processing steps applied to raw SET data include converting upper-case letters into lower-case; removing special characters to increase the accuracy of the results and reduce the errors in the data; removing stop words ("and" or "the") and punctuation from comments; performing lemmatisation to normalise words into their base form (so that words with the same meaning but different suffixes are treated as the same word), stemming (using the porter stemming algorithm) which does not always result in a dictionary word, Part-Of-Speech (POS) (weather as a word is a NOUN, VERB, ADJECTIVE etc), n-grams (convert words to uni-gram, bi-grams, etc.), and Named Entity Recognition (NER) (extracting all important entities in a given sense.) (Altrabsheh et al., 2014; Chathuranga et al., 2018; Cunningham-Nelson et al., 2019; de Oliveira et al., 2019; Gottipati et al., 2018b; Lalata et al., 2019; Nitin et al., 2015; Pyasi et al., 2018; Shah and Pabel, 2019; Tseng et al., 2018). For example, Altrabsheh et al. (2014) highlighted a set of pre-processing steps that they applied to prepare the raw SET data for analysis. Also, Cunningham-Nelson et al. (2019) used POS tagging to remove and keep some key aspects (adjectives, verbs and adverbs) from each statement in student comments.

Feature engineering is a central task in data preparation for ML (Nargesian et al., 2017). Usually, ML models cannot receive raw data as input for model training and testing until transformation functions have been applied to transform the raw data into vector forms, suitable for the models to access (Kanter and Veeramachaneni, 2015). Bag of Words is a simplified model representation for the data used in NLP and Information Retrieval (IR) (Zhang et al., 2010). Bag of Words is a common feature engineering technique applied to convert raw student texts into vectors (de Oliveira et al., 2019; Nguyen et al., 2018). A Bag of Words model was proposed by de Oliveira et al. (2019) that first transforms students' raw comments into tokens and subsequently matches them to numeric vectors.

Term Frequency-Inverse Document Frequency (TF-IDF) is a statistical measure used to evaluate important words in student comments and represent them as a bag-of-words matrix of TF-IDF values (Rajaraman and Ullman, 2011). It uses the Term Frequency (TF) vector to convert student comments into vector space representations (Beel et al., 2016). TF-IDF is another form of feature engineering technique applied on SET data (Altrabsheh et al., 2014; Koufakou et al., 2016; Sindhu et al., 2019; Tseng et al., 2018) to transform them to vectors before applying ML models. For example, Tseng et al. (2018) used TF-IDF to convert the students' comments into weighted values and derive all the important words.

Model training is when an ML algorithm is fed with sufficient training data to learn from (Moore and Lewis, 2010). The process for model training includes the following stages: split the input data randomly for modelling into training and test data sets; build the models by using the training data set; evaluate the training and the test data set; make predictions or recommendations. Sometimes, the feature engineering and model training can be combined and embedded (operating in fully autonomous mode and providing outputs without substantial human interventions) through intelligent systems (Lemos et al., 2013; Lughofer, 2018; Sayed-Mouchaweh and Lughofer, 2012).

Some researchers have used ML to train real SET data such as Naive Bayes (Altrabsheh et al., 2014; de Oliveira et al., 2019; Gottipati et al., 2018a; Koufakou et al., 2016; Lalata et al., 2019), Complement Naive Bayes (CNB) (Altrabsheh et al., 2014), Support Vector Machine (SVM) (Altrabsheh et al., 2014; de Oliveira et al., 2019; Lalata et al., 2019; Nitin et al., 2015), k-Nearest Neighbor (k-NN) (Koufakou et al., 2016), Maximum Entropy (Altrabsheh et al., 2014), Logistic Regression (de Oliveira et al., 2019; Lalata et al., 2019), Random Forest (Lalata et al., 2019), and Decision Tree (Lalata et al., 2019), to perform sentiment classification. Others have applied NLP (topic modelling, summarisation, and clustering) using techniques such as LDA (Cunningham-Nelson et al., 2019; Gottipati et al., 2018a), k-means (Luo et al., 2018a), agglomerative clustering (Nitin et al., 2015), and graphical methods (LexRank, PhraseSum, SequenceSum, SimSum, and CDSum) (Luo et al., 2018a,b) to summarise, group or extract aspects or topics from students comments. A few others have employed deep learning models on SET data such as Long Short-Term Memory (LSTM), Bi-Directional Long Short-Term Memory (Bi-LSTM), and Recurrent Neural Network (RNN) (Nguyen et al., 2018; Sindhu et al., 2019; Tseng et al., 2018).

2.8.5 Data Understanding

This stage aims to understand everything about the data by adopting techniques such as Visualisation, Exploratory Data Analysis (EDA), Usability Testing, and A/B testing. Visualisation refers to having an adequate visual representation of the data to make meaningful decisions (Wu et al., 2016). Visualisations make patterns and trends plotted on a graph easy to communicate rather than looking at thousands of rows of figures (Hansen and Johnson, 2011). Several researchers have used visualisations such as dashboards and reports to communicate SET data ideas (Gottipati et al., 2018a,b; Nitin et al., 2015; Pyasi et al., 2018; Santhanam et al., 2018). Gottipati et al. (2018a) recommended an interactive dashboard that can help faculty perform a more in-depth analysis of students' comments to generate topics or sentiments. Nitin et al. (2015) used a user-friendly interactive report to visualise topics and sentiments for SET data.

EDA primarily focuses on visually inspecting the data to understand the patterns inherent in the data (Idreos et al., 2015). Data Exploration applies descriptive statistics (frequency, central tendency, variation, or position) to get a bigger picture of the data (Cunningham-Nelson et al., 2019; Lalata et al., 2019; Luo et al., 2018a), or diagnostic analytics (univariate or multivariate analysis to identify correlation and trends between the dependent and independent features) (Lalata et al., 2019). For instance, Lalata et al. (2019) used the frequency and central tendency to explore the total number of positive and negative sentiments and the average numerical rating per teacher. Also, Cunningham-Nelson et al. (2019) explored the various POS tagging that existed in students comments.

Usability testing is a method of testing the functionality and usefulness of a digital product by observing real users as they try to complete a task on it (Nielsen and Loranger, 2006). In the educational sector, usability and usefulness evaluations have been conducted most often with teachers (Ali et al., 2012; Govaerts et al., 2012; Holstein et al., 2017). For instance, Ali et al. (2012) conducted usability testing for teachers on the LOCO visualisation dashboard and found its perceived usefulness to be high.

2.9 Conclusion

This chapter presents the results of the review aiming to provide insights into TA. The literature also revealed that several tools and methods are available for extracting digital traces associated with teaching, in addition to traditional student evaluation tools. However, one of the main challenges recognised was the cost associated with some devices used to capture in-class activities, and analytics of SET can also provide an alternative to minimise this challenge.

Visualisation of SET data presents possibilities for greater meaning-making and improving the quality of teaching practice and learning outcomes. Additionally, using dashboards to communicate insights into learning and teaching experiences generated from SET data provides stakeholders with personalised, customisable, adjustable and dynamic visualisation that can be easily manipulated and understood by teachers.

While researchers have presented diverse ways to process SET data to generate visualisation reports such as word clouds, sentiments or text summarisation, very little work has been done on presenting both the qualitative and quantitative SET data on dashboards. Dashboards are dynamic and easy to manipulate and drill down into data to focus more on particular aspects of the data to generate greater insights.

The results of this review have led to the development of a conceptual framework for processing SET data. From the analysis of the literature, the researcher proposed a Data Science Life Cycle Framework for processing SET data to guide teachers and researchers to engage with data relating to teaching activities to improve the quality of teaching. This framework is robust and explicitly recognises the need for integrating data, software and other artefacts to generate meaningful insight and benefit the teaching profession.

CHAPTER 3

Methodology

3.1 Overview

This chapter provides a general overview of the different Data Science Life Cycle stages applied to process SET data, and the various methods and phases implemented at each stage. An extensive literature review was carried out to understand the domain and issues around SET data in the business understanding stage in the Data Science Life Cycle model. Ethics was also sought and approved by the University of Otago Human Ethics Committee (Non-Health) to use institutional SET data.

The data acquisition stage followed. However, during the data acquisition stage, the Quality Assurance committee denied access to institutional SET data due to privacy concerns. Hence, using a data simulation technique, SET scores were generated. Subsequently, the simulated SET scores informed the development of preliminary teacher dashboards. The literature also revealed that teachers wanted more than SET scores; they wanted to analyse free-text comments in SET data. This information then motivated the investigation and surveying of teachers' underlying perceptions about the value of SET and visualisation of SET using dashboards. The investigation showed that free-text comments in SET data provided feedback that created valuable opportunities for teachers to reflect on their teaching and improve teaching effectiveness and teaching quality. Real free-text SET data was collected from one academic staff member who voluntarily offered for this research the student comments about teaching experiences and the value and the effectiveness of a course this staff member taught for five years). The qualitative SET data was used to validate the simulated data and preliminary dashboards.

The dashboards were redesigned to Teacher's Evaluation Dashboard (TED)—the deployment and modelling applied at this stage. The deployment stage used the Splunk ® "big data" infrastructure to deploy the data and models and generate dashboards. The modelling stage applied some pre-processing, NLP and ML models to visualise the SET data. Finally, usability testing was carried out on TED, where teachers tested and explored the TED dashboards—the data understanding stage. The rest of the chapter will elaborate on each of the study phases and the methods used.

3.2 Ontology and Epistemology

This thesis has employed a multi-phase research approach that is underpinned by a data science approach. Gruber (1993) defined ontology as an explicit specification of a conceptualisation. Ontology encompasses problems about the most general features and relations of the entities that exist and how researchers use a particular conceptualisation to establish their beliefs within a theoretical perspective (Grix, 2002; Vázquez, 2018).

Over the last decades, educational problems have become more complicated and conventional educational research methods can no longer address them. For instance, Hey et al. (2009) proposed science's evolution through four paradigms. Furthermore, he argues that even though the fourth paradigm is still in its infancy stage, it results in the unfolding data revolution. Contrary to the proposition of Kuhn (2012) on paradigm shifts, which is based on the predominant method of science that cannot account for specific phenomena or answer critical questions, thereby obliging the formulation of new ideas, Hey's transitions are based on advances in data forms and the development of new analytical methods.

Daniel (2017) also supported this view of data science as the fourth research tradition to attend to new approaches and procedures for undertaking scientific research. Emerging research architecture influenced by the technological environment, and complex and diverse data define the latest types of empiricism. The latest style of empiricism reflects the fourth tradition of research methodology (data science). The first tradition of scientific methods is focused on quantitative empirical methods, and is identified as positivist epistemology. Moreover, the second tradition consists of theoretically situated research activities in interpretivism (qualitative methods). Mixed approaches form the third tradition with their underlying epistemology of pragmatism.

> ..."data-ism" is an effort to capture everything as data and derive some knowledge from it, "that everything that can be measured should be measured; that data is a transparent and reliable lens that allows us to filter out emotionalism and ideology; that data will help us do remarkable things —like foretell the future" (Brooks, 2013, para. 1).

The fourth tradition underpins much of the speculation around "data science" and is typically articulated through an empiricist interpretation—that data can speak for themselves. This empiricism is best reflected in the argument of Anderson (2008, para. 13), the former Wired magazine editor-in-chief, whose rallying cry that "big data" represents "the end of science" struck a chord with many commentators. In a controversial essay, Anderson argues that "the data deluge makes the scientific method obsolete", and ultimately the trends and relationships found within "big data" generate substantive and informative knowledge of social, political, and economic processes and complex phenomena, as follows:

> There is now a better way Petabytes allow us to say: "Correlation is enough" We can stop looking for models. We can analyse the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot [...] Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all. There is no reason to cling to our old ways.

Likewise, Prensky (2009, p. 4) argues: "Scientists no longer have to make educated guesses, create hypotheses and models, and check them with experiments and examples based on the data". Dyche (2012) therefore claims that "mining big data reveals relationships and patterns that we did not even know to look for". Similarly, Steadman (2013, para. 14) argues:

> The "big data" approach to intelligence gathering allows an analyst to get the full resolution on worldwide affairs. Nothing is lost from looking too closely at one particular section of data; nothing is lost from trying to get too wide a perspective on a situation that the fine detail is lost [...] The analyst does not even have to bother proposing a hypothesis anymore.

In this thesis, these ideas were realised in the development of teacher dashboards underpinned by a data science approach that focused on SET data as a form of TA. It represents five critical stages relying on ongoing negotiation of knowledge working iteratively, from the initial business understanding stage, data acquisition, deployment, and modelling to the data understanding stage. The purpose of this approach is to underscore the importance of engaging in new ways of undertaking educational research to empower the teachers with tools to help them make informed decisions to improve the teaching profession.

3.3 Methods

This section describes the methods used in this research to collect data, perform analysis and usability testing on the dashboards developed to improve teaching quality.

3.3.1 The Business Understanding Stage

At this stage, ethics permission was also sought from the University of Otago Human Ethics Committee to collect and use institutional SET data. An extensive literature review was carried out concurrently to understand the domain of SET data. Subsequently, part of the literature findings led to surveying teachers' perceptions of SET data. In this study, the researcher carried out an extensive literature review to understand the domain of SET data.

3.3.1.1 Survey

Previous studies in the literature have enumerated many benefits and shortcomings using SET to improve teaching quality. The analysis, interpretation and reporting of SET data are still crucial concerns for both institutions and educators (Boring, 2015; Braga et al., 2014; Hessler et al., 2018; Rathke and Harmon, 2011; Spooren et al., 2017; Stupans et al., 2016; Subramanya, 2014).

Accordingly, further investigation was carried out by surveying academics perceptions of SET data and reporting them using dashboards. Questionnaires employed both closed-ended and open-ended question types in this study to investigate how teachers used teaching evaluation data and what they perceived about the use of dashboards as a tool to visualise SET data for better data interpretation. This questionnaire contained 12 questions, including single-choice questions, multiple-choice questions, open-ended questions and Likert-scale questions that were rated on various scales ranging from 3 to 5 points (see Appendix B).

3.3.1.2 Sampling Procedures

This research employed a purposive expert sampling method (Etikan et al., 2016) to solicit and recruit the academic participants in this study. An online questionnaire was designed using an online data collection program (Qualtrics ®) to request that participants fill out their teaching evaluation perceptions of SET data and dashboards. Between August 2019 and October 2019, academic participants were solicited to provide more effective ways to present the SET data results to advance teaching quality. Solicitations were sent to all the academic staff at the University of Otago with teaching responsibilities, an estimated number no less than 800 (*refer to University of Otago Quick Staff and Student Statistics*), alongside soft copies of the consent forms and information sheets. A total of 86 participants responded to this questionnaire. They were then thanked and invited to include their email address in the questionnaire to express interest in participating in a further usability study. Subsequently, about two to three months later, between October 2019 and February 2020, interested participants received emails inviting them to schedule a time to participate in the usability study.

3.3.1.3 Data Analysis and Interpretation

This study performed analysis using SPSS ®, Python and R programming languages to examine the data distributions. Then, the Spearman's rho correlation analysis and chi-square test of association were employed to determine the relationships that existed between the various features in the data, including teaching experience, frequency of evaluation, beliefs about using SET data to improve teaching quality, and experience in using dashboards on SET data. Furthermore, thematic analysis was carried out using a constant comparative technique (Dye et al., 2000) to investigate responses to open comment questions. The process involved seeking descriptive and theoretical links between teachers' perceptions and beliefs about SET data's value. For each open-ended question, key themes were identified (alongside a statistical analysis of their occurrences), and then similar themes that emerged were grouped.

3.3.2 The Data Acquisition Stage

This stage reveals how the data used for this research was acquired. SET scores were simulated to generate the "Number Rating" dashboards while real free-text data were collected from an academic to generate the "Open-Ended" dashboards.

3.3.2.1 Data-Driven Simulation

After the University of Otago Human Ethics Committee approved the ethics to use institutional SET data (see Appendix C), the data acquisition stage commenced. However, this stage faced the challenge of access denial to institutional SET data by the Quality Assurance Committee board. Ethical issues, such as consent, privacy disclosure, and the need to de-identify data were paramount concerns. Consequently, SET scores were simulated using a random probability distribution technique (Ndukwe et al., 2018), and preliminary dashboards generated. Appendix D, Appendix E, and Appendix F describe the probability sampling method used to randomly generate SET scores.

3.3.2.2 Real Data

The survey results reinforced results in the literature where academics found the free-text comments in SET more informative than the SET scores, and provide useful feedback to inform improvements in teaching, curriculum design and assessment (Brockx et al., 2012; Gottipati et al., 2018a). Hence, teachers wanted to go beyond visualising SET scores to also include comments in SET.

One academic staff member voluntarily offered real students' free-text data (on student comments about teaching experiences on the value and the effectiveness of a course this staff member taught for five years). This free-text data was collected and used for this research. Data pre-processed and cleaned were carried out on this dataset using Python programming language to anonymise persons and places (department or faculty names) that appeared in the comments. The qualitative SET data was used to validate the simulated data and preliminary dashboards.

3.3.3 The Deployment and Modelling Stage

The survey also revealed that teachers wanted to visualise SET data using dashboards. Hence, the development of TED resulted in a modification and redesign of the preliminary dashboards created to visualise the simulated SET scores. TED adopted and adapted the Worsham's NLP text analytics application to visualise the real students' free-text feedback (Worsham, 2018). This combination makes TED a powerful tool that visualises both SET scores and the free-texts in SET.

TED app was deployed in the Splunk ® "big data" platform, which automatically performed the data pre-processing, ran the ML models in the background and displayed the visualisation dashboards. The data pre-procession was automatically applied to the raw text SET data, including tokenisation, upper-case conversion, special character removal, stop words removal, stemming, POS, and NER. The Bag-of-Words matrix of the TF-IDF feature engineering technique was also applied to evaluate important words in student comments. Furthermore, NLP and ML models such as LDA, Latent Semantic Analysis (LSA) and Non-Negative Matrix Factorisation (NMF) for topic modelling; k-means, birch, spectral clustering, density-based spatial clustering and X-Means for clustering; and the VADER for sentiment classification were used.

3.3.4 The Data Understanding Stage

Two-iteration usability testing was carried out to evaluate teachers' perceived user experience regarding the usability and usefulness of TED. A total of twenty-three (N = 23) participants partook in this study, with twelve participants (N = 12) in the first iteration and eleven participants (N = 11) in the second iteration.

The usability study had each participant perform a simple task and undergo a short interview. Subsequently, collective feedback from the usability study on the initial version of the TED prototype (first iteration usability study) was used to improve the second version of the TED prototype's design and development. After that, the usability study was performed on the improved TED prototype (second iteration usability study). A System Usability Scale (SUS) (Brooke, 1996) questionnaire was administered to participants at the end of individual sessions.

The usability testing session lasted between 30 minutes to 1 hour per participant. However, two participants requested merging their usability study session on two occasions, which resulted in 21 separate usability testing sessions altogether. These participants were identified as p18–19 and p22–23. Both the In-person Moderated Usability Testing and the Remote Moderated Usability Testing were organised and recorded on Zoom®. A total of 20 participants had In-person Moderated Usability Testing, while 3 participants had Remote Moderated Usability Testing. During each usability testing session, the researcher introduced participant(s) to the TED and gave them a short tutorial on using the TED Number-rating dashboard (made up of both the Aggregate dashboard and the Comparison dashboard). Participants were given short tasks to perform on the dashboards (see Appendix G, for more information about the Task Protocol). Subsequently, the researcher performed another collaborative task with the participants, where they collectively explored the Open-ended Comment dashboard of the TED (see Appendix G).

Consequently, the participants were interviewed about what they thought of this dashboard at the end of each task. In addition, questions about how useful the dashboard would be to them, and how the dashboard can be improved guided the investigation. At the end of each session, participants were emailed a link to the SUS questionnaire to determine the usability score of TED. A total of 20 participants responded to the SUS questionnaire; eleven respondents (N = 11) participated in the first iteration, while nine respondents (N = 9) participated in the second iteration. See Appendix H for the sample SUS questionnaire used for this study. The interview data were transcribed using the Otter.ai® online transcription software. Additionally, abstracts from individual transcripts were extracted and mailed electronically to each of the participants involved to eliminate misinterpretations and ensure data quality.

3.4 Summary

This chapter discussed the researcher's ontological and epistemological terrain, and the various methods and tools the researcher used in the thesis. The researcher applied a random probability technique to generate SET scores and preliminary dashboards. Then preliminary and further investigations were conducted on the dashboard by surveying academics' perceptions of SET and dashboards for visualising SET. Because simulated SET scores are limited and do not reflect reality, real SET data was collected from an academic who volunteered five years of free-text SET. Subsequently, TED was designed to visualise students' free-text in addition to SET scores. Finally, a usability study was conducted on the dashboard to determine the perceived usability and usefulness of TED.

CHAPTER 4

Using Student Evaluation of Teaching Data

4.1 Overview

Prior studies have shown that teachers wanted more than SET scores;they wanted to analyse the free-text comments in SET data. This chapter presents the results of the further investigations carried out by surveying of academic staff to gain an overview their perceptions of SET data. Key findings suggest that participants view SET data as a useful tool for improving teaching quality; however, they stated some limitations of using SET. This research also explored teachers' familiarity with the use of dashboards as a tool for better data interpretation. The results indicated that participants perceived that using dashboards to report SET data would be valuable; however, they had very little or no experience with the use of dashboards.

4.2 Introduction

SET is globally adopted in the educational sector (Macfadyen et al., 2016). According to Kite et al. (2015), SET is a teaching evaluation where students can evaluate essential aspects of teaching. SET data present results that are primarily used by the teachers to improve on teaching practice (Mortelmans and Spooren, 2009), assess how effectively courses are taught (Beleche et al., 2012), and inform academic promotion decisions (Oon et al., 2017).

This chapter explores how teachers engage with the results of students' teaching evaluations to improve their teaching practice. A questionnaire consisting of 18 items (see Appendix B) was sent to 818 academic staff at the University of Otago, where a total number of eighty-six (N = 86) participants responded. Demographic information provided by SPSS® suggested that 12% of respondents are professors, 24% associate professors, 35% senior lecturers, 18% lecturers and 10% others, including research fellows, professional practice fellows, senior teaching fellows and part-time, fixed-term lecturers. Additionally, 23% of the participants came from the Sciences division, 28% Humanities, 23% Health Sciences, 8% Commerce and 19% others, such as social sciences, arts, academic and freelance. Also, 4% of respondents were under 25 years, 4%

aged 26–30, 22% 31–40, 20% 41–50, 46% 51–55 and 4% were older than 66. Finally, 4% of respondents had 0–5 years of teaching experience, 28% 5–10, 13% 10–15, 17% 10–15 and 39% had more than 20 years of teaching experience.

4.3 The Utility of SET Data

SET is still the primary measure that institutions use to get students' opinions about teaching. Academics use the information SET provides to make changes to their teaching. A critical issue concerning SET is the extent and manner to which they are actually used. This section investigates participants' engagement level and frequency of evaluating SET data.

Participants were asked how often they evaluate their teaching. The responses ranged from 1 (always perform teaching evaluations) to 5 (never perform teaching evaluations). Results show that more than three-quarters of the participants indicated that they frequently evaluate their teaching (90% 1 or 2 ratings). On the other hand, about a quarter of the participants indicated that they rarely evaluate their teaching (10% 3, 4 or 5 ratings). Six themes were identified from the responses participants provided for or against frequently performing teaching evaluations (see Appendix I).

Participants who perform frequent evaluations indicated that improving teaching practice is an essential contribution SET brings. They like to have feedback on whether their changes have had the desired and anticipated effect on students. For instance, participant p9 said, "I am constantly tweaking my workshops, and student evaluation data is one way of getting some feedback to see if these changes have the desired effect". Others reported that SET results are used to judge academics on their performance, which directly affects their careers. For example, participant p43 stated, "we are required to do so for confirmation, progress and promotion".

It is also important to note that several participants who frequently collect evaluation data indicated that evaluation data provides them with the opportunity to gather informal feedback on how their students learn. For example, participant p25 stated, "I believe it is important to be in touch with the student experience. However, I also rely on informal conversations with students, and a journal assessment to monitor how things are going". They also argued that this sort of informal feedback from students contributes immensely to course content development. For instance, participant p53 said, "... giving students opportunities to express views/feedback, to be aware of any reactions to changes in course content, assessment or teaching approaches, to take into account differences in class profiles (e.g., change in the mix of domestic and international students or background of students)".

Participants who performed frequent evaluation also indicated that frequent evaluation is likely to trigger survey fatigue and affect response rates. For example, respondent p37 said, “I try to evaluate 1–2 blocks of teaching each year, bearing in mind that students are overwhelmed with all of their lecturers seeking [the] same feedback”. In contrast, those participants who rarely conduct evaluations believe that low response rate is likely to encourage teachers, cause them to give high grades, and lessen course work to win the few students’ approvals that respond to the survey. For example, participant p44 said, “response rates are low, the surveys are somewhat limited and what you can find out, and there is more to good teaching than good evaluations, so other things need to be considered”. As a result, some respondents indicated that they preferred to evaluate other teaching components. For instance, respondent p32 remarked, “I try to evaluate innovations frequently and established courses at least every three years”. Also, comments of participants who rarely performed SET suggested that once the desired comfort level has been achieved with the teaching role, complacency creeps in. For example, participant p42 stated, “I already have a good idea of what the outcome will be”. Pre-empting outcome may be a sign of complacency, which is not good enough to motivate change.

Furthermore, respondents were asked how often they use the insights generated from teaching evaluation data to inform their teaching. Responses ranged from 1 (always use information from teaching evaluation data to inform teaching) to 5 (never use information from teaching evaluation data to inform teaching). Results show that approximately three-quarters of the participants indicated often using teaching evaluations to inform their teaching (71%, 1 or 2 ratings). Another quarter of the respondents reported that they rarely use teaching evaluation to inform their teaching (24%, 3 or 4 ratings). A small number, less than a quarter, indicated that evaluations never inform their teaching in any way (5%, 5 ratings). Three themes were identified from the respondents’ responses about the last time teaching evaluation data was used to improve their teaching practice (see Appendix J).

Participants who frequently performed evaluation said it helped them in understanding student learning. For instance, one participant p51 said: “the results allow me to determine if my current actions are creating an environment of significant learning that will meet the intended learning outcomes. The results will alter my actions accordingly to meet this outcome”. However, some participants who rarely performed evaluations indicated some level of hostility towards SET, as expressed by one participant:

> *. . . I now view the student evaluations as more akin to online trolling than anything else. Although we are required to do them, a couple of years ago, I made an active decision never to look at them. This decision has made me happier and far more confident in my teaching. I seek personal, face-to-face feedback, feedback through class reps, and I assess my teaching success through students’ results—all these are useful.* (p15)

This statement confirms that academics' acquiescence to SET as part of current tertiary environments often does not convert into using them to improve teaching.

Many participants valued SET and the significance of undertaking it regularly, participant p31, "I use evaluations to gauge student opinion about my teaching. Based on this, I adjust what I do in the classroom. It is a continuous and iterative process". However, some negative views were also expressed, especially with regard to students' open responses. For example, respondent p18 recounted, "constructive criticism or suggestions are rare". Hence, participants questioned the quality of student judgements at the institution. One participant who rarely performed evaluation voiced reservations concerning the quality of student feedback:

> *. . . It depends on the quality of the data. If they just say that they liked/did not like the course, there is no real feedback to implement changes. They need to provide some constructive feedback, such as "I particularly liked x topic/exercise" or "I struggled with z" so then I know which aspects of the course work/does not work.* (p45)

Others remarked that students' response rates using the online system is low and may result in low-quality data. For instance, one participant said:

> *I consider all the feedback, but some of it I find more useful than others. I do not trust the student evaluations as much as other forms of evaluations, because there is often such a low response rate—especially after going to online-only versions. The low response rate seems to polarise the evaluations, and there is no way to know whether students filling them out have attended the class.* (p21)

Data fusion to link students anonymously to their attendance, grades, activities and SET is a possible solution that could be implemented to help the teacher know if students who filled in the evaluation attended the classes and their activities, and the level of their performances in class.

Respondents were asked the last time they used teaching evaluation data to improve their teaching. The scale range used is as follows: 1 (last week), 2 (last month), 3 (last semester), 4 (last year) and 5 (never). The result shows that approximately three-quarters of respondents indicated using evaluations to improve teaching within the last week, month or semester (73% 1 or 2 or 3 rating). About a quarter chose last year (22%, 4 rating). A small fraction indicated that they have never used SET to improve their teaching (5%, 5 rating). Three themes were identified from the respondents' responses about the last time teaching evaluation data was used to improve their teaching practice (see Appendix K).

Participants who recently used evaluations to inform teaching claim to be interested in checking their teaching and course experiences with students. Most of them were interested in ideas to fine-tune their courses towards improving the experiences of students; participant p44 explained, "I think about my course and each time I teach it, I wonder and consider ways to improve the content, make it more relevant and think about better ways of communicating what I would like students to think about, and to find out their expectations of the course". However, participants who rarely perform evaluations expressed some concerns about carrying out SET towards the end of the course work, making it difficult to report SET results back to current students. This tradition of disconnected relationship with SET data is evident, and a strong reason not to give students feedback. For instance, participant p43 noted, "most of our classes are only taught yearly, so there is no chance to implement them until the next teaching period". Moreover, respondent p47 reported they "had just heard from class reps". This expression is not surprising; in other words, some participants prefer to receive additional feedback to mitigate some challenges of the formal evaluation, in addition to the formal evaluation being performed less frequently, compared to other forms of evaluations, such as Peer Observation (POB) or Informal Discussions with Colleagues (IDC).

A pertinent recurring theme in this study is the value of SET in improving the quality of teaching. According to Stiggins (2017), SET data are routinely used in several educational sectors for instructional improvement and personal decisions. This accountability trend has put considerable pressure on academics aware of the need to use the information in their decision-making processes. Participants who support SET data tend to favour SET data as a source of valuable feedback. However, some participants continue to challenge SET data, claiming that it is used more for personal decisions, such as confirmation, promotion and progress. In general, the data thus confirmed a gap in the quality of engagement with SET data. There is a deliberate and systematic use for professional development and ongoing engagement with students regarding their feedback and how it is valued and used.

4.4 The Value of SET Data

SET is a common practice across the educational sector and was initially introduced in the 1920s to improve teaching practice. The instrument continued to evolve, and its use was extended to performance management practices (Galbraith et al., 2012).
Table 4.1 illustrates the detailed information on the types of teaching evaluations collected by participants. Respondents were asked the different categories of data they collect and use for teaching evaluation. All of the participants (n = 58, 47%) indicated that they collect SET data, three-quarters of the participants (n = 35, 29%) indicated that they collect IDC data in addition to SET data, a third of the respondents (n = 18, 15%) indicated that they collect POB data in addition to SET data. Only a few participants

Table 4.1 Multiple Response Analysis on Categories of Teaching Evaluation Data.

Teaching Evaluation Categories	**n (%)**
Student Teaching Evaluation (SET)	58 (47)
Informal Discussions with Colleagues (IDC)	35 (29)
Peer Observation (POB)	18 (15)
Others	11 (9)
Total	**122 (100)**

(n = 11, 9%) indicated that they collect other teaching evaluation data (such as online activities, teaching and learning circles, informal student evaluations, self-reflection data, attendance data, session feedback data, dialogue and discussion with students). Six themes were identified from the participants' open-ended responses concerning the different categories of data they collect for teaching evaluation (see Appendix L).

Several participants expressed attitudes suggesting that university requirements and standard practices of SET have affected their perception of the standard teaching evaluation process. For example, participant p39 said, "I do not find the current student teaching evaluations useful, but am forced to collect them". Some participants gave other reasons they carry out the standard evaluation process, such as promotion, progression, and confirmation requirements. For example, p46 stated, "I collect student teaching evaluation data because it is a requirement for confirmation and promotion". However, most respondents showed endorsement of the formative application of SET; they claimed that it could be a useful tool for gathering feedback from students. For instance, this comment from p31: "I would like to know the opinion of my audience in regard to the usefulness and effectiveness of my teaching".

Participants who perform SET have strong beliefs in TPD and teaching quality. For example, respondent p16 stated "I believe that teaching is the most valuable thing that the academic staff of our University do, even if the irrational demand for constant research is both monetised and weaponised against the need for constant improvement in our teaching and approaches". Additionally, participants like to garner a diverse range of feedback to improve TPD. Some of the reasons are given in remarks made by participant p51: "the wide range of data allows me to understand the students' current, as well as past and future status, understanding, concerns and expectations about the learning experience. Multiple sources help triangulate and get views from a wide range of students".

Although participants highlighted some benefits inherent in the institution's online evaluation system, some interests were also expressed in giving preference to traditional evaluation, due to the perceived view that it will increase the students' response rate. For example, one respondent stated:

. . . we collect feedback via a paper survey two to three times during the semester. Paper surveys in lab increase response rates. However, in the didactic course with multiple instructors, using the University online end of paper evaluation is preferable, due to ease of use in evaluating a paper, with many instructors and a diversity of topics. (p19)

More detail on the perceptions of teachers on SET was seen by examining how the participants rated SET data to improve teaching; responses ranged from 1 (very high) to 5 (very low). Approximately half of the participants (55%, 1 or 2 rating) indicated high satisfaction with SET as a means of improving teaching, a quarter chose the middle ground (16%, 3 rating) and a third of the respondents (29%, 4 or 5 rating) indicated low satisfaction with SET as a means of improving teaching.

Further insights were also gained from what participants thought was the most important reasons for using teaching evaluation. Results show that approximately three-quarters of participants (69%) indicated that "improving teaching outcomes" is the most important reason for using teaching evaluation, and a quarter (25%) chose "promotion". However, a minimal number of respondents (6%) indicated that "learning about teaching" is the most important reason for using teaching evaluation. Three themes were identified from the participants' open-ended responses concerning the most important reasons for using teaching evaluation (see Appendix M).

Most participants identified SET as a means to improve teaching outcome. They noted positive changes, such as those noted by participant p3 who remarked, "evaluations help me improve the course so that it benefits students better in terms of what they are looking for when they attend the course". Another participant p19 elaborated: "the goal of teaching is helping students to learn. Hence, improving teaching outcomes (that is, students' learning) is most important. As part of this, you learn about teaching". Other participants against SET questioned students' ability to judge teaching, indicating student bias as a significant concern. A view was expressed by participant p54: "it would be useful if students put their name to comments so that the comments were more constructive feedback—when you get groups of students saying the same thing out of anger or frustration it is not helpful". Some proponents of SET for improving teaching, also acknowledged that most people believe that promotion is the primary purpose of SET, with some valid reasons. For example, participant p21 recounted that "I think to improve teaching is the most important reason, but also think that this is impeded by the poor quality of data in teaching evaluations. So it turns into a process for promotion rather than about teaching quality". On the other hand, some participants debated against this view of SET for promotional purposes:

I find it most useful when students give open-ended comments about what do or do not work for supporting their learning. The numerical metrics *used in promotion are not particularly meaningful as it depends on how many people respond, and who responds. Usually, it is the people who particularly loved or hated their experience. I would find it more useful to capture the average students in the middle who just thought it was "kinda OK".* (p45)

Other concerns mentioned were related to unease about SET data's institutional use, including data quality issues, students' low response rates, and institutional dependence on one measurement source to base assumptions and decisions. For example, one participant noted:

> *we have an online system with mainly generic items that are not useful to me as a teacher, the repetitive nature of these bore students, and this combined with low response rates means that despite being an advocate for gathering student feedback and being interested in student feedback I do not think the current system works to provide quality feedback.* (p53)

It is also important to mention that a few other participants viewed teaching evaluations as part of a shared teacher-learner relationship in which both parties have a significant stake. For example, one participant remarked:

> *My role is about creating a space for significant learning to happen that meets the intended learning outcomes. The data collected via teaching evaluations helps me determine whether I am doing that, what I am doing wrong, and what I could do differently to improve my facilitation of an environment that creates significant learning moments in the students.* (p51)

SET data is a standardised method for assessing and rating a teacher's effectiveness and teaching performance. It is the most immediate and the most widely used strategy administered by the education sector to improve teaching quality, and provides teachers with constructive feedback to direct their TPD and personal decisions, such as promotion and confirmation (Huxham et al., 2008; Linse, 2017). In line with much of the current literature, the results discussed in this section showed that most participants thought collecting SET data was worthwhile for improving teaching practice (Neumann, 2000). Additionally, the high responses for improving teaching, followed by promotions, indicate that these two primary purposes for which evaluations exist, namely, for accountability and professional development, are sensitive.

4.5 The Application of SET data

The use of tools such as teaching analytics dashboard in the educational sector is an emerging field. Epp and Bull (2015) proposed data visualisation that minimises ambiguity and maximises effect and intervention through graphical variables (such as graphs and maps), visualised data interpretability, and learner modelling (ensuring user groups understand how to interpret data). Also, Wojton et al. (2014), in their study of visualisation dashboards, observed that, apart from transparency, data visualisations that provide context relating to previous experience are more likely to be successful.

Participants were asked to indicate the forms of teaching evaluation data they accessed to investigate the various teaching evaluation data forms. Table 4.2 illustrates that approximately a third of the respondents (n = 47, 38%) indicated that they access the summary statistics part of the teaching evaluation data. Approximately another third of the respondents (n = 35, 29%) indicated that they access the text part of teaching evaluation data. A quarter of the participants (n = 30, 25%) indicated that they access evaluation data in its raw form. Additionally, another 10, (8%) of the participants indicated that they accessed other forms of teaching evaluation data such as peer observations, discussion with colleagues and informal evaluations.

Table 4.2 Multiple response analysis on the forms of teaching evaluation data accessed by participants

Teaching Evaluation Forms	**n (%)**	**%cases**
Summary statistics (tables, graphs, proportions, etc.)	47 (38)	81
Textual form	35 (29)	60
In raw data	30 (25)	52
Others	10 (8)	17
Total	**122 (100)**	**210**

Furthermore, participants were asked to indicate specific tools they used to access teaching evaluation data. Table 4.3 shows that approximately half of the respondents (n = 40, 46%) indicated that they engage with teaching evaluation data in PDF forms. About a third of the participants (n = 29, 34%) indicated that they engage with teaching evaluation data in MS Word format. Approximately a quarter of the participants (n = 15, 17%) indicated they use Excel in engaging with student evaluation data. Only a tiny percentage of the participants (n = 1, 1%) indicated that they use other applications, such as SPSS, Tableau and Python, to engage with teaching evaluation data.

Respondents were then asked how confident they were in interpreting SET data. A scale that ranged from 1 (Confident) to 5 (Not Confident) was used. Results show that approximately one-third of the participants (85%, 1 or 2 rating) reported that they were confident interpreting SET data, a quarter of the respondents (13%) chose the middle range. A minimal number (2% 4 or 5 rating) reported that they had no confidence in interpreting SET data.

Furthermore, participants were asked how likely it was that they would require others' support to help with the interpretation of SET data. A scale that ranged from 1 (very likely) to 5 (very unlikely) was used. The results show that approximately a quarter of

Table 4.3 Multiple response analysis on tools used to engage with teaching evaluation data.

Teaching Evaluation Tools	**n (%)**	**%cases**
PDF Document	40 (46)	69
Word Document	29 (34)	50
Excel	15 (17)	26
SPSS	1 (1)	2
Tableau	1 (1)	2
Python	1 (1)	2
Total	**87 (100)**	**151**

the participants (13%, 1 or 2 rating) reported that they were likely to need assistance to interpret SET data. About another quarter of respondents (20%) were indecisive. One-third of the participants (68% 4 or 5 rating) reported that they were unlikely to need assistance to interpret SET data.

Additionally, participants were asked about their knowledge of dashboards. The result shows that approximately three-quarters of the participants (82%) indicated that they do not know anything about dashboards. The remaining quarter had some sort of experience of dashboards (18%). To further understand the result, three themes were identified from the open-ended responses provided by respondents about their knowledge of dashboards (see Appendix N).

Only a few participants had some knowledge of dashboards. However, they maintained that they had not used teaching analytics dashboards in the institution. For instance, participant p33 stated, "I have not experienced the use of dashboards at this institution". However, several participants echoed that they were not aware of dashboards. Additionally, some of them seemed not to be satisfied with how SET results are presented in the institution and reported limited SET data access. For instance, participant p36 agreed: "never heard of it. We usually just get emailed the student evaluation summaries in PDF form".

Furthermore, participants were asked if they would use a teacher's dashboard to visualise SET data. The scale used ranged from 1 (very likely) to 5 (very unlikely). The result showed that approximately, a third of the participants (32%, 1 or 2 ratings) indicated that they were likely to use TED to visualise SET data. About half of the respondents (43%) were indecisive. Only a quarter of the participants (25%, 4 or 5 ratings) indicated that they were unlikely to use TED to visualise SET data. Four teams were identified from respondents' open-ended responses about using TED (see Appendix O).

TED supporters claimed that qualitative input would bring significance to quantitative data issues that arise. They agree with the idea of incorporating qualitative student feedback data into TED to determine the level of student opinion positivity and negativity to enable academics to analyse discrepancies with quantitative outcomes, or even make it possible for academics to compare qualitative scores with quantitative scores. For example, participant p8 said, “sounds a good idea, and I am guessing presents a useful summary of the data. However, it is still the comments that are the most informative”. Some participants remained sceptical and feared that TEDs could be another form of institutional auditing tool to monitor performances rather than improve teaching performance. For example, participant p44 responded, “I would like to know more. It would be good if TED helped assess my teaching, but there is a risk that it is just another way of collating information for promotion and progression”.

Participants who had a neutral opinion about TED expressed concerns that students could be easily swayed by accessible courses and likeable teachers, affecting the data quality and visual representation of SET via TED. A view was expressed in a statement by participant p19: “concerned that teaching evaluations from students alone sometimes can be like popularity contests. Students sometimes do not distinguish the likeability of a teacher from the effective teaching and learning”. A few others argued that SET data was simple enough and did not require an extra tool to interpret their outcomes. For example, one participant stated:

> *The data I collect from the InForm system is sufficiently straightforward that I can use the statistics and the free form comments for the ends that I need to meet. If the evaluations were more complex, or if my data collection methods were limited to the numbers/bubbles, or if my classes were so large as to make the collection of those statistics meaningless (because above a certain size, classroom learning has not been proven more useful than no learning at all), then I suppose I could use more robust analytical tools.* (p16)

Another portion of participants in this neutral category indicated conditional acceptance of TED based on perceived usefulness. For example, participant p51 said, “It depends how user-friendly it is, and can it collate and interpret the sources of evaluation I have just described before for me”. Additionally, respondents who indicated they were not interested in TED were either not keen to migrate to other platforms or valued metrics other than SET. Interestingly, they questioned the usefulness of TED, for instance, p46 echoed other comments: “as I am unfamiliar with it, I am not sure if it will be useful”.

Comprehension and enhancement of teachers’ professional vision (Goodwin, 2003) are essential yet recognised research subjects within analytics in the educational sector. TA (Vatrapu et al., 2011) aims to tackle this research gap by involving teachers in the design, development, and evaluation of visualisation dashboards for academics to visualise institutional data and enhance their pedagogical decision-making, informed by data literacy skills.

Adequately reflecting on teaching practice without sufficient information is not an easy process for teachers. This lack of data access is a system gap that could be mitigated by introducing visualisation dashboards to link multiple data, including SET data, to encourage teacher data literacy. From the responses, it is evident that several of the participants have not experienced any form of dashboard. Visualisation dashboards could be used to dynamically present SET data in real-time, instead of using static PDFs, sent as emails on request. Consequently, the institutional systems need to encourage academics to use dashboards to reflect on their teaching practice, critical for TPD.

4.6 Chi-Squared Test of Association and Fisher's Exact Test

A chi-squared test of association was conducted to explore associations between group participants' perceived value of teaching evaluation and why they perform evaluations. The results of the test (3 $\times$ 3) shows that there is a statistically significant association between participants' perceived value of teaching evaluation and why they perform evaluations $(\chi^2(2, 48) = 11.11, p < 0.05)$. The association was strong (Cohen, 1988; Green and Salkind, 2016), Cramer's V = .340 (see Appendix P).

However, the chi-squared test of association demonstrated a violation of the assumption, where all cells should have expected counts greater than or equal to five. This violation means there is not an adequate sample size to run the test; hence, the risk of making a wrong decision is maximised (Kroonenberg and Verbeek, 2018). Consequently, to minimise this risk, Fisher's Exact Test was later conducted. The result showed a statistically significant association between teaching evaluation and why they perform evaluations, p = .022 (see Appendix P).

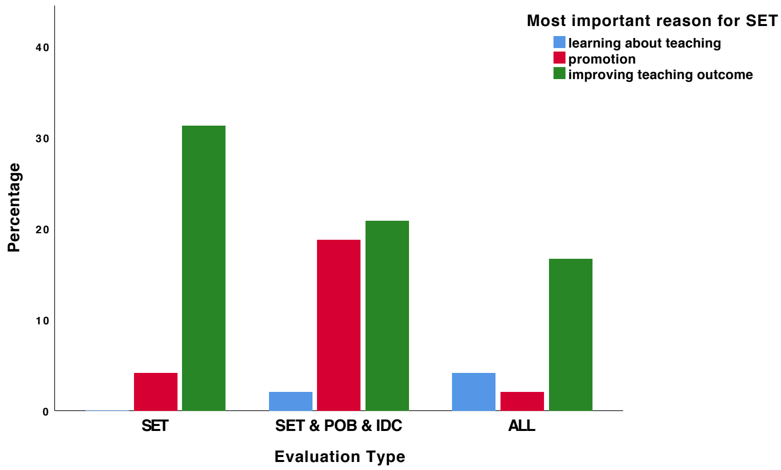


Fig. 4.1 The value of SET data towards enhancing teaching quality. Key: SET: Student Evaluation of Teaching, POB: Peer Observation, IDC: Informal Discussions with Colleagues

To further demonstrate the value of SET data, Figure 4.1 shows that even though few participants use SET only for promotion purposes, and are not necessarily concerned about improving teaching quality, a significant number of respondents valued SET data and acknowledged using it to improve teaching quality.

4.7 Spearman's Rank Correlation Analysis

Spearman's rho correlation test was conducted to understand the data and determine if there is a relationship between teaching experience and teaching evaluation data. The result of the Spearman's rank correlation analysis using an alpha value of 0.05 showed that there is a significant positive correlation observed between the teaching experience and teaching evaluation data collected by participants $(r_s(56) = .336, p < 0.05)$ (see Appendix Q). The correlation coefficient between teaching experience and the frequency of evaluating teaching was 0.34, indicating a moderate effect size. This correlation indicates that as teaching experience increases, SET data's use to improve teaching quality also increases. Additionally, participants who used other forms of evaluation data argued that SET data could be fused with diverse data (such as student attendance, performance and engagement) to enhance teaching quality and support teachers in making informed decisions. For instance, one participant revealed:

> *I think it is very important to be able to match responses in different questions. For example, if I ask how many lectures the student attended and how effective I have been as a teacher, I want to pair the responses. I suspect that students who attend very few lectures tend to rate my effectiveness lower than those who come to class; this is important to know when assigning credit for good teaching.* (P34)

However, some participants expressed concerns about potential manipulation of the evaluation process by academics, and it is essential to have multiple points of feedback. For example, participant p7 recounted that "it has to be triangulated to be valid. It is a human factor and skills to get good scores. I know what to do to manipulate and get high scores; pretend to be extremely nice to students, pretend you care, and you get good scores". Others raised the issue of student evaluation bias. For example, respondent p15 noted that "The online evaluation system now closely mirrors the conditions of anonymous online comment. These conditions seldom produce thoughtful feedback, and they far too often invite vengeful, abusive and discriminatory comments".

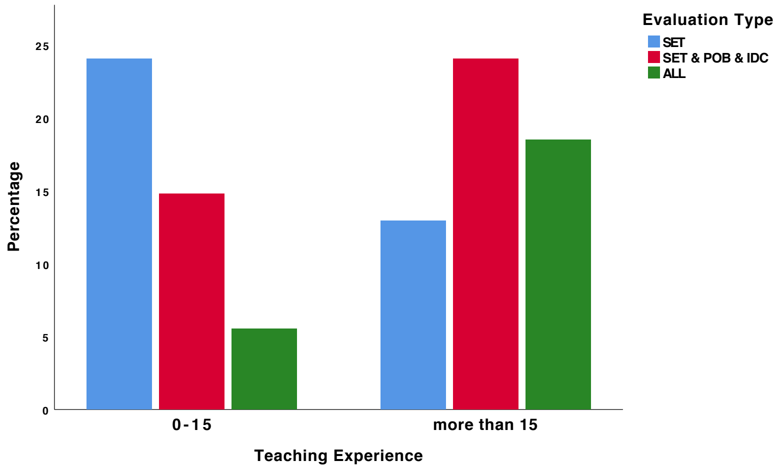


Fig. 4.2 Participants with less Teaching Experience carried out Teaching Evaluation more using SET. Key: SET: Student Evaluation of Teaching, POB: Peer Observation, IDC: Informal Discussions with Colleagues.

Figure 4.2 further illustrates that participants with more teaching experience collect more of other forms of teaching evaluation data. On the other hand, participants with less teaching experience collect more SET data. Hence, this result reveals that even though SET data is a valuable tool that could improve teaching quality, teachers with less experience are more likely to be more interested in SET data for promotional purposes. In comparison, teachers with more teaching experience are more likely to understand different aspects that can better inform their teaching practice. Consequently, revealing another dimension of linking SET with other data sources.

4.8 Discussion and Conclusion

This study investigates academics' perspectives on the usefulness of SET data for both summative and formative purposes. As seen in previous sections, SET remains a controversial subject among teachers in the educational sector. A commonly repeated caution in this study relates to the value of diversifying the approach to evaluation by fusing SET with other data sources. Veteran scholars all accept that SET can provide only one aspect of a more extensive system to provide unbiased evaluations to promote TPD (Abrami et al., 2007; Marsh, 2007; Feldman, 2007). Similarly, Harvey (2011) endorsed SET as one of the most important tools in the educational sector's ongoing advancement, but emphasises that it is just one tool and should never be used as the sole source of evidence.

Many influential scholars and professionals make a case for a more comprehensive approach. For instance, Ramsden (2003) forewarned against over-dependence on any collection of SET results, suggesting that greater reliability is achieved when SET data are used in combination with information from other sources to achieve linked data and overall improved data quality. Similarly, Berk (2005) reported the use of SET as only one of several possible methods for assessing teaching performance, although recognising that SET data is a critical source of evidence for both formative and summative decisions. He proposed that fusing SET data with other data sources could enhance evidence-based decision making, using educational data to assess teaching quality.

The inconsistency of the participants' responses is consistent with the debate concerning the validity and quality of the SET data found in the literature (Coffey and Gibbs, 2001; Flaherty, 2016; Harland and Wald, 2018; Linse, 2017; Palmer, 2012; James et al., 2015). The results seem to point to a state of general uncertainty surrounding SET data, which affects the respondents' attitudes. For instance, tension can be seen from the respondents who did not favour using SET data for personal gain such as promotion or confirmation. In their view, they can only support it if the data are used for administrative decisions. The stress expressed here is possibly related to this situation's uniqueness for participants in the institution and their general uncertainty about promotion.

According to Marsh and Roche (1993), the motive of teaching evaluation is to assess the teachers for the following purposes: to provide formative feedback to teachers on the quality of teaching to bring about change (Oon et al., 2017); as a summative measure of teaching effectiveness to be used in promotion and tenure decisions (James et al., 2015); as a source of information for students in the selection of instructors (Huxham et al., 2008) and courses; and as a source of data for research on teaching. Overall, participants who participated in this study displayed moderately positive attitudes towards the validity of SET data and their usefulness for improving teaching quality. Advocates of SET data claim that it provides feedback that creates a valuable opportunity for teachers to reflect on their current teaching, leading to improved teaching effectiveness. Hence, SET data are useful for improving teaching, encouraging teacher TPD, and reflecting on learning needs.

Ballantyne et al. (2000) noted that as SET data are used to enhance the quality of teaching, a great deal of attention needs to be paid to professional follow-up and accurate details on the application, process and validity of SET data. Some participants have questioned the validity of SET carried out in the institution. As one of the participants remarked:

> *There is no official guide of what counts as "good" in the evaluations or guidance on how people should use the information provided. I've figured it out on my own, but I talk to senior colleagues who are baffled by it. It is also hard to get a good response from students since moving online. I have gone from 80–90% responses to 10–25% responses, which is considered*

> *good amongst my peers. It raises the question of whether the data is valid, or even worth interpreting as you tend to get responses from opposites, those who love you or hate you.* (p38)

Some respondents actually believe that SET data do not help improve their professional skills and question using such data in making an informed decision about teaching practice. On the other hand, several respondents argue in favour of SET data as a source of improving professional teaching practice. However, if SET is used as part of a quality assurance programme, it is clear that the teacher's feelings need to be considered in the process. Further analysis may be required to conclude that teachers, as practitioners, use negative reviews as a positive means of change. Specifically, those new to the teaching field may demand extra help in coming up with input from SET data to benefit from their teaching practice assessment and development.

Additionally, in the institution where this research was carried out, SET data is routinely used for instructional development and personal decisions. However, it is not always accessible to teachers, limiting SET data for professional development. There is a set of principles and expectations that allows teachers to make sense of input from SET data. Although this is a small-scale analysis that may not be generalisable, it nevertheless sheds light on an element of SET data that has received little attention in the literature. To understand more about the constructions of SET data, teachers need to focus on the value of SET data, which emphasises how teachers can draw on the input of SET data to enhance their teaching profession (Arthur, 2009).

The ideology developed in this chapter indicates the need to shift teachers from a sense of weakness—that they cannot influence SET data—to a sense of strength, control and professional development capacity. Additionally, more effort can be directed to developing visualisation dashboards, which have the dual capacity to present SET data, and fuse SET data with multiple data sources to provide teachers with data-informed decisions.

CHAPTER 5

Teacher's Evaluation Dashboard (TED)

> *Consistent systems will make the user dare to explore the system and learn more; the same info should always be in the same position and look similar. If the user knows that the same command always has the same effect, he will be more confident when using the system.* (Nielsen, 1993, ch. 5)

5.1 Overview

At the University of Otago (UO), SET is intended to inform continuous improvement in teaching and learning, support quality assurance of courses (papers) taught, and provide an avenue for the students' voices to be heard concerning their educational experiences. The Otago InForm is a survey instrument used by UO to ask students to respond to questions on a 5-point Likert-scale (strongly disagree, disagree, neutral, agree and strongly agree), and open-ended questions with a free-text response for additional feedback. The survey described in the previous chapter indicated that a significant number of participants found SET valuable and wanted to visualise it using dashboards. This result resonates with the literature indicating analysis and reporting of SET data as one of the paramount concerns for both institutions and educators. Hence, this research discussed the design and development of a Teacher's Evaluation Dashboard (TED) to visualise SET scores (Number-ratings dashboard) and free-texts in SET data (Open-ended comments dashboard), to provide meaningful insight into SET data.

5.2 Introduction

TED is a visualisation tool designed to support a teacher's visualisation of SET data. This study designed TED to visualise SET data. In the context of this research, SET data is a combination of students' quantitative number ratings and qualitative opinions about their experiences on the value and effectiveness of teaching on the various courses taught.

The researcher argues that lesser effort leads to more motivation-effective system interventions. Therefore, ease of use encourages positive attitudes towards technology and promotes usage. TED was designed on top of Splunk® enterprise's "big data" infrastructure to allow for availability and interoperability. TED should be available and accessible to teachers at any time, as server shutdowns may discourage users from using the system and affect their tasks. TED should also respond to users when they request expanded exploration of the data in real time or almost real time.

The main purpose of TED is to provide insightful information that can support teachers in decision-making associated with learning and teaching activities (Ifenthaler et al., 2018; Mangaroska and Giannakos, 2017). Prior research indicates that TA and LA applications concentrate on offering insights to improve decision-making and assist learning and teaching interventions (Dyckhoff et al., 2012; Nguyen et al., 2017). In addition to collecting and analysing educational data to provide valuable insights, TA should consider the intervention of teaching to effectively support the process of learning and learning design (Ifenthaler et al., 2018; Xing et al., 2015). In conjunction with this point, we argue that TED is only useful by empowering teachers to be autonomous, and hence, assisting TED to become more effective in problem-solving, decision-making and decision implementation so that the system can continue to be increasingly effective in these activities and decrease the need for the intervenor (Argyris, 1970, p. 15).

The literature informed the conceptualisation of the design of TED, which focused on the analytics process discussed in the business understanding stage of the Data Science Life Cycle framework. TED embodies two main parts consisting of number and text. Descriptive analytics informs the qualitative aspect of TED (also known as "Number-ratings") to provide the teachers with useful metrics and reports. (see Appendix R and Appendix S). On the other hand, diagnostic analytics informs the qualitative part of TED (also known as "Open-ended Comments") and is informed by diagnostic analytics to provide teachers with detailed, in-depth information to particular aspects of students' comments. This form of analytics enables teachers to conduct early diagnostics interventions to detect anomalies in the data and provide additional support. See Appendix T, Appendix U, Appendix V, Appendix W, Appendix X, Appendix Y and Appendix Z for diverse ways to meaningfully present students perceptions about the taught course(s) to the teachers for decision making and appropriate action taking. The rest of this chapter describes the design principles (Shneiderman's Eight Golden Rules standards) applied to influence the design and development of TED.

5.3 Design Principles

Fig. 5.1 Shneiderman's (2004) Eight Golden Rules

TED applied the guidelines detailed by Shneiderman's Eight Golden Rules of Interface Design (see Figure 5.1); subsection 5.3.1 to 5.3.8 details the highlights of how insights from Shneiderman's rules have applied to some TED features such as filter functions, graphs and visuals to reflect the usability and usefulness of TED.

5.3.1 Strive for consistency

In computer science terms, consistency means achieving or striving for the programme interface's accuracy, how the interface appears, and the programme structure in detail.

According to Cha and Romli (2010, p. 6), "Consistency interfaces let users familiarise themselves with the system, thus helping them use the system well." Similar layout or content will assist the users to quickly familiarise themselves with the user interface. This rule's meaning is to achieve or strive for an object's consistency based on a rough description. For instance, the "Search" time range selector in the TED dashboard is consistent in all the pages' filter panels. This kind of consistency in TED design aims to prevent confusion when the teachers are using the tool.

5.3.2 Enable frequent users to use shortcuts

Shortcuts save time and prevent significant user errors in the system (Cha and Romli, 2010). The justification for using shortcut is to save time or optimise efficiency for experienced users. Hence, to speed up the operation output, expert users typically know to use shortcuts such as special keys, hot keys or default options. This feature is important in the system but not compulsory because novice users may not be familiar with the more complex functions available in the interface.

When choosing how many alternatives there should be for user interaction, you should make the easiest alternative visible so new users can use that and do not have to choose between alternatives. When users become more experienced, they can be offered more alternatives. This way, they can choose the most effective way when they are more confident with the system (Nielsen, 1993).

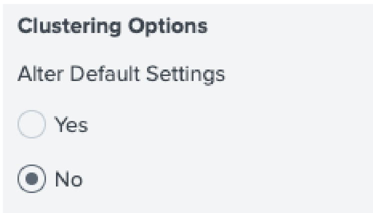


Fig. 5.2 Clustering Options automatically sets to default values and other functions hidden from novice user.

For instance, TED automatically sets some default value options. Figure 5.2 shows the "Clustering Options" filter, which is automatically assigned the default "No" option. The aim is to simplify the "Cluster" interface page so that teachers' do not feel lost or worried when using TED. However, it is possible to enable the more experienced teachers to have access to other parameters and functionalities.

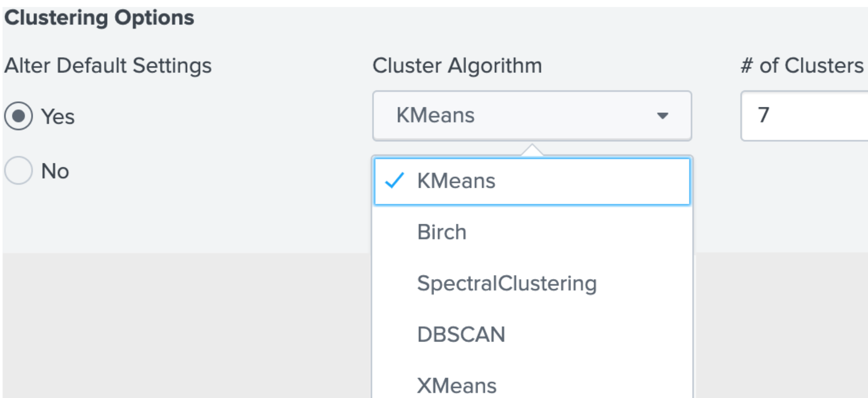


Fig. 5.3 Clustering Options changed from the default settings by an experienced user to reveal other hidden fields

A user can select the "Yes" option to generate other input fields automatically (see Figure 5.3). Accordingly, the "Cluster Algorithm" single select field and "# of Clusters" text field auto-generates to allow more experienced users to choose the clustering algorithm to use and the number of clusters to generate.

5.3.3 Offer Informative Feedback

Informative feedback reports that indicate a rise or drop, percentage increase or decrease, patterns and trends are essential indications that can inform users of various states (Cha and Romli, 2010). Feedback is instrumental since it would be a form of system notification to users. The system should prompt feedback to inform the user which state they are in for all user behaviour. Additionally, visibility of the feedback and its format is critical for users to quickly notice feedback.

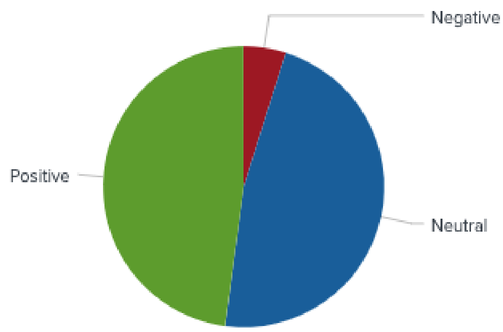


Fig. 5.4 Pie chart showing the portion of positive, neutral and negative sentiments.

The Pie Chart in Figure 5.4 illustrates one of the visualisations used in the "Sentiment" page of the TED dashboard used to represent the students' open-ended comments. The Pie Chart visualisation divides into three clickable parts with three primary colours, green, blue and red. The green piece represents positive sentiments, blue represents neutral sentiments, and the red represents negative sentiments.

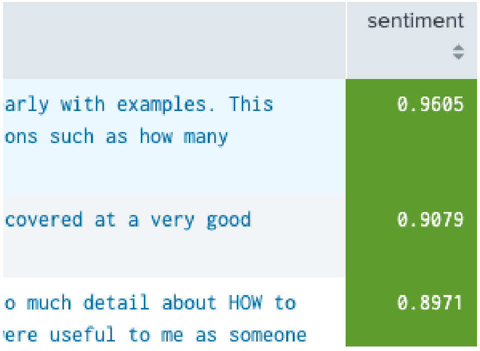


Fig. 5.5 Positive sentiments displayed when Positive portion of the pie chart is clicked

These clickable components also offer users informative feedback. For instance, a user can click on the green part of the "Pie Chart", also known as the positive sentiment to auto-generate a summary table of the positive comments below the pie chart visualisation (see Figure 5.5).

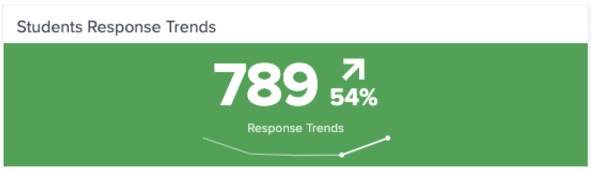


Fig. 5.6 Green "Students' Response Trends" Single Value Visualisation

The "Students' Response Trends" (see Figure 5.6) is a single Value Visualisation in the "Number-Ratings" dashboard in TED. This visual automatically turns green when the current students' response rate is higher than previously. The line indicates the pattern of students' response rate for the last five years. The bold number indicates the total number of students who responded to the survey. The up-arrow indicates a rise in response rate, and the percentage indicates the percentage increase.

Fig. 5.7 Red "Students' Response Trends" Single Value Visualisation

In contrast, Figure 5.7 automatically turns red when the current students' response rate is lower than the previous. The down-arrow indicates a drop in response rate, and the negative percentage indicates the percentage decrease.

Fig. 5.8 Radial Gauge Visualisation for the Overall Percentage Score

Figure 5.8 shows a "Radial Gauge" visualisation in the "Number-Ratings" page found in TED which uses the traffic light colours to represent students' perceptions about the teacher's overall performance ranging from 0 to 100. The green colour ranges from 60 to 100%, indicating good enough, orange ranges from 20 to 60% indicating fair, and red ranges between 0 and 20%, indicating unsatisfactory.

5.3.4 Design Dialogue to Yield Closure

Message dialogue to yield closure is an important element in designing interfaces. The Open-ended Comments visualisations in TED used simple labels and descriptive titles to make it easy for the teachers to understand and prevent them from guessing.

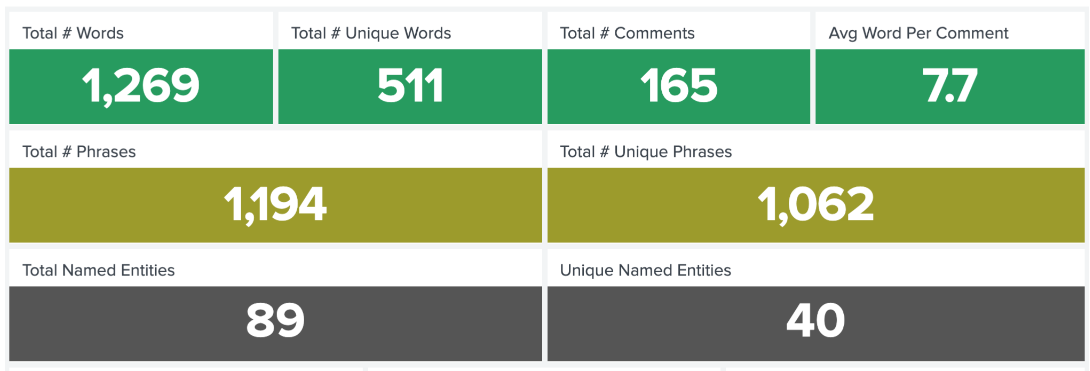


Fig. 5.9 The Set of Single Value Visualisations for the "Words and Phrases" Dashboard

For example, the "Words and Phrases" Open-ended Comment dashboard used the "Total # Words" rather than the initial "Total # Terms", "Total # Comments" rather than "Total # sentence", and "Total # Phrases" rather than "Total # ngrams" (see Figure 5.9).

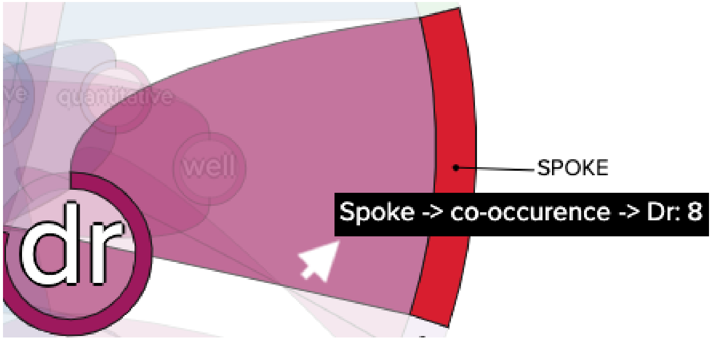


Fig. 5.10 Hovering mouse over a portion to display some information

TED also uses hover messages to provide useful information to guide the teachers properly on dashboard usage. For instance, the chord diagram in the "Named Entity" Open-ended Comment dashboard is used to visualise the inter-relationship between identified objects and actions. In this visualisation, the various objects are arranged radially around circles, while arcs connecting the data represents the relationships between objects and data points. A user can hover the mouse over a particular portion of the chord diagram visualisation to provide some descriptive information useful to the teachers (see Figure 5.10).

5.3.5 Permit Easy Reversal of Actions

According to Cha and Romli (2010), introducing an easily reversible feature relieves anxiety, since the users know that the system supports "undo" and "redo" to always return to the previous state. This notion that users can always undo actions encourages users to explore unfamiliar options. TED provides teachers with features that enable them to undo actions to revert to their original states. For instance, the "Multiple Select" and "Time Range Picker" filters in TED "Number-Ratings" dashboard help users drill down into the data and easily return to original states.

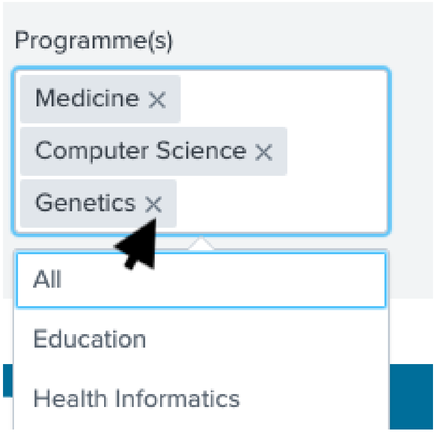


Fig. 5.11 "Student Ratings" dashboard "Programme" filter with selected "Genetics", "Medicine" and "Computer Science"

The "Programme(s)" filter in Figure 5.11 can be used to filter students based on their study programmes. The "×" icon on any selected programme option can be clicked, (say "Genetics") to remove it from the option of programmes to visualise, consequently causing the visualisation not to render the removed programme, and rendering the other programmes "Medicine" and "Computer Science". Continually removing programmes until all the options are removed will eventually cause the "All" option to load into the "Programme(s) Multiple Select" filter, which automatically triggers the visualisation to revert to its original state.

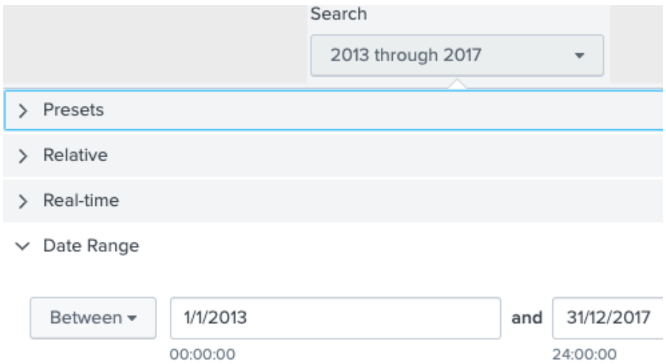


Fig. 5.12 The "Time Picker" filter for the "Student Ratings" dashboard

The "Time Range Picker" filter (see Figure 5.12) is used in the TED to perform a search based on time to visualise historical and present trends in SET data, presenting teachers with a summary of students' perception about courses over a period of time.

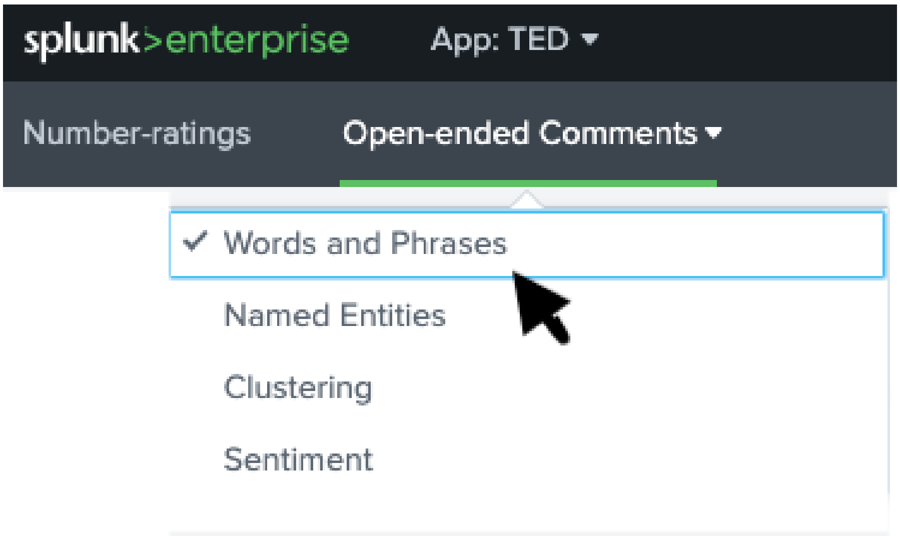


Fig. 5.13 The "Menu"

The TED menu (see Figure 5.13) provides teachers with shortcuts to navigate to particular parts of the dashboard without experiencing multiple procedures depending on the actions they wish to perform. Also, descriptive labels that can be easily understood by the teachers are used in the menu items, since some teachers who are inexperienced with TED will tend to explore unfamiliar things and different parts and functions in the dashboard itself.

5.3.6 Support Internal Locus of Control

Handing control to the users is vital to ensure that they can use the system to their satisfaction and so achieve their aims. Hence, giving users the feeling they are in charge and in control will maximise the system's usability.

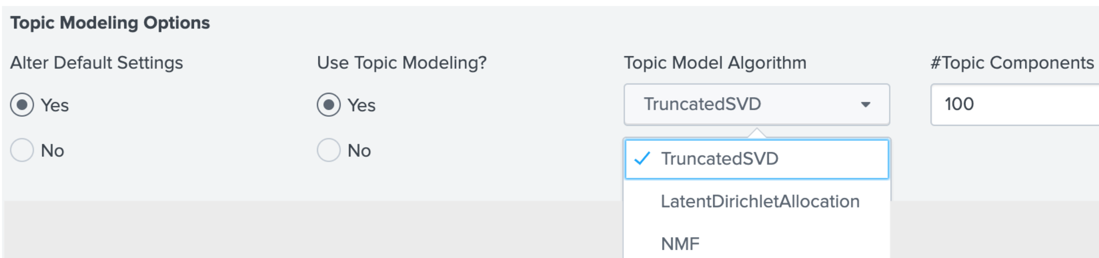


Fig. 5.14 The "Topic Modelling Options" in the Cluster Filter panel with "Alter Default Settings" Radio Button, "Use Topic Modelling" Radio Button, "Topic Model Algorithm" Drop Down and "#Topic Components" Text Field for the "Cluster" Dashboard.

TED hands control to the teachers once they are provided with valid and useful information to enable proper interventions and actions. For instance, Figure 5.14 illustrates how the "Topic Modelling" filter in the "Cluster" visualisation page gives teachers the control to choose which topic modelling algorithm to use, depending on what works for them and based on their data.

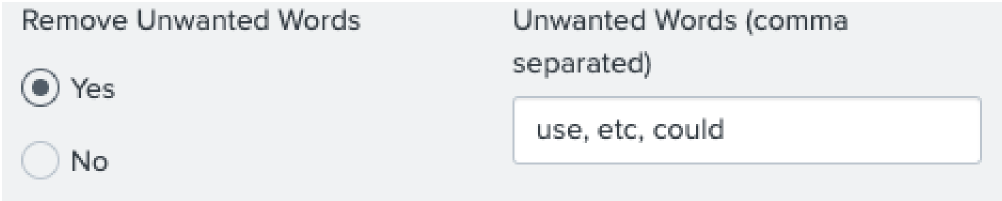


Fig. 5.15 Input Filter Panel with the "Remove Unwanted Words" Radio Button and the "Unwanted Words (comma separated)" Text Field for the "Cluster" Dashboard

Additionally, the "Unwanted Words" text field in the "Words and Phrases" visualisation dashboard filter (see Figure 5.15) gives teachers the power to enter the list of words they consider noise and wish them to be removed to further narrow down the focus.

5.3.7 Reduce Short-term Memory Load

According to Todd and Marois (2004), an average human can only maintain about five items in their short-term memory at one time. Hence, interface design should be as easy and straightforward as possible to minimise short-term memory load and facilitate positive user responses (Cha and Romli, 2010).

Fig. 5.16 Top "Phrase Cloud" Visualisation for "Words and Phrases" Dashboard

TED used simple visual representations such as Word Clouds, Sentiments and Bar Plots to streamline students' open-ended comments to make it easy for the teacher to understand data. For instance, the "Phrase Cloud" visualisation in the "Words and Phrases"

Open-ended Comment dashboard (see Figure 5.16) presents teachers with a summary of top phrases commonly used by the students in a word cloud visualisation.

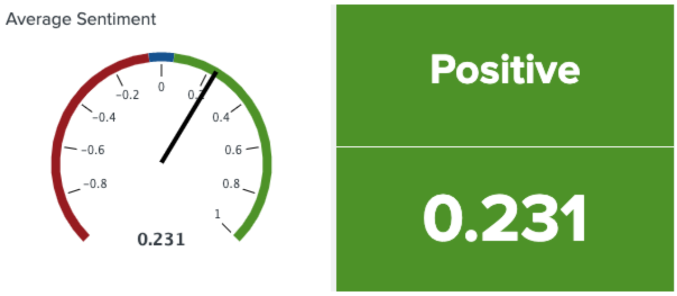


Fig. 5.17 "Radial Gauge" and "Single Value" Sentiment Visualisation

Additionally, Figure 5.17 illustrates the "Radial Gauge" used to quantify sentiments into simple number visualisations for easy understanding of the positive, negative, or neutral sentiments about a taught course.

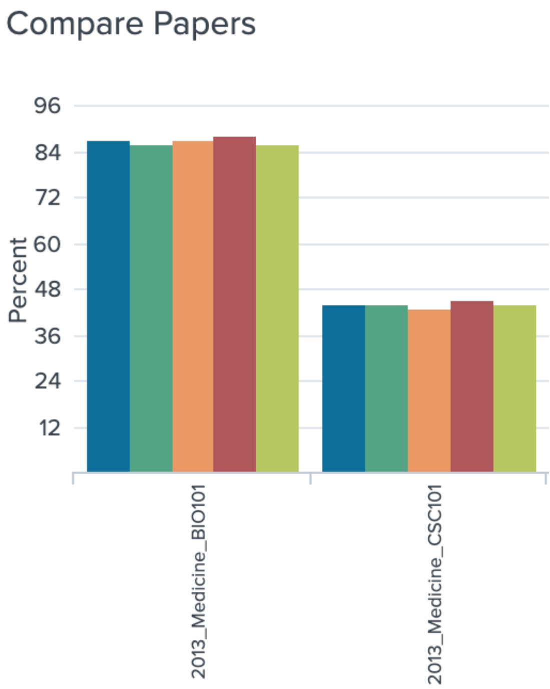


Fig. 5.18 The Papers Comparison Column Chart Visualisation

Furthermore, the "Bar Plot" is one of the visualisations used in the "Number-ratings" dashboard and compares the students' ratings on courses for the teachers to illustrate the trends and patterns over a period. Figure 5.18 illustrates a simple visual that presents Medicine students' perceptions in BIO101 compared with CSC101 in 2013.

5.3.8 Offer Simple Error Handling

Error prevention and error handling is one way to prevent a user from making mistakes. Further, systems should provide clear and informative instructions to enable users to recover from errors that may occur.

Fig. 5.19 Poor Error Handling Example

TED requires the users to input their username and password before accessing their dashboards. Error prevention occurs when the user inputs an invalid username or password, with a message displaying that the user has keyed in the wrong username and password. Additionally, a message box always shows when the search query experiences delay in processing the data or when the user keys in invalid data (see Figure 5.19).

5.4 Summary

This chapter discussed the framework that informed the design and development of a novel visualisation tool known as TED. TED embodies two main functions: visualisation of students' number-ratings and open-ended comments. The inspiration for the TED design was based on two analytical process models, Descriptive Analytics and Diagnostic Analytics. Descriptive Analytics informed the number-ratings dashboard, and Diagnostic Analytics informed the open-ended comments dashboard. This chapter also discussed the Shneiderman's Eight Golden Rules of design principles applied to TED's interface design to improve usability and usefulness of visualised SET data using dashboards.

CHAPTER 6

Usability Studies of Teachers Evaluation Dashboard (TED)

6.1 Overview

This chapter presents the data understanding stage of SET data. TED is a visualisation tool designed to represent SET data. A usability study was carried out to explore and test the validity of TED. This study followed an iterative research process that allowed for the development and improvement of TED design. The average System Usability Scale (SUS) score indicated an acceptable perceived usability of TED. As TED development moved from the first iteration to the second iteration, the SUS score increased from 65.50 to 81.3, indicating that feedback from the first iteration improved the design and development of TED. Lastly, the thematic analysis further indicated that teachers perceive TED as a usable and useful tool for improving their teaching quality.

6.2 Introduction

The collection of evidence from authentic classroom practice is essential for teachers' professional development (TPD) and education research (Roschelle et al., 2013). Also, teacher inquiry and reflection on educational data are considered essential TPD elements (Clarke and Erickson, 2003; Hansen and Wasson, 2016). However, because teachers frequently have several tasks to perform and maintain in their classes, their time for reflection and inquiry on educational data is limited (Emin-Martnez et al., 2014).

The value of a dashboard increases the rate of adoption for a data system. According to Nielsen (2012), usefulness is a result of utility and usability. Perceived usability is the extent to which an individual trusts that a specific system would be easy to use, independent of effort. In other words, it can also be perceived as the extent to which an individual trusts that the features or use of a specific system would improve performance.

Usefulness is the extent to which a product can be specified by users to achieve specified goals with efficiency and satisfaction in a specified context of use (Bevan, 2001, p. 536). Usefulness refers to whether a product provides the features needed independent of how the system is implemented, while, usability refers to the ease of use

(Nielsen, 2012).

$$Value = Usability + Usefulness \tag{6.1}$$

Equation 6.1 implies that users will accept that a system is valuable based on how easy it is to use, and the features it provides to users to carry out their profession (Nielsen, 2012)). Hence, individuals are likely to find an application valuable if they believe that the system is easy to use and will assist them in doing their work.

6.3 Iterative Design

The concept of iterative designs can apply to educational technology to explore and implement real-life concepts while serving a dual function of researching the current opportunities available to promote educator reflection, and iteratively design practical prototypes for teachers (Emin-Martnez et al., 2014). This study follows an iterative methodological framework to enhance the design of TED through the first and second iterations of our design, as described by Barab (2014) (see Figure 6.1). The first and second iterations followed in this research explains the design principles inspired by the Learning Awareness Tools - User eXperience (LATUX) approach for technology development in teaching and learning analytics (Martinez-Maldonado et al., 2015; Saar et al., 2018), which promotes the standards of prototype reliability and validity in testing conditions.

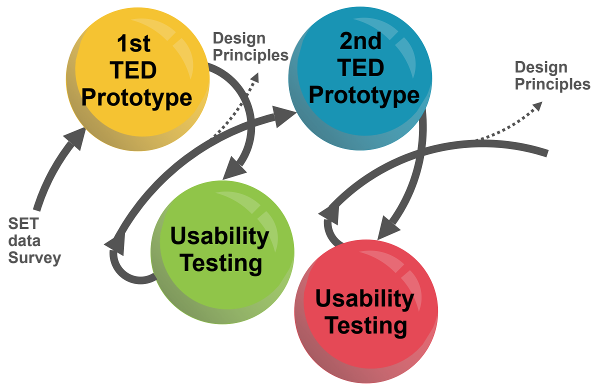


Fig. 6.1 The Iterative Design Process

A usability study was conducted to ascertain the perceived usability and usefulness of TED. Short interviews were conducted with 23 participants who performed short tasks on TED to determine its perceived usefulness, difficulties, and concerns. The usability study followed an iterative process. Twelve participants (n = 12) participated in the first iteration and eleven participants (n = 11) in the second iteration. Feedback from the usability study conducted in the first iteration was used to improve the TED design. Subsequently, another usability study was conducted in the second iteration to verify that the design of TED was improved.

6.3.1 The Perceived Usability of the First TED prototype (First Iteration)

Participant teachers found the TED Number-ratings (Aggregate) dashboard to be usable. However, some participants complained about the "traffic light" colours on the **Overall Students Rating Score** radial gauge visualisation used on this dashboard. One of the participants, p1, suggested changing the "traffic light" colours to a single colour. Another participant, p4, complained that the red colour used in the **Student Responses** single value visualisation on this dashboard conflicted with one of the radial gauge colours and advised that the colour on the single value visualisation be changed to avoid misrepresentation. One participant advised having some form of metric to represent the true reflection of the proportion of the class response rate on this dashboard, as follows:

> *... it will be helpful to have some metric to know what the response rates are. For instance, a 25% class response rate may not represent the complete information, considering some factors such as the representative sample of the class, the selected sample that likes and hates the course, and the method of selection; a 25% response rate that is randomly selected may represent a true reflection of the class more than a non-random selection of 25% response rate. However, I still think it will be good to know the proportion of the class that responded.* (p21)

One other participant complained about the **Search Time Range Picker**:

> *Search time picker may be a little tricky to know which academic year to select. For instance, using the time picker to select 2013/2014 academic year, how do I know which of the year options to choose from; either 2013 or 2014.* (p4)

Participants found the TED Number-ratings (Comparison) dashboard usable; however, more than half of the participants said that the line chart representation used for the **Programmes Comparison** visualisation on this dashboard was not a sensible way to represent multiple programmes. For example, one of the participants, p2, said that "line charts is not a sensible way of illustrating this kind of data, the bar charts would be a preferable way to represent this kind of information." Additionally, many participants preferred the **Average Percent Score per Question** bar chart visualisation to be transferred from this dashboard to the Number-ratings (Aggregate) dashboard. Two participants also requested that one of the **View Type** dropdown input options be renamed from **Relational** to **Comparison**. One participant requested an additional

Papers Comparison visualisation that compared papers and the **Programmes Comparison** visualisation. This participant taught the same set of students two papers and wanted to visualise how they would rate teaching performance in one paper compared with the other.

> *Can you compare Papers in addition to comparing programmes? I think it would also be useful to compare the responses from the same student programme for two different courses. For example, I might want to compare how Ecology students evaluated my teaching in PAPER 101 and PAPER 202.* (p17)

Participants found the TED Open-ended Comments (Words and Phrases) dashboard usable. However, many participants suggested simplifying the terminologies used for some labels and titles. For example, one participant, p8, said, "some names have to be reworded; n-grams can be renamed as Words to Phrases, stop words renamed to unwanted words." A few others said that the **Word Count**, **Unique Word Count**, **Phrases**, **Unique Phrases** single value visualisations would not make sense to them in their teaching. However, one of them admitted that the **Avg Words per Comment** single value visualisation could be useful for the teachers to have an idea of the average number of words students made per comment.

Some participants expressed some level of satisfaction on the TED Open-ended Comments Named Entity dashboard. Others had some usability concerns regarding information interpretation and information overload (see Appendix AM for detailed information on this dashboard).

Several participants advised hiding some of the numerous input fields present on the TED Open-ended Comments Cluster dashboard to simplify the dashboard. One other respondent, p5, recommended that the **# of Clusters** text input field found in the **Clustering Options** section of this dashboard can be modified to automatically determine the maximum number of clusters based on the size of the data rather than the default seven clusters.

More than half of the participants were impressed with the TED Open-ended Comments Sentiment dashboard. One participant said that the **Average Sentiment** visualisation simplified and summarised the students' comments by quantifying them. However, some participants recommended making the pie chart visualisation on this dashboard clickable.

6.3.2 Actions Taken on the First Prototype of TED (First Iteration)

For most people, red means "stop" or "bad" and green represents "good" or "go", and this was not the case for the **Student Responses** single value visualisation, unlike the colour distinctions in the **Overall Students Rating Score** radial gauge visualisation.

Hence, the recommendation to change the **Student Responses** single value visualisation colour from red to blue was adopted to avoid misrepresentation. Additionally, the proposal to have a metric to reflect the class proportion participating in a survey accounted for the **Student Response Rate** single value visualisation.

The participant who had issues with the **Search Time Range Picker** disclosed being a newly employed staff member who had transferred from another institution and admitted that the academic year calendar used in the former institution ran differently. Hence, the participant needed more time to get familiarised with the way the academic year calendar of their current institution runs to comfortably use the time picker.

The issues raised concerning the Number-ratings Comparison dashboard were carefully considered and implemented by the researcher, including changing the **Programmes Comparison** line chart to a bar chart, relocating the **Average Percent Score per Question** bar chart to the Number-ratings (Aggregate) dashboard, and renaming the **View Type** drop-down options from **Relational** to **Comparison**. For a detailed description of the themes generated for this dashboard, see Appendix AC.

On the concerns raised about the TED Open-ended Comments Words and Phrases dashboard, most of the suggestions raised on this dashboard were addressed. However, the **Word Count**, **Unique Word Count**, **Phrases**, **Unique Phrases** single value visualisations were still left untouched as it may require more time to learn how to change the code that generated them. The researcher hopes to handle this in future iterations.

Following the recommendations that were given about the Open-ended Comments Cluster dashboard, most of the fields were made to be hidden by default when this dashboard loads. More experienced users can decide to open more options for additional functionalities. A new algorithm was also introduced to determine the number of clusters based on the amount of data. This algorithm can be selected from the **Cluster Algorithm** drop-down list, which provides several clustering algorithms and allows experienced users to select and use a specific clustering algorithm.

The researcher implemented the recommendation made on the Open-ended Comments Sentiment dashboard. Partitions of the **Average Sentiment** pie chart visualisation was made clickable to drill down into the various sentiment categories of comments (positive, negative and neutral), depending on the partition clicked.

6.3.3 The Perceived Usability of the Second TED prototype (Second Iteration):

Without a shadow of a doubt, colours are one of the most important components of dashboards. Participants were happy with the colours used on the Number-ratings (Aggregate) dashboard. One participant, p9, supported the "traffic light" colours used in the **Overall Students Rating Score** radial gauge visualisation on this dashboard: "I like the traffic light colours of the radial gauge and the interactivity going on there." A few other participants recommended including confidence intervals in this dashboard. As one of the participants said:

> *... incorporate confidence intervals. If confidence intervals are wide and there are differences between questions, that may not mean anything. However, if the confidence interval is narrow and there are differences between questions, that might be an indication that the lecturer needs to do something. One thing that is of concern is a tool providing bad data, points users effort in the wrong direction may be making things worse. Consequently, costing users to invest time and resources that could have been spent on other things.* (p8)

Many participants expressed satisfaction with the features provided on the Open-ended Comments (Words and Phrases) dashboard. One of the participants said:

> *... dashboard helps isolate what the students are saying and create opportunities for the teacher to easily collect the most common words or phrases used by the students, and to know how many times they were used.* (p10)

In addition to the earlier complaints made about the Open-ended Comments (Named Entity) dashboard, participants further recounted that this dashboard information is overloaded and hard to interpret. As one of the respondents said:

> *The information is dense. This dashboard would require a little bit more time to play around with the content. So we would need to, I suppose, inform ourselves of the types of keywords that we would be looking at ourselves.* (p20)

One other participant asked if the **Chord** visualisation on this dashboard can be more interactive, as follows:

> *Can the Chord visualisation that presented the relationships between the Named Entities and their co-occurrences be clickable to get the words with the habits for the individual response?* (p7)

Several participants expressed some degree of satisfaction for the Open-ended Comments (Cluster) dashboard. However, one participant suggested that each of the bars in the *Column Chart* cluster visualisation in this dashboard be clickable to extract comments about the words' contexts in that cluster. For a detailed description of the themes generated for this dashboard, see Appendix AD.

Participants were impressed with the pie chart visualisation on the Open-ended Comments (Sentiment) dashboard, especially the clickable feature. However, one participant argued that the bar chart would have been a better visual representation:

> *Generally speaking, people interpret pie charts terribly. So people have a lot of difficulty with this sort of proportional thinking when they are faced with a pie chart, you would probably be better with a little bar chart here because then people can see the relative difference more easily for the percentages in the pie chart, bar charts may be preferable.* (p20)

One participant also suggested that having a time range picker will be helpful for monitoring trends in this dashboard:

> *In future, implementing the time range feature will be useful for monitoring improvement in certain parts of teaching over time. For example, knowing that you speak too fast in lectures and looking for getting feedback speaking fast over time.* (p20)

Another participant advised that this dashboard's visualisations could be problematic if it does not consider the comments' weights, explaining that a few negative comments could be weightier than many positive comments. For a detailed description of the themes generated for this dashboard, see Appendix AE.

6.3.4 Decisions and Future Actions on the Second prototype of TED (Second Iteration):

The feedback on colours used on the **Overall Students Rating Score** radial gauge visualisation confirmed that the choice of colours used on the Number-ratings (Aggregate) Dashboard was meaningful. The confidence intervals visualisation feature was not implemented on this dashboard due to time constraints; however, it may be implemented in future. For a detailed description of the themes generated for the TED Number-ratings (Aggregate) Dashboard, see Appendix AF.

The researcher may consider deleting the Open-ended Comments (Named Entity) dashboard from TED in future design if nothing can be done about introducing clickable visualisation, so reducing the information overload and making the dashboard easier to interpret.

Feedback about making the bars in *Column Chart* visualisation for the Open-ended Comments (Cluster) dashboard clickable was valid; however, due to time constraints it was not implemented, the researcher hopes to implement in future design.

The suggestion about implementing a time range picker in the Open-ended Comments (Sentiment) dashboard was also valid and will be considered in future iterations. The real free-text data used to test this dashboard did not come with timestamps. However, including timestamps in the data in future iterations will enable this feature to be easily implemented.

6.4 System Usability Scale

The System Usability Scale (SUS) questionnaire is a validated instrument used to evaluate the usability of an application and provides "quick and dirty" and reliable measuring tools (Brooke, 1996). John Brooke developed the System Usability Scale in 1986 as a method to assess and evaluate the usability of products, systems, or services. SUS is in the form of a questionnaire consisting of 10 questions with answer options of 1 (strongly disagree) to 5 (strongly agree).

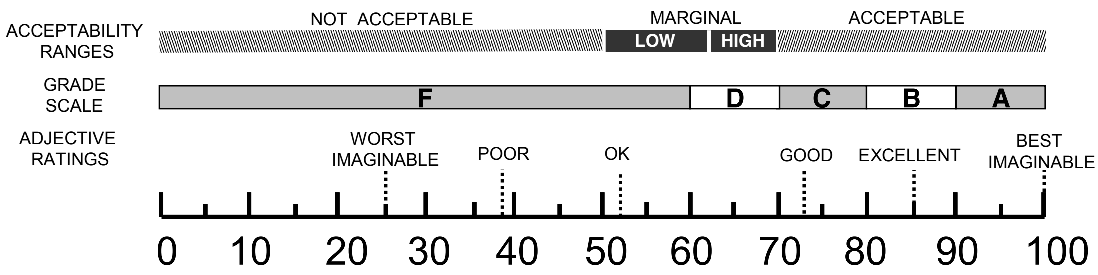


Fig. 6.2 The SUS Score (Bangor et al., 2009)

The method for calculating SUS values is odd items minus 1, and 5 minus even-numbered items, then the results of odd items and even-numbered items are added and multiplied by 2.5. If the SUS score is above 68, then it is considered above average, and if it is below 68, then it is considered below average (Brooke, 1996). However, Bangor et al. (2009), in their article, further explained the SUS grading scales and score ranges (see Figure 6.2). The SUS score is not a percentage, even though the value obtained is in the range of 0–100. Recent research states that SUS can be divided into two sub-scales, namely, usability (items 1, 2, 3, 5, 6, 7, 8, 9) and learning abilities (items 4 and 10) (Lewis and Sauro, 2009).

SUS was chosen and used to answer questions about the usability of TED because the questions are quick and easy, and would encourage participants to answer within a short period. The questionnaire consisted of only ten statements, and each survey result was a single score (0–100) (Bangor et al., 2009). Apart from SUS producing reliable results, it can also be used on small sample sizes to reduce the cost of research and testing (Dumas and Loring, 2008; Jordan et al., 1996; Pradini et al., 2019). A total of 20 participants responded to the SUS questionnaire: eleven respondents (n = 11) participated in the first iteration, while nine respondents (n = 9) participated in the second iteration. This action was to verify if the usability of TED improved as it updated from the first version (first TED prototype) to the second version (second TED prototype).

The average SUS score for for the first TED iteration gave a value of 62.50 (OK/Fair). In contrast, the average SUS score for the modified version of TED used for the second iteration gave a value of 81.39 (Good). Hence, this increment in the SUS scores indicates that as the feedback from the first iteration was used to improve TED during the developmental process, the usability of TED improved even further (see Appendix AG).

6.5 The Perceived Usefulness of TED

This section discussed the perceived usefulness of TED to determine how useful this dashboards could be towards improving teaching quality.

6.5.1 The Perceived Usefulness of the Number-ratings (Aggregate and Comparison) Dashboard

Actionable information should provide reports about teaching and learning with flexible granularity in reporting (Gašević et al., 2016; Schumacher and Ifenthaler, 2018; Wolff et al., 2013). The flexibility of the questions on the TED Number-ratings (Aggregate and Comparison) dashboard was a useful feature used to drill down into granularity levels in reporting. For instance, one of the participants said:

> *The flexibility of the questions on this dashboard is a good idea, rather than restricting the questions to the five standard questions used in the institution. This question flexibility helps the teacher's dashboard address both compliance and learning. For example, a teacher might want to investigate a question about critical thinking; and that may be the focus of his teaching. The questions filter indicates to the teacher what specific questions are saying and where to improve. For example, a teacher may get 80% in Q1, 45% in Q2, 85% in Q3, 90% in Q4, and 70% in Q5, although the overall score was 74%, this kind of representation points that Q2 is where the teacher needs to improve.* (p3)

Accordingly, the purpose of TED as a new class of teacher assistance tool was described by our design principles to provide reports of actionable information about teachers and their learning with flexible granularity in reporting.

Participants acknowledged that the presentation of SET on the Number-ratings dashboard is more comprehensive than the way SET results are currently presented, mostly in PDF formats and sent to them via email. Some of them said that it was cumbersome and time-consuming to search out the PDF SET results for the previous years, and almost impossible to visualise trends and compare patterns. As one of the participants explained:

> *I think this dashboard provides a much more useful summary than the summary that I currently receive. The summary I received gave those statistics;*

> *however, it is quite a complicated document and difficult to compare across the years. One has to go to different resources to find the previous statistics to compare across the years.* (p20)

Participants reported that the filters were innovative. For example, participant p6 said, "Programmes filter is really handy; to be able to sort students by their programmes is particularly helpful for me because I teach a first-year course, which is quite large, and students are coming from a wide variety of programmes."

However, a few other participants expressed that the programme's filters could create room for manipulating the system. For example, one of the participants expressed concern:

> *The Programmes filter is very innovative. However, the programmes filter could have some implications, such as isolating and excluding a particular group of students who always rate instructor low, and making a case to the University to justify why that is the case. On the other hand, it may also assist the instructor in cherry-picking only the groups with a high rating and present it to the University for promotion.* (p3)

Data privacy is a key challenge for using analytics at higher education institutions (Daniel, 2015). Concerns have been raised about recording student activities on the system and profiling student learning. Jones and Gregor (2007) revealed that institutional executive offices are likely to be concerned about privacy and security concerns when the system is up and running. Two participants raised some identity and privacy concerns that may require some attention in a future design. As one of the participants explained:

> *... for privacy concerns, the programmes filter should have strict rules around not drilling down to groups that are too small. Hence, it will be good to create a rule to turn off the programmes filter features when the groups are too small to prevent easy identification of individual students in smaller groups.* (p18)

For a detailed description of the themes generated for this dashboard, see Appendix AH.

6.5.2 The Perceived Usefulness of the TED Open-ended Comments (Words and Phrases) Dashboard

Evaluative comments are useful for teaching evaluations to promote or improve teaching performance (Danielson and McGreal, 2000). One of the participants, p4, said, "I may be interested in the verbs and adjectives from the Parts-of-Speech visualisation because those are words that might be related to evaluations." However, another participant argued differently, saying that the content provided by these dashboard visualisations was not enough evidence for both qualities of teaching and promotions.

> *It will be more useful for quality of teaching if the dashboard could count only the top evaluative words like good, not good, different, useful, in order to track the evaluative comments. I do not find it useful to make a case for promotions; most of the information has to do with the class's content rather than promotion.* (p1)

According to Brown (2020), instructors in larger courses rely more on the dashboards than the smaller courses' instructors. Hence, dashboards are useful for large classes that handle many students. As one of the participants explained:

> *If you were trying to analyse a very large class over a period, I guess, the word clouds will be helpful to see how things changed. And especially if you were looking at the content of what you were doing, then you could see whether they are being reflected. For example, if the XY framework is really important in what you are doing and it shows up in your dashboard, then that is really helpful.* (p21)

For a detailed description of the themes generated for this dashboard, see Appendix AI.

6.5.3 The Perceived Usefulness of the Open-ended Comments (Named Entity) Dashboard

Most participants did not find this dashboard useful and applicable to help teacher performance evaluation and improve teaching. As one of the participants noted:

> *... would take a while to think about the nature of the comments that would be intuitive to understand own specific subject area, as to the types of things that are useful to my teaching. We were interested in this particular type of data because we want to be making changes to our teaching, either to how we are performing, ourselves or to the material that we are presenting to the students.* (p20)

For a detailed description of the themes generated for this dashboard, see Appendix AJ.

6.5.4 The Perceived Usefulness of the TED Open-ended Comments (Cluster) Dashboard

The online discussion forum's common conception is that it is a virtual learning environment where both the students and teachers learn as much from one another as from course materials or lectures (Thomas, 2002). It has been shown that online discussion forums increase participation and collaborative thinking by providing asynchronous, non-hierarchical and reciprocal communication environments (Dehler and Porras-Hernandez, 1998). Furthermore, online discussions provide a perfect forum for an academic discourse that promotes increased student engagement, critical analysis and reflection, and social construction knowledge (Dehler and Porras-Hernandez, 1998). Participants considered this dashboard to be a handy tool for analysing text from online discussions forums, as one of the participants described:

> *... dashboard would be the sort of thing that would be useful to have on something like a discussion forum. So in my teaching, we are starting to look at using discussion forums with remote teaching. Using cluster visualisation would be quite a good way to work out common questions amongst students; if a teacher has got much feedback and has no staff to assist in dealing with reading through the comments and trying to work out what are the common themes to come up with answers that can address the majority of people's question. However, one of the difficulties is that the students often have difficulty structuring their questions, so they don't know what they don't know. Having this sort of interface; to have keywords to cluster decision options students have for certain types of problems will be useful. It would enable the teachers [to] work out which students have managed to understand the words that we regularly use during the course and those who do not understand how to structure their questions. Therefore, separating those top students who know what they are doing from those struggling with structuring. This separation will enable the teacher to know how to address their questions, such as answering top students differently, compared to those struggling with structuring the question itself and will need a lot more description of any solution [the] teacher can offer.* (p20)

For a detailed description of the themes generated for this dashboard, see Appendix AK.

6.5.5 The Perceived Usefulness of the TED Open-ended Comments (Sentiment) Dashboard

The primary focus of the sentiment analysis is to determine the writer's feeling (attitude, emotion, or opinion) from a given text (Balahur et al., 2014; Rajput et al., 2016). This dashboard automatically performs analyses on textual feedback and generates quantitative and qualitative metrics that can help the teacher highlight significant areas of appreciation and concerns. More than half of the participants found this dashboard powerful and useful. For example, participant p15 recounted, "This dashboard is really valuable. All of the quantitative dashboards you have are valuable, but I think this one appears to be the most useful." Another participant said:

> *This is the most useful. This dashboard could help the teacher know his strengths and weaknesses. What would be most valuable for me is identifying the problems; a teacher always wants to see what aspects to improve. This dashboard will be most valuable to identify what kinds of problems exist. It will also be a good idea to have a dashboard that combines the clustering Algorithm with the Sentiment Algorithm to tell me what areas I need to improve in groups/clusters). For example, it should take the negative comments and cluster them into themes or take the positives and cluster them into themes of good or jobs well done. This dashboard can further issue advice to the teacher and highlight areas of good performance, and areas of improvement; to tell the teacher the group of things they might need to improve upon from the negative comments as well as the group of strengths from the positive comments; to show what themes have emerged, to know what needs to be improved, such as the structure of the course or if it is the articulations that need to be improved. As a lecturer, I look at how I can deliver value-added services in the form of things that my students say I need to improve upon.* (p5)

Participants applauded this dashboard and supported the idea that it could improve teaching quality; they said it could refocus teachers' attention. As one of the participants stated:

> *I think the nice thing about this dashboard is, when reading through the comments, teachers can be discouraged by the first few comments they see. Moreover, I remember one year the first comment was negative, and it just made me feel bad immediately. We know that these things cannot be taken too personally, of course, not with a large class. However, regardless, the first comment was negative. Whereas if I can filter, I can look at what went well first, giving me much more resilience to take on board the things that did not go so well and to be more constructive about responding to those things.* (p20)

Trust in data processors can, for instance, alleviate concerns with opaque personal data-processing (Mazoué, 1990; Shackelford and Raymond, 2014). Some participants expressed some trust concerns about the algorithm that predicted the sentiments. For instance, participant p4 reiterated, "It will be interesting to drill down into the positive and negative comments; it will enable me to judge for myself whether the system is interpreting the comments correctly or not."

Some participants argued for and against using the sentiment dashboard for promotions. The expressions of one of the participants, p13, advocating using the sentiment dashboard for promotion is as follows: "I can sort of say that this would be quite useful, particularly to serve as evidence of your teaching for promotion purposes." Another participant supported using this sentiment dashboard for promotion purposes, but raised concerns that the dashboard cannot handle negatives that are not the teacher's fault:

> *...in situations where there are too many negatives, how would this dashboard distinguish between negatives that are not the teacher's fault and those that are? It does not affect promotions, for example, too many negatives that were not the teacher's fault.* (p14)

However, another participant who was against using it claimed that technology advancement had not reached that level yet:

> *I will be wary of using the sentiment index for promotion. For instance, teacher 'A' has a sentiment index of 0.7; [they] should get a promotion and teacher 'B' has a sentiment index of -0.3; [they] should not get promoted. I think technology probably is not quite at the level [where] we should use sentiment for promotion purposes.* (p9)

For a detailed description of the themes generated for this dashboard, see Appendix AL.

6.6 Discussion

One obvious thing is that educators have busy schedules, and time is never enough (Masood et al., 2018). Time is needed to plan for classes, complete different tasks required by several education roles, and spend time with students. A systematic review of educational data visualisations conducted by Vieira et al. (2018) noted that far too little attention was paid to delivering information to users in classroom settings. For instance, handing a collection of students' open-ended comments may be overwhelming. For teachers to be at their best, they need to take advantage of tools like dashboards that can be very efficient for teachers handling large classes to advance their teaching professional development (Molenaar and Knoop-van Campen, 2017).

> *... I do not often have much time to go through all the comments from my evaluations. Occasionally, I scan through and highlight a couple of positive and negative comments to attend to later to make some changes to my course. However, I have a massive class of about 500 to 800 students, and time is never on my side. Dashboards like this would be absolutely magic for identifying particular problems that I know come up on the course, such*

as finding keywords very easily in the comments, seeing how people respond to particular problems already identified, or particular areas that may have changed. So I might specifically be asking for comments in that area. (p20)

Despite the practicality and everyday use of quantitative data from surveys, student text responses on open-ended questions help instructors with actual input (Koufakou et al., 2016). These surveys' written comments are associated with many drawbacks and challenges, such as misspellings, jokes, or short or irrelevant statements. Low return rates are reported in the literature, as only 10% to 60% of students answer these questions (Jordan, 2011).

Improving teaching with the std five questions by getting good scores is not helping; it is the text that gives one something about the lecturer's presentation. Even during peak lectures, some students still find it boring or think it was a poor lecture. However, some other students may find it fantastic, really good. The difference is how they see it, and there is not a lot the teacher can do about that. Nevertheless, if many students said that the lecture was disorganised and did not understand what the teacher was talking about, that would be helpful. (p21)

However, without the constraints of the carefully worded numerical rating questions, it is the open-ended nature of the questions that enables students to relay what is on their mind and what they believe is significant (Gottipati et al., 2018b). Some studies indicate that instructors prefer written comments in SET (as opposed to feedback based on statistical data) (Cunningham-Nelson et al., 2020; Nitin et al., 2015; Santhanam et al., 2018). To find places for change, a faculty teaching a course should read these comments and analyse how students view their teaching.

... it would allow me to prompt the students to use certain words when making comments, and that way, the system would find those comments very easily, such that I would then be able to search. For instance, a teacher who wants the students to provide feedback on how you felt about the changes to "comparison mean" could prompt them to use the word "comparison mean" in their comments. Making it easy for the teacher to use the word, "comparison mean" as the search term to find all the comments associated with the phrase "comparison mean", would make it very useful. Particularly, if I was to add that sort of collaborative notes to inform the students about how I make the best use of their comments. (p20)

The relationship between students expected grades and SET has been controversial (Isely and Singh, 2005). For instance, there is a suspicion that instructors of a particular course that students perceive as easy may result in higher grade expectations may also indicate a more favourable average SET result.

> *It will be good to compare the students' number ratings against students pass/fail rates. For instance, A students compared to B students, compared to C students and those that failed. (It will be useful to filter out to see how those two are associated). Because it is perceived that teachers that teach a course too hard might get low ratings compared to those that teach a course too easy.* (p9)

Dashboards can facilitate the linking of educational data (data fusion) to make it more effective in enhancing teacher professional development. Thomas (2018) examined the connection between student ratings on teaching and student physiological data. Similarly, Schmidlin et al. (2015) established how to analyse and cross-reference data without decrypting the data sources. Hence, there is a need to explore the connection of SET data with other forms of data in future research, such as attendance, grades and student engagement data.

> *From my perspective, if this dashboard could provide a means to anonymously match examples of people who perform very well and examples of people that perform poorly. This kind of information would be helpful for teachers to compare anonymous performances and trends. For instance, to examine what the third-year papers look like compared to second-year papers to see if they are the same patterns in all of them or if there are differences, as well as how the students responded.* (p21)

Teacher dashboards are emerging as a key way TA and LA might positively influence educational practice (Buzhardt and Heitzman-Powell, 2005). Despite the often reported benefits of educational technology, educators often find it challenging to integrate these applications and devices into typical school practices (Doering et al., 2003; Genet, 2013; Lu, 2018). Additionally, Fullan (2007), noted that embracing change is highly complex and challenging, and educators see the use of technology as a barrier to teaching professional development. Hence, many teachers are already in their comfort zone and may be neglectful in increasing their knowledge of technology or may have difficulty adapting to technology (Mardiana, 2020, 2018).

> *I am experienced, and I have been evaluating my teaching for a long time now using what is already existing [...] So I am set in my ways, and I do not need a dashboard to do that.* (p3)

Elsaadani (2013) showed a significant relationship between teachers' age and positive attitudinal difference towards technology as a function of age. Furthermore, Purcell et al. (2013) argued that older lecturers might have difficulty using technology and keeping up with modern teaching approaches, practical acquisition and the use of new teaching technologies.

> *This dashboard necessitates a certain degree of computer competence, for someone of my generation, we are far more used to a paper project. I think this is a future-oriented thing, and probably, it will become more valuable as the years go by, and people become more and more attuned to it, in that respect, I think it is valuable. Older lecturers may struggle to get a hold of it.* (p7)

Usability has been found to rely mostly on users' abilities to explore their data and customise the display in a way that closely corresponds to the learning design (Gruzd and Conroy, 2020). Hence, using dashboards to access personal teaching data is preferable and more effortless than visualising unknown datasets. Teachers should be enabled to conduct their inquiries by formulating and testing their explicit models of how their practice works and how it could be enhanced (Griffiths, 2017).

> *... might not fit the kind of feedback I have looked at in the past, and how I have learned from that feedback. I can see the potential, but I guess I would love to be able to play with my own data, to be able to say how useful [it is].* (p13)

According to Marsh and Roche (1993), one of the motives of teaching evaluation is to assess the teacher's summative measure of teaching effectiveness to be used in promotion and tenure decisions (James et al., 2015). Teachers need justification for using SET in staff appraisals and providing tools that can also provide effective visualisations that can serve as some form of evidence for promotions, building portfolios, internal quality assurance and performance management (Blackmore, 2009; Seldin et al., 2010).

> *... people who are on perhaps confirmation pathway who are needing particular things, such as promotions, would find this dashboard very useful for them because they need to be able to break down their feedback far more than perhaps somebody who is just looking at it for refining their teaching or feedback on content.* (p20)

There is a need for levels of recursion in the system which present themselves as a "black box" that allows faculty staff (teachers, researchers and educational managers) to effectively operate while remaining essentially isolated or anonymous (Tatto et al., 2006). Articles reported a dysfunctional impact of the dashboards related to concerns of surveillance and anxiety about what data were being captured, who had access, and how they were interpreted (Crooks, 2017; Faiola et al., 2015; Roberts et al., 2017b;

Yigitbasioglu and Velcu, 2012). When dashboards become portals to the information system, there are questions of access and agency, who can see the data at all levels? Is access managed by role? (Kerzner, 2017). Xiong and Li (2007) proposed a role-based access control system to control user access and permissions. Authoring a personalised dashboard presentation specific for each user's domain of responsibility, privileges and data restriction could eliminate issues of privacy, trust or possible system manipulation (Malik, 2005). Hence deans should not have access to the head of department's dashboard views, and the head of the department should not have access to the teacher's dashboard views.

> *There are two uses of course evaluations, and we only see one of them applied effectively in the institution, which is using it to keep the staff in check. Course evaluations are like the grade-books for teachers, and it is common to think about course evaluations as something that will positively contribute to teaching. I think it makes perfect sense for us (the teachers) to figure out how to improve things. In my view, to make this dashboard more useful as a personal tool for the teachers; then teachers will have to perform this analysis independently of the department's head. This separation will also avoid teachers who did not get a perfect rating, trying to game the system.* (p5)

Static reports offer an insight into data that is relevant to a specific periods to support decisions. They are generated in Word, Excel or PowerPoint and exported mostly into PDFs, usually found in print or emails and include static data about a specific area (Graham, 2020). Even though these reports provide valuable information about a specific period, there is no way to further investigate insights they present. In contrast, dynamic reports or dashboards offer real-time and dynamic reports that provide access to the most up-to-date information or real-time information. Furthermore, they allow users to interact effectively and efficiently with data through interactive features and other functionalities to conduct basic and advanced data analysis (Sarikaya et al., 2018). They allow the user to further investigate the information they know and spot potential business opportunities they thought never existed (Malik, 2005). This investigation will help them make better decisions and respond effectively to changes (Franklin et al., 2017). Static reports offer specific values; however, dynamic reports such as dashboards allow the user to squeeze every last drop of value from the data presented.

> *... I have been co-teaching on the teacher training with XXX, and each time we start the programme, we have to go through the troublesome process of scrolling through previous evaluations to pull them together. There is no way of actually putting them all together without going through the whole laborious task of reading every comment and making notes and then*

connecting it. However, with a user-friendly tool like this, it would be so much easier for a teacher to plan their next session without having to scroll through and finding loads of previous PDF files that have been sent and having to find them. Having one's SET data sitting in one place and then getting access to data using dashboards that can perform other functions like comparing SET scores over the years makes a big difference. (p17)

6.6.1 Conclusion

The increment in the SUS scores indicates that the participants perceived usability of TED improved moving from the first iteration to the second iteration in the development process. Overall, this study shows that perceived usefulness and usability of TED was acknowledged by the participants as a tool that will indeed be valuable to teachers. Their responses potentially support the use of dashboards to positively influence teaching quality for educational effectiveness. Furthermore, when interpreting our results in the light of efficiency and effectiveness, this research can conclude that information in the dashboard connects to teachers' professional developmental practice; teachers will be able to successfully use this new tool. Moreover, developments in teacher usage of dashboards over time and the role of experience and possible interactions with professional skills, need to be explored in future research. Consequently, this study concludes that dashboards benefit teaching practice and provide ample opportunities to improve teacher–technology interaction and dashboard usage to improve teacher data inquiry and data literacy.

CHAPTER 7

Discussion, Future Directions and Conclusion

7.1 Discussion

7.1.1 Practical Implications

I began this thesis by stating that the study was about how SET data can be analysed and presented in dashboards as a useful form of TA. I suggested the Data Science Life Cycle as a framework that can be used to process SET data from the initial business understanding stage, data acquisition, deployment, modelling through to the final data understanding stage of the cycle. I then located this study within the nodes of these five critical stages in the cycle.

I proposed that at the core of this perspective was the need to address the shift from linear processing synonymous with the previously proposed models for processing SET data (Gottipati et al., 2018a,b; Nitin et al., 2015; Pyasi et al., 2018; Shah and Pabel, 2019), to the iterative or cyclic model that pays critical attention to following SET data from business understanding, through data acquisition, deployment, modelling, and data understanding. Data understanding, while recognising the need to expand the area of focus beyond SET data, to the complete bundle of artefacts integrated into an iterative fashion to generate meaningful insight for teaching professional development.

As discussed in Chapter 2, previous research acknowledges that visualisation can facilitate the process of generating insight from SET data (Cunningham-Nelson et al., 2017; Downer et al., 2019; Luo et al., 2018a; Palmer and Campbell, 2015). However, these visualisations are usually static reports such as PDFs, Word documents, and Excel spreadsheets. It is important to note that the University of Otago used these static reports to represent SET data at the time of this study; SET data are extracted into Excel spreadsheets and exported into PDFs and printed or sent as emails to teachers. Despite their disadvantages, these static reports provide valuable insight that enables teachers to inspect and understand their teaching processes and progress. Printed or emailed reports are very difficult (or in some cases impossible) to handle, especially when performing a drill down into historical SET data or diving deep into data to investigate or focus on particular aspects of SET data.

There have been increasing advances in dynamic reporting and teaching dashboard design (Greller and Drachsler, 2012; Holstein et al., 2017, 2018; Schwendimann et al., 2016; Verbert et al., 2014). This advancement agrees with one of the themes discussed in Chapter 4; teachers indicated they wanted SET data to be visualised using dashboards. I have proposed a Teacher's Evaluation Dashboard (also known as TED) as the new way of presenting SET data (or student perception data) to help the teachers drill down to discover trends and patterns in historical SET data and dive deeper into particular aspects of student perceptions to improve upon course content or areas where students may be struggling, and reflect more on SET data.

TED provides new ways of visualising SET scores and the comments in SET, all from one place. Prior studies have focused only on one aspect, either the SET scores (Hajizadeh and Ahmadzadeh, 2014; Kitto et al., 2019) or SET comments (Altrabsheh et al., 2014; Gottipati et al., 2018a). The design of TED combined Descriptive Analytics and Diagnostic Analytics. Descriptive Analytics informed the number-ratings dashboard, and Diagnostic Analytics informed the open-ended comments dashboards. This combination of two tools in one place provides teachers with the opportunity to discover trends and drill down on SET data to focus on particular areas. The number-ratings dashboard allows teachers to search or discover trends and patterns in historical SET scores. Furthermore, the open-ended comments dashboard creates a platform for teachers to explore, drill down, and dive deep so as to focus on particular aspects of student comments.

Students' comments in SET are important for teachers (Altrabsheh et al., 2014; Cunningham-Nelson et al., 2019; Gottipati et al., 2018a; Nitin et al., 2015; Santhanam et al., 2018) as they reflect the complexity of teaching environments and the way each student reacts to specific environments or teaching strategies. In this light, multiple writers have pointed out that comments would provide more knowledge, and better insights to strengthen critical educational problems (Alhija and Fresko, 2009; Hodges and Stanton, 2007; Smith and Welicker-Pollak, 2008). As discussed in Chapter 6, the usability study I conducted indicated that teachers perceived TED as a useful tool that contributes to the teaching profession, especially for teachers who handled large classes and experienced many student comments (or struggled with time to read all the comments).

7.1.2 Theoretical Implications

As discussed in Chapter 4, teachers with more teaching experience are interested in different forms of data in addition to SET data. The significance of this outcome for this thesis rests on the exploratory fusion of SET data with other forms of educational data, such as physiological, assessment, and engagement data. TED can be extended to integrate multiple data and provide a combination of rich, live and continuous data exclusively to enrich the teacher's life, adding more value to the teaching profession and practice and raising the teaching profession to a new level.

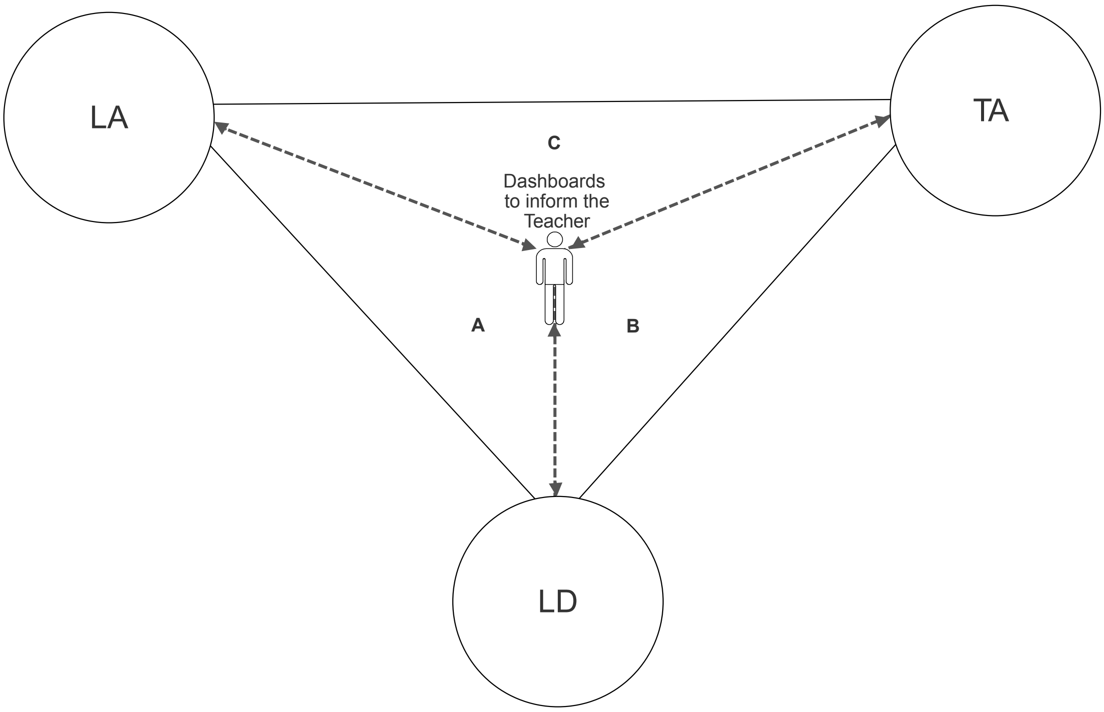


Fig. 7.1 Triadic network between TA, LA and LD to inform teaching

As illustrated in Figure 7.1, the interconnection of LA, TA and Learning Design (LD) form a triadic network with the teacher at the centre, performing value-added interactions to make informed decisions based on dashboards. Each part of this interconnection forms a triangle, totalling three triangles (A, B and C).

7.1.2.1 Triangle A

This triad illustrates the interaction between the teacher, the LA and the LD, to inform TPD. Hernndez-Leo et al. (2019) argued that LD could contribute to structuring and orchestrating the design intent with learners' digital trace patterns, advancing the knowledge and interpretation of LA. LA tailored to fit the design intent could be considered by teachers as contributing to the enhancement of the LD in subsequent design interactions. For example, LA could be an information tool to inform the tutors or designers of pedagogical decision-making (Persico and Pozzi, 2015). Hence, a teacher may want to use LA to make just-in-time pedagogical decisions, such as grouping students based on their performance.

Similarly, a teacher may want to investigate if the estimated time taken for students to carry out learning tasks is reasonable or whether adjustments need to be made to the course design (Hernndez-Leo et al., 2019; Pozzi and Persico, 2013). This domain can also provide teachers with analytics regarding the challenges and difficulties students face in the problem-solving phase while performing a task. In return, they give the teacher information in the form of the Teaching Analytics Dashboard (TAD) summarising the various challenges students encountered with that activity. They may also provide solutions on how to address them; for example, an early alert system that instantiates a dashboard for instructors using metrics calculations such as login counts and page views (Thille and Zimmaro, 2017). The data sources in the LA node can improve teachers' awareness, which could also lead to the improvement of LD and distinguish design elements that could modify future designs. Data collection in this domain is mostly automatic through virtual learning environments (e.g., LMS, MOOCs). Other forms of data collection may include social media platforms (e.g., Facebook, Twitter), wearable sensors (e.g., eye-trackers, EEGs), and software tools that support and collect data related to specific student activities and attendance (Bakharia et al., 2016; Bos and Brand-Gruwel, 2016).

7.1.2.2 Triangle B

This triangle represents the relationship between the teacher, the LD and TA. While experiencing LD, TA endeavours to handle continuous teacher engagement, progression, achievement and learner satisfaction Bakharia et al. (2016); Sergis and Sampson (2017), for example, when exploring the impact of recorded video on instructor performance and student learning. Using MOOC AB testing, teachers could experiment with whether a difference in video production setting would affect the instructors acting performance, or whether any changes in format and instructor performance will cause differences in student viewing behaviour (Chen et al., 2016).

Further, data sources in TA could assist teacher reflection on the impacts of their LD. Data collection could also be automatic , with the teachers using wearable sensors while performing teaching activities, also known as in-class analytics. Several institutions now record video content of their face-to-face classes. Some others even go a step further by collecting physiological data. As mentioned earlier, these datasets have a way of exemplifying and illustrating things that, ordinarily, a book of pedagogy cannot convey, in providing systematic feedback for the teachers. It involves capturing data during a traditional in-class, face-to-face teacher-centric instruction or teacher–student interaction (where students learn by directly or indirectly interacting with instructors in a lab or lecture hall) and analysing data to identify areas of possible improvements. The data usually captured in this setting are audio, video, body movement, brain activity, cortex activity, to mention just a few. For example, a teacher can perform diagnostic analysis on class recorded videos to expose intrinsic motivation during his lecture. This kind of diagnostic analysis could help teachers understand more about their teaching and discover areas of further improvement. SET is another form of data about the teachers; they are collected via the institutional application platforms (Hernndez-Leo et al., 2019) and can be visualised to improve teaching performance.

Analytics that happens in the LD involves visualising teaching design to facilitate teacher reflection on the lesson plan, visualising the extent to which the lesson plan aligns with the educational objectives, and finally, validating the lesson plan to highlight potential teaching design inconsistencies. For example, a teacher can visualise the number of assessment activities of the lesson plan or the various types of educational resources used in the lesson plan, to ascertain if they are still valid or obsolete. Similarly, a teacher could analyse the time allocated for each lesson activity, find out if the time allocated for each activity is good enough, or visualise the level of inconsistencies of time misappropriations and imbalances between the overall lesson plan and the individual lesson activities.

7.1.2.3 Triangle C

This area presents the communication between the teacher, the LA and the TA. Thomas (2018) explored the correlation between student ratings on teaching and student physiological data. Similarly, Schmidlin et al. (2015) established how to analyse and cross-reference data without decrypting the data sources. Hence, we argue that SET could be linked with LA such as student digital traces from LMS (Stier et al., 2019) and other forms of data (such as attendance data), without compromising privacy. This claim for data fusion could support the teachers to make informed decisions in new ways. For example, analytics performed on linked datasets could quickly reveal students' opinions that may not count at the end of the semester courses.

Visualisations could quickly recognise students with low participation rates and link it to their opinions, without revealing any identity. Additionally, teachers may be interested in comparing students with a low participation rate with those having high participation rates. This kind of information may lead teachers towards making explicit judgements with evidence. A tutor may choose to disregard the opinions of those students who participated in less than 20 per cent of in-class activities and assignments and had a low attendance rate, and hence narrowing the focus on the opinions of students who participated in improving teaching practice.

However, regarding ethical concerns, data fusion at the individual level still requires explicit and informed consent from the students whose data are collected (Menchen-Trevino, 2016). For other issues such as privacy concerns, data fusion can be problematic as this usually requires that the teachers know student identities. However, from a programmatic perspective, extra measures can be put in place to address this concern.

Algorithms can be interfaced to mask student identities to some other unique identities to make them anonymous but linked (Schmidlin et al., 2015) to provide a richer set of data for the teacher to make informed decisions.

7.1.2.4 Dashboards to Inform the Teacher

Primarily, TA is the centrepiece that links LA and LD, remodelling them to address teaching challenges. More specifically, TA argues that connecting insights generated from LA methods with those generated from in-class methods, using TA concepts, produces evidence-based professional teaching development (Ndukwe and Daniel, 2020). This interconnection provides the teachers with a better picture for improving the context in which learning happens, and enables them to be more informed about teaching and learning decisions. In other words, the method is continually providing teachers with interesting information from intelligent feedback based on data generated from a learning context to improve teaching and learning outcomes. Hence, there is the continuous provision of inspiring information for teachers from intelligent feedback, based on data generated from the learner's activities, the teacher's activities, data about the learner, data about the teacher and the learning context, to continually improve learning outcomes, LD, teaching practice and finally, to add value to TA.

7.1.3 Limitations

Data access, ethical, and privacy concerns are still significant issues faced in the educational sector (Daniel, 2015; Miyares and Catalano, 2016). Even after the University of Otago ethics committee approved this project, Quality Assurance committee denied access to use institutional SET data for this research. Ethical issues, such as consent, privacy disclosure, and the need to de-identify data were paramount concerns. As a result of this challenge in accessing actual institutional SET data, this study was limited to using simulated SET scores and free-text SET feedback volunteered by one academic staff member.

Some studies have addressed the need to de-identify academic analysis data before making it available to institutions for operational functions (Petersen, 2012; Prinsloo and Slade, 2013). TED in small classes poses a potential privacy threat, leading to identifying an individual student or group. As discussed in Chapter 6, a significant risk of using TED in small groups includes teachers deciding to isolate or exclude particular groups for a low rating. This concern can be reduced by making the future design of TED include an optional programme's filter that can be disabled when visualising SET data for smaller groups to protect confidential information and privacy.

Finally, even though the dashboards provided teachers with easy access to data and presented them with useful information to inform teaching practice, especially those who take multiple courses with many students, using TED to visualise SET data for a small group of students may be an unnecessary waste of time and effort.

7.2 Feature Direction

Future research is expected to improve on the design of TED, collect more real SET data, and integrate other forms of educational data such as physiological, engagement, and assessment data to enrich the teacher's life in making more informed decisions. I also hope to test a proposed Teaching Outcome Model (TOM) (Ndukwe et al., 2018) with TED.

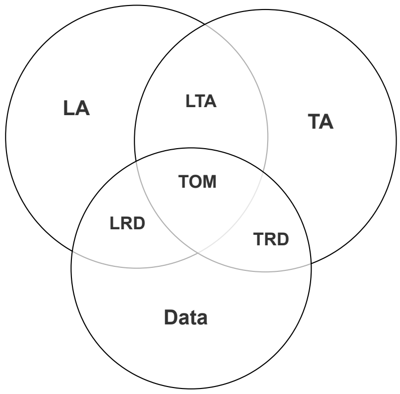


Fig. 7.2 Conceptualisation of TA

As illustrated in Figure 7.2, data is collected as a result of students or teachers interacting with learning or teaching contexts, such as LMS, LD, course materials, pedagogy and teaching styles. Learner Data (LRD) is data collected from the students; they include the following: student activity data (for example, time spent per page, number of logins, and clickstream logs), student performance data (for example, percentage of correct answers, course grades, and course completion rates) and student learning process data (for example, student interactions within an activity or task, and information about the processes the student followed to solve a problem) (Thille and Zimmaro, 2017). Teacher Data (TRD) represents collected data from the teacher's perspective (such as automatic physiological data from teacher activities) or data about the teacher (such as SET data). LA focuses on the learner and performs analytics on the LRD to improve learning outcome. On the other hand, TA focuses on the teacher and performs analytics on the TRD to improve teaching outcome. Learning and Teaching Analytics (LTA) is the common ground and realises teaching and learning outcomes. The goal of TA is to help teachers decide what to do next by adaptively organising instructional activities and learning resources according to learners' needs. It requires teachers to apply knowledge of analytics on education data for pedagogical reflection and improvement.

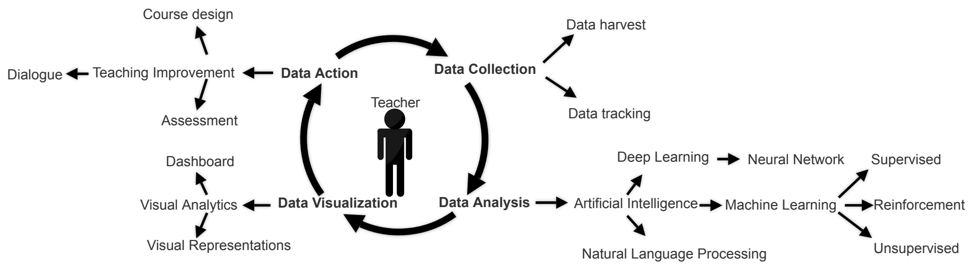


Fig. 7.3 TA Life cycle

TOM is an iterative process that guides the teachers, researchers, faculty, and administrators to use the data to improve teaching quality and learning outcomes. This model will enable teachers to investigate and evaluate their work using data, consequently improving their use of data to inform teaching practice. To build more awareness about teaching data, TOM models TA through iterative cycles of data collection, data analysis, data visualisation and action stages that are independent of each other (see Figure 7.3). As a pragmatic methodology, design-based research can guide TOM while generating insights that can support teachers' reflections on teaching and student learning. Conversely, TOM ensures that design-based research methodologies can be operational and systemised. Following the various stages outlined in the model, teachers can regularly identify, match and adjust teaching practice and learning design for all the learners' needs.

7.2.1 Data Stage

In the data collection stage, a constant stream of data accumulates from the digital traces of daily teaching activities and engagements (Kumar, 2013), including structured and unstructured data, visual and non-visual data, and historical and real-time data. It is also important to note that the rate at which diverse data accumulates in our educational system will keep growing. According to Voithofer and Golan (2018), there are several ways to mine teaching and learning data without professional knowledge beyond the necessary teacher training experience in data literacy, administering learning design and class orchestration. Subscribing to this school of thought, adopting "big data" technology in our institutions will guarantee easy access to data by the various stakeholders. Implementing big data technology will also mitigate the bottleneck of disparate data points existing in our educational sector. It therefore enables educators to focus more on instruction, set up interactive class activities, and participate more in discussions so as to create more data for evidence-based decision making. Also, the misuse of data is a broad primary concern (Roberts et al., 2017a). One critical matter is identifying the data collected, analysed, and visualised to ensure that the right people can access the data for the right purpose. Executing data governance policies around institutional data such as "open definition of purpose, scope and boundaries, even if that is broad and in some respects, open-ended" is critical (Gitelman et al., 2013, p. 6). This measure will introduce clarity and address issues around who controls data, and security and privacy.

7.2.2 Analysis Stage

This step involves the different ways of working with data to ensure data quality. Professionals such as data scientists, programmers, engineers and researchers need to work together with the teachers. They can apply data mining techniques, statistical methods, complex algorithms, and Artificial Intelligence (AI) techniques (such as NLP, ML, deep learning) to adequately transform data into a useful analytical process. Analytics in the education space presents in diverse forms: descriptive, diagnostic, predictive and prescriptive. These different forms of analytics can be used to offer a high-level view or fine-grained view of individual learners, teachers, faculties and their various activities, engagements and behaviours. Unravelling data analytics' value empowers teachers and researchers to identify problems and transform challenges into opportunities that can be exploited to support teacher reflection and enrich teacher data-literacy experiences. For example, teachers can apply NLP on text data to gather topics from discussion posts, the contributions participants have made within collaborative projects and their sentiments.

Furthermore, ML techniques could be combined with TA to enhance teaching outcome. For instance, chatbots could support the teacher by acting as a teacher assistant in large classes. However, an essential consideration in analytics is that data can be identified easily (Roberts et al., 2017a; Feldman, 2000), especially when data sets increase in size and scope. To resolve these concerns, a particular university introduced a two-stage method of data de-identification coupled with data governance to restrict data access (Garud et al., 2018).

7.2.3 Visualisation Stage

This stage ensures data is presented in useful and meaningful ways to teachers, empowering teachers with interactive visual interfaces and dashboards that facilitate teacher cognition and promote reflection about pre-processed and fine-grained teaching and learning activities. TAD can project real-time and historical information from different data sources that might not be necessarily interoperable (Moore, 2018). However, visualisation is necessarily "what you see is what you get"; meaning that the information presentation method may affect its interpretation, and consequently, may influence decision-making. Hence, it is necessary to address visualisations in diverse forms such as visual analytics and exploratory data analysis to create room for visual interactivity and exploratory visualisation to discover trends, patterns, relationships, and behaviours. For example, a teacher can use a TAD to monitor student engagement. When the student engagement is poor, it may prompt the teacher to take necessary actions such as changing their teaching material and making it more interactive. Additionally, there are questions around privacy, such as who has access to visualisations relevant to an instructor—other faculty members participating in the course, directly or indirectly, administrators, researchers, or potential employees of other institutions.

7.2.4 Action Stage

At this stage, informed decision leads to action, and actions unavoidably reshape our environment to subsequently regenerate new data. Teachers' actions are used to improve the course design and assessment (value-added formative assessment). The epistemological question to be addressed at this stage will ensure effective actions and interventions by the teacher.

7.3 Conclusion

The overall thesis focused on SET as a kind of TA. This study set out to use SET data to inform teaching practice and proposed a Data Science Life Cycle Framework for processing SET data to inform teaching practice. The Data Science Life Cycle follows the business understanding stage through data acquisition, deployment, modelling, and data understanding.

In the business understanding stage, the literature showed that the analysis, interpretation and reporting of SET data are still crucial concerns. Furthermore, visualising SET data using dashboards presents opportunities to communicate insights on learning and teaching experiences and provide a personalised, customisable, adjustable, and dynamic visualisation that teachers can easily manipulate and understand. Ethics was sought and approved by the University of Otago ethics committee to use institutional SET data.

The data acquisition stage followed. However, during the data acquisition stage, the Quality Assurance committee denied access to institutional SET data due to privacy concerns. Hence, using a data simulation technique, SET scores were generated. This data simulation approach was only used because access to institutional data was denied. Simulated SET scores informed the development of preliminary teacher's dashboards. The literature also revealed that teachers wanted more than SET scores; they wanted to analyse free-text comments in SET data. That information motivated this research to further survey teachers' perceptions of SET data and visualisation using dashboards. The investigation showed that free-text comments in SET data provided feedback that created valuable opportunities for teachers to reflect on their teaching and improve teaching effectiveness and teaching quality. Real students' free-text comments in SET data was collected for this research from one academic staff member who voluntarily offered their (student comments about teaching experiences on the value and the effectiveness of a course the staff member taught for five years). The qualitative SET data was used to validate the simulated data and preliminary dashboards.

The dashboards were redesigned (to what is known as Teacher's Evaluation Data or TED) to include free-text comments dashboards in addition to the SET scores dashboards that already existed—the deployment and modelling stages. The deployment stage used the Splunk® "big data" infrastructure to deploy the data and models that generated the dashboards. Furthermore, the modelling stage applied some pre-processing, NLP and ML models to visualise the SET data.

Finally, usability testing was carried out on TED, where teachers tested and explored the TED dashboards—the data understanding stage. These usability testing results showed that teachers perceived TED to be usable and useful to inform teaching practice and improve teaching quality.

This study has shown that, despite the controversy concerning the validity and quality of the SET data, visualisation dashboards can be used to shift teachers from a sense of weakness—that they cannot influence SET data—to a sense of strength, control and professional development capacity.

REFERENCES

Abel, T. D. and Evans, M. (2013). Cross-disciplinary participatory and contextual design research: Creating a teacher dashboard application. *IxDandA*, 19:63–76.

Abrami, P. C., dApollonia, S., and Rosenfield, S. (2007). *The dimensionality of student ratings of instruction: What we know and what we do not*, pages 385–456. Springer.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R. J. (2011). Sentiment analysis of twitter data. In *Proceedings of the workshop on language in social media (LSM 2011)*, pages 30–38.

Alhija, F. N.-A. and Fresko, B. (2009). Student evaluation of instruction: what can be learned from students written comments? *Studies in Educational evaluation*, 35(1):37–44.

Ali, L., Hatala, M., Gašević, D., and Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers and Education*, 58(1):470–489.

Altrabsheh, N., Cocea, M., and Fallahkhair, S. (2014). Learning sentiment from students feedback for real-time interventions in classrooms. In *International conference on adaptive and intelligent systems*, pages 40–49. Springer.

Altrabsheh, N., Gaber, M. M., and Cocea, M. (2013). Sa-e: sentiment analysis for education. In *International conference on intelligent decision technologies*, volume 255, pages 353–362.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, 16(7):16–07.

Anderson, R. J., Anderson, R., VanDeGrift, T., Wolfman, S., and Yasuhara, K. (2003). Promoting interaction in large classes with computer-mediated feedback. In *Designing for change in networked learning environments*, pages 119–123. Springer.

Argyris, C. (1970). *Intervention theory and method: A behavioral science view*. Addison-Wesley Reading, MA.

Arnold, K. E. and Pistilli, M. D. (2012). Course signals at purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 267–270. ACM.

Arthur, L. (2009). From performativity to professionalism: lecturers responses to student feedback. *Teaching in Higher Education*, 14(4):441–454.

Assuno, M. D., Calheiros, R. N., Bianchi, S., Netto, M. A., and Buyya, R. (2015). Big data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, 79:3–15.

Atterer, R., Wnuk, M., and Schmidt, A. (2006). Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the 15th international conference on World Wide Web*, pages 203–212.

Babik, D., Stevens, S., and Waters, A. E. (2019). Comparison of ranking and rating scales in online peer assessment: Simulation approach. In *Proceedings of the 9th International Conference on Learning Analytics and Knowledge*, pages 205–209.

Bakharia, A., Corrin, L., De Barba, P., Kennedy, G., Gaevi, D., Mulder, R., Williams, D., Dawson, S., and Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, pages 329–338. ACM.

Balahur, A., Mihalcea, R., and Montoyo, A. (2014). Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications.

Ballantyne, R., Borthwick, J., and Packer, J. (2000). Beyond student evaluation of teaching: Identifying and addressing academic staff development needs. *Assessment and Evaluation in Higher Education*, 25(3):221–236.

Banerjee, A., Bandyopadhyay, T., and Acharya, P. (2013). Data analytics: Hyped up aspirations or true potential? *Vikalpa*, 38(4):1–12.

Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123.

Barab, S. (2014). Design-based research: A methodological toolkit for engineering change. *Handbook of the learning sciences*, 2:233–270.

Barmaki, R. and Hughes, C. E. (2015). Providing real-time feedback for student teachers in a virtual rehearsal environment. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 531–537. ACM.

Bartlett, J. and Tkacz, N. (2017). Governance by dashboard: A policy paper.

Beel, J., Gipp, B., Langer, S., and Breitinger, C. (2016). paper recommender systems: a literature survey. *International Journal on Digital Libraries*, 17(4):305–338.

Beleche, T., Fairris, D., and Marks, M. (2012). Do course evaluations truly reflect student learning? evidence from an objectively graded post-test. *Economics of Education Review*, 31(5):709–719.

Benner, J. and McArthur, J. (2019). *Lessons Learned from a Multi-year Initiative to Integrate Data-Driven Design Using BIM into Undergraduate Architectural Education*, pages 857–864. Springer.

Benton, S. L. and Cashin, W. E. (2014). *Student ratings of instruction in college and university courses*, pages 279–326. Springer.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International journal of teaching and learning in higher education*, 17(1):48–62.

Berman, F., Rutenbar, R., Hailpern, B., Christensen, H., Davidson, S., Estrin, D., Franklin, M., Martonosi, M., Raghavan, P., Stodden, V., et al. (2018). Realizing the potential of data science. *Communications of the ACM*, 61(4):67–72.

Bevan, N. (2001). International standards for hci and usability. *International journal of human-computer studies*, 55(4):533–552.

Blackmore, J. (2009). Academic pedagogies, quality logics and performative universities: Evaluating teaching and what students want. *Studies in higher education*, 34(8):857–872.

Boring, A. (2015). Gender biases in student evaluation of teachers. *Paris, France*.

Bos, N. and Brand-Gruwel, S. (2016). Student differences in regulation strategies and their use of learning resources: implications for educational design. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, pages 344–353. ACM.

Braga, M., Paccagnella, M., and Pellizzari, M. (2014). Evaluating students evaluations of professors. *Economics of Education Review*, 41:71–88.

Brockx, B., Van Roy, K., and Mortelmans, D. (2012). The student as a commentator: students comments in student evaluations of teaching. *Procedia-Social and Behavioral Sciences*, 69:1122–1133.

Brooke, J. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

Brooks, D. (2013). The philosophy of data. *New York Times*, 4:2013.

Brown, M. (2020). Seeing students at scale: how faculty in large lecture courses act upon learning analytics dashboard data. *Teaching in Higher Education*, 25(4):384–400.

Bueckle, M. G. N. S. A. and Brner, K. (2017). Empowering instructors in learning management systems: interactive heat map analytics dashboard. *Retrieved November*, 2:2017.

Buzhardt, J. and Heitzman-Powell, L. (2005). Stop blaming the teachers: The role of usability testing in bridging the gap between educators and technology. *Electronic Journal for the integration of Technology in Education*, 4(1):13–29.

Calders, T. and Pechenizkiy, M. (2012). Introduction to the special section on educational data mining. *Acm Sigkdd Explorations Newsletter*, 13(2):3–6.

Casey, K. and Azcona, D. (2017). Utilizing student activity patterns to predict performance. *International Journal of Educational Technology in Higher Education*, 14(1):4.

Cha, A. P. and Romli, A. (2010). Human-computer interaction of design rules and usability elements in expert system for personality-based stress management. *International Journal of Intelligent Computing Research (IJICR)*, 1(1/2):33–42.

Chathuranga, J., Ediriweera, S., Hasantha, R., Munasinghe, P., and Ranathunga, S. (2018). Annotating opinions and opinion targets in student course feedback. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC)*.

Chen, Z., Chudzicki, C., Palumbo, D., Alexandron, G., Choi, Y.-J., Zhou, Q., and Pritchard, D. E. (2016). Researching for better instructional methods using ab experiments in moocs: results and challenges. *Research and practice in technology enhanced learning*, 11(1):9.

Cheng, Y. C. and Tam, W. M. (1997). Multi-models of quality in education. *Quality assurance in Education*.

Chounta, I.-A., McLaren, B. M., Albacete, P. L., Jordan, P. W., and Katz, S. (2016). Analysis of human-to-human tutorial dialogues: Insights for teaching analytics. In *IWTA@ EC-TEL*, pages 9–17.

Clarke, A. and Erickson, G. L. (2003). *Teacher inquiry: Living the research in everyday practice*. Psychology Press.

Coffey, M. and Gibbs, G. (2001). The evaluation of the student evaluation of educational quality questionnaire (seeq) in uk higher education. *Assessment and Evaluation in Higher Education*, 26(1):89–93.

Cohen, J. D. (1988). Noncentral chi-square: Some observations on recurrence. *The American Statistician*, 42(2):120–122.

Cox, M. and Ellsworth, D. (1997). Application-controlled demand paging for out-of-core visualization. In *Proceedings. Visualization'97 (Cat. No. 97CB36155)*, pages 235–244. IEEE.

Crooks, R. (2017). Representationalism at work: dashboards and data analytics in urban education. *Educational Media International*, 54(4):289–303.

Cunningham-Nelson, S., Baktashmotlagh, M., and Boles, W. (2019). Visualizing student opinion through text analysis. *IEEE Transactions on Education*, 62(4):305–311.

Cunningham-Nelson, S., Baktashmotlagh, M., and Boles, W. W. (2017). Visualising student satisfaction. *Proceedings of the 28th Australasian Association for Engineering Education*.

Cunningham-Nelson, S., Laundon, M., and Cathcart, A. (2020). Beyond satisfaction scores: visualising student comments for whole-of-course evaluation. *Assessment and Evaluation in Higher Education*, pages 1–16.

Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5):904–920.

Daniel, B. (2017). Big data and data science: A critical review of issues for educational research. *British Journal of Educational Technology*.

Daniel, B. K. and Butson, R. (2013). Technology enhanced analytics (tea) in higher education. *International Association for Development of the Information Society*, pages 89–96.

Danielson, C. and McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Ascd.

Davenport, T. H. et al. (2006). Competing on analytics. *Harvard business review*, 84(1):98.

de Oliveira, A. B., Alves, A. L. F., and de Souza Baptista, C. (2019). Using opinion mining in student assessments to improve teaching quality in universities. In *International Conference on Intelligent Systems Design and Applications*, pages 225–234. Springer.

Dean, A. and Gibbs, P. (2015). Student satisfaction or happiness? *Quality Assurance in Education*.

Dehler, C. and Porras-Hernandez, L. H. (1998). Using computer mediated communication (cmc) to promote experiential learning in graduate studies. *Educational Technology*, 38(3):52–55.

Dietz-Uhler, B. and Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning*, 12(1):17–26.

Dix, A. J. and Leavesley, J. (2015). Learning analytics for the academic: An action perspective. *J. UCS*, 21(1):48–65.

Doering, A., Hughes, J., and Huffman, D. (2003). Preservice teachers: Are we thinking with technology? *Journal of Research on Technology in Education*, 35(3):342–361.

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4):745–766.

Donovan, J., Mader, C. E., and Shinsky, J. (2010). Constructive student feedback: Online vs. traditional course evaluations. *Journal of Interactive Online Learning*, 9(3):283–296.

Downer, K., Wells, C., and Crichton, C. (2019). All work and no play: A text analysis. *International Journal of Market Research*, 61(3):236–251.

Drachsler, H. and Greller, W. (2012). The pulse of learning analytics understandings and expectations from the stakeholders. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 120–129.

Ducheva, Z., Pehlivanova, M., and Dineva, S. (2013). Possibilities for students to evaluate and improve electronic courses. In *The 8th International Conferemnce on Virtual Learning ICVL*, pages 135–141.

Dumas, J. S. and Loring, B. A. (2008). *Moderating usability tests: Principles and practices for interacting*. Elsevier.

Dyche, J. (2012). Big data eurekas!dont just happen. *Harvard Business Review Blog*, 20.

Dyckhoff, A. L., Zielke, D., Bltmann, M., Chatti, M. A., and Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology and Society*, 15(3).

Dye, J. F., Schatz, I. M., Rosenberg, B. A., and Coleman, S. T. (2000). Constant comparison method: A kaleidoscope of data. *The qualitative report*, 4(1):1–10.

Edstrm, K. (2008). Doing course evaluation as if learning matters most. *Higher education research and development*, 27(2):95–106.

Elliott, K. M. and Shin, D. (2002). Student satisfaction: An alternative approach to assessing this important concept. *Journal of Higher Education policy and management*, 24(2):197–209.

Elsaadani, M. A. (2013). Exploring the relationship between teaching staff age and their attitude towards information and communications technologies (ict). *International Journal of Instruction*, 6(1):215–226.

Emin-Martnez, V., Hansen, C., Triana, R., Jess, M., Wasson, B., Mor, Y., Dascalu, M., Ferguson, R., and Pernin, J.-P. (2014). Towards teacher-led design inquiry of learning. *eLearning Papers*, (36).

Epp, C. D. and Bull, S. (2015). Uncertainty representation in visualizations of learning analytics for learners: current approaches and opportunities. *IEEE Transactions on Learning Technologies*, 8(3):242–260.

Etikan, I., Musa, S. A., and Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, 5(1):1–4.

Faiola, A., Srinivas, P., and Duke, J. (2015). Supporting clinical cognition: a human-centered approach to a novel icu information visualization dashboard. In *AMIA Annual Symposium Proceedings*, volume 2015, page 560. American Medical Informatics Association.

Feldman, K. A. (2007). *Identifying exemplary teachers and teaching: Evidence from student ratings*, pages 93–143. Springer.

Feldman, M. S. (2000). Organizational routines as a source of continuous change. *Organization science*, 11(6):611–629.

Flaherty, C. (2016). Bias against female instructors. *Inside Higher Ed*, 11.

Flanders, N. A. (1970). *Analyzing teacher behavior*. Addison-Wesley P. C.

Franklin, A., Gantela, S., Shifarraw, S., Johnson, T. R., Robinson, D. J., King, B. R., Mehta, A. M., Maddow, C. L., Hoot, N. R., Nguyen, V., et al. (2017). Dashboard visualizations: Supporting real-time throughput decision-making. *Journal of biomedical informatics*, 71:211–221.

Fullan, M. (2007). *Leading in a culture of change*. John Wiley and Sons.

Galbraith, C. S., Merrill, G. B., and Kline, D. M. (2012). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? a neural network and bayesian analyses. *Research in Higher Education*, 53(3):353–374.

Garud, R., Berends, H., and Tuertscher, P. (2018). Qualitative approaches for studying innovation as process. *The Routledge Companion to Qualitative Research in Organization Studies. London: Routledge*, pages 226–247.

Gašević, D., Dawson, S., Rogers, T., and Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28:68–84.

Gauthier, G. (2013). Using teaching analytics to inform assessment practices in technology mediated problem solving tasks. In *IWTA@ LAK*. Citeseer.

Genet, D. (2013). Teachers' fear of technology–how does it impact the classroom? In *Society for Information Technology and Teacher Education International Conference*, pages 1309–1314. Association for the Advancement of Computing in Education (AACE).

Gibson, A. and Martinez-Maldonado, R. (2017). That dashboard looks nice, but what does it mean? towards making meaning explicit in learning analytics design. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction*, pages 528–532.

Ginon, B., Johnson, M. D., Turker, A., and Kickmeier-Rust, M. (2016). Helping teachers to help students by using an open learner model.

Gitelman, L., Jackson, V., Rosenberg, D., Williams, T. D., Brine, K. R., Poovey, M., Stanley, M., Garvey, E. G., Krajewski, M., Raley, R., et al. (2013). "facts and FACTS": Abolitionists' database innovations. *Raw Data Is an Oxymoron*, pages 89–102.

Goggins, S. P., Galyen, K., Petakovic, E., and Laffey, J. M. (2016). Connecting performance to social structure and pedagogy as a pathway to scaling learning analytics in moocs: an exploratory study. *Journal of Computer Assisted Learning*, 32(3):244–266.

Goodall, P., Sharpe, R., and West, A. (2019). A data-driven simulation to support remanufacturing operations. *Computers in Industry*, 105:48–60.

Goodwin, C. (2003). Professional vision, in'american anthropologist', 3. *trad it. Visioni professionali, in Goodwin*, pages 17–68.

Gorham, J. (1988). The relationship between verbal teacher immediacy behaviors and student learning. *Communication education*, 37(1):40–53.

Gottipati, S., Shankararaman, V., and Lin, J. (2018a). Latent dirichlet allocation for textual student feedback analysis. *International Conference on Computers in Education (ICCE)*, pages 220–227.

Gottipati, S., Shankararaman, V., and Lin, J. R. (2018b). Text analytics approach to extract course improvement suggestions from students' feedback. *Research and Practice in Technology Enhanced Learning*, 13(1):6.

Govaerts, S., Verbert, K., Duval, E., and Pardo, A. (2012). The student activity meter for awareness and self-reflection. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 869–884.

Graham, A. (2020). Dynamic cost reporting. *Construction Journal*, pages 16–17.

Green, S. B. and Salkind, N. J. (2016). *Using SPSS for Windows and Macintosh, books a la carte*. Pearson.

Greer, J. and Mark, M. (2016). Evaluation methods for intelligent tutoring systems revisited. *International Journal of Artificial Intelligence in Education*, 26(1):387–392.

Greller, W. and Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology and Society*, 15(3):42–57.

Griffiths, D. (2017). The use of models in learning design and learning analytics. *Interaction Design and Architecture(s) Journal*, 33:113–133.

Grix, J. (2002). Introducing students to the generic terminology of social research. *Politics*, 22(3):175–186.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220.

Gruzd, A. and Conroy, N. (2020). Learning analytics dashboard for teaching with twitter. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Gummer, E. and Mandinach, E. (2015). Building a conceptual framework for data literacy. *Teachers College Record*, 117(4):n4.

Gunn, C., McDonald, J., Donald, C., Nichols, M., Milne, J., and Blumenstein, M. (2017). *Building an evidence base for teaching and learning design using learning analytics*. Ako Aotearoa (National Centre for Tertiary Teaching Excellence), Wellington.

Hajizadeh, N. and Ahmadzadeh, M. (2014). Analysis of factors that affect the students academic performance-data mining approach. *International Journal of advanced studies in Computer Science and Engineering (IJASCSE)*, 3.

Hansen, C. D. and Johnson, C. R. (2011). *Visualization handbook*. Elsevier.

Hansen, C. J. and Wasson, B. (2016). Teacher inquiry into student learning:-the tisl heart model and method for use in teachers professional development. *Nordic Journal of Digital Literacy*, 11(01):24–49.

Harland, T. and Wald, N. (2018). Vanilla teaching as a rational choice: the impact of research and compliance on teacher development. *Teaching in Higher Education*, 23(4):419–434.

Harvey, L. (2011). *The nexus of feedback and improvement*, pages 3–26. Elsevier.

Hernndez-Leo, D., MartinezMaldonado, R., Pardo, A., MuozCristbal, J. A., and RodrguezTriana, M. J. (2019). Analytics for learning design: A layered framework and tools. *British Journal of Educational Technology*, 50(1):139–152.

Hessler, M., Pöpping, D. M., Hollstein, H., Ohlenburg, H., Arnemann, P. H., Massoth, C., Seidel, L. M., Zarbock, A., and Wenk, M. (2018). Availability of cookies during an academic course session affects evaluation of teaching. *Medical Education*, 52(10):1064–1072.

Hey, A. J., Tansley, S., and Tolle, K. M. (2009). *The fourth paradigm: data-intensive scientific discovery*, volume 1. Microsoft research Redmond, WA.

Hodges, L. C. and Stanton, K. (2007). Translating comments on student evaluations into the language of learning. *Innovative Higher Education*, 31(5):279–286.

Holstein, K., McLaren, B. M., and Aleven, V. (2017). Intelligent tutors as teachers' aides: Exploring teacher needs for real-time analytics in blended classrooms. In *International Learning Analytics and Knowledge Conference*, pages 257–266.

Holstein, K., McLaren, B. M., and Aleven, V. (2018). Student learning benefits of a mixed-reality teacher awareness tool in ai-enhanced classrooms. In *International conference on artificial intelligence in education*, pages 154–168. Springer.

Huxham, M., Laybourn, P., Cairncross, S., Gray, M., Brown, N., Goldfinch, J., and Earl, S. (2008). Collecting student feedback: a comparison of questionnaire and other methods. *Assessment and Evaluation in Higher Education*, 33(6):675–686.

IAU (2006). *World List of Universities and Other Institutions of Higher Education*. Palgrave MacMillan.

Idreos, S., Papaemmanouil, O., and Chaudhuri, S. (2015). Overview of data exploration techniques. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 277–281.

Ifenthaler, D., Gibson, D., and Dobozy, E. (2018). Informing learning design through analytics: Applying network graph analysis. *Australasian Journal of Educational Technology*, 34(2).

Isely, P. and Singh, H. (2005). Do higher grades lead to favorable student evaluations? *The Journal of Economic Education*, 36(1):29–42.

James, D. E., Schraw, G., and Kuch, F. (2015). Using the sampling margin of error to assess the interpretative validity of student evaluations of teaching. *Assessment and Evaluation in Higher Education*, 40(8):1123–1141.

Jones, D. and Gregor, S. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems*, 8(5):1.

Jordan, D. W. (2011). *Re-thinking student written comments in course evaluations: Text mining unstructured data for program and institutional assessment*. PhD thesis.

Jordan, P. W., Thomas, B., McClelland, I. L., and Weerdmeester, B. (1996). *Usability evaluation in industry*. CRC Press.

Jung, W. I. (2018). A methodology on weapon combat effectiveness analytics using big data and live, virtual, or/and constructive simulations. *Electronic Theses and Dissertations*, 5889:2004–2019.

Kanter, J. M. and Veeramachaneni, K. (2015). Deep feature synthesis: Towards automating data science endeavors. In *2015 IEEE international conference on data science and advanced analytics (DSAA)*, pages 1–10. IEEE.

Käser, T., Hallinen, N. R., and Schwartz, D. L. (2017). Modeling exploration strategies to predict student performance within a learning environment and beyond. In *Proceedings of the Seventh International Learning Analytics and Knowledge Conference*, pages 31–40.

Kerzner, H. (2017). *Project management metrics, KPIs, and dashboards: a guide to measuring and monitoring project performance*. John Wiley and Sons.

Khedher, A. B., Jraidi, I., and Frasson, C. (2019). Tracking students mental engagement using eeg signals during an interaction with a virtual learning environment. *Journal of Intelligent Learning Systems and Applications*, 11(1):1–14.

Kim, J., Jo, I.-H., and Park, Y. (2016). Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement. *Asia Pacific Education Review*, 17(1):13–24.

Kite, M. E., Subedi, P. C., and Bryant-Lees, K. B. (2015). Students perceptions of the teaching evaluation process. *Teaching of Psychology*, 42(4):307–314.

Kitto, K., Williams, C., and Alderman, L. (2019). Beyond average: Contemporary statistical techniques for analysing student evaluations of teaching. *Assessment and Evaluation in Higher Education*, 44(3):338–360.

Kolko, J. (2010). *Thoughts on Interaction Design*. Morgan Kaufmann.

Koufakou, A., Gosselin, J., and Guo, D. (2016). Using data mining to extract knowledge from student evaluation comments in undergraduate courses. In *International Joint Conference on Neural Networks (IJCNN)*, pages 3138–3142. IEEE.

Kroonenberg, P. and Verbeek, A. (2018). The tale of cochran's rule: My contingency table has so many expected values smaller than 5, what am i to do? *The American Statistician*, 72(2):175–183.

KU, O., LIANG, J.-K., CHANG, S.-B., and WU, M. (2018). Sokrates teaching analytics system (stas): An automatic teaching behavior analysis system for facilitating teacher professional development. In *Proceedings of the 26th International Conference on Computers in Education. Philippines: Asia-Pacific Society for Computers in Education*.

Kuhn, T. S. (2012). *The structure of scientific revolutions*. University of Chicago press.

Kumar, S. (2013). Big data: A game changer for e-learning. `https://learnnovators.com/blog/big-data-a-game-changer-for-e-learning/`. Online; accessed 01 June 2018.

Lalata, J.-a. P., Gerardo, B., and Medina, R. (2019). A correlation analysis of the sentiment analysis scores and numerical ratings of the students in the faculty evaluation. In *International Conference on Artificial Intelligence and Pattern Recognition*, pages 140–144.

Lemos, A., Caminhas, W., and Gomide, F. (2013). Evolving intelligent systems: Methods, algorithms and applications. In *Emerging paradigms in machine learning*, pages 117–159. Springer.

Lempinen, H. (2012). Constructing a design framework for performance dashboards. In *Scandinavian Conference on Information Systems*, pages 109–130. Springer.

Lewis, J. R. and Sauro, J. (2009). The factor structure of the system usability scale. In *International conference on human centered design*, pages 94–103. Springer.

Libbrecht, P., Kortenkamp, U., Rebholz, S., and Mller, W. (2013). Tales of a companion teacher analytics. In *IWTA@ LAK*.

Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54:94–106.

Lu, L. (2018). Teacher, teaching, and technology: The changed and unchanged. *International Education Studies*, 11(8):39–50.

Lughofer, E. (2018). Model explanation and interpretation concepts for stimulating advanced human-machine interaction with expert-in-the-loop. In *Human and Machine Learning*, pages 177–221. Springer.

Luo, W., Liu, F., and Litman, D. (2018a). An improved phrase-based approach to annotating and summarizing student course responses. *arXiv preprint arXiv:1805.10396*.

Luo, W., Liu, F., Liu, Z., and Litman, D. (2018b). A novel ilp framework for summarizing content with high lexical variety. *Natural Language Engineering*, 24(6):887–920.

Macfadyen, L. P., Dawson, S., Prest, S., and Gašević, D. (2016). Whose feedback? a multilevel analysis of student completion of end-of-term teaching evaluations. *Assessment and Evaluation in Higher Education*, 41(6):821–839.

MacNell, L., Driscoll, A., and Hunt, A. N. (2015). Whats in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4):291–303.

Malik, S. (2005). *Enterprise dashboards: design and best practices for IT*. John Wiley and Sons.

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, 47(2):71–85.

Mangaroska, K. and Giannakos, M. (2017). Learning analytics for learning design: Towards evidence-driven decisions to enhance learning. In *European conference on technology enhanced learning*, pages 428–433. Springer.

Mardiana, H. (2018). Lecturer's attitude towards advance technology and its impact to the learning process: Case study in tangerang city campuses. *Online Submission*, 4(1):12–25.

Mardiana, H. (2020). Lecturers adaptability to technological change and its impact on the teaching process. *JPI (Jurnal Pendidikan Indonesia)*, 9(2):275–289.

Marsh, H. W. (2007). *Students evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness*, pages 319–383. Springer.

Marsh, H. W. and Roche, L. (1993). The use of students evaluations and an individually structured intervention to enhance university teaching effectiveness. *American educational research journal*, 30(1):217–251.

Martinez, F., Taut, S., and Schaaf, K. (2016). Classroom observation for evaluating and improving teaching: An international perspective. *Studies in Educational Evaluation*, 49:15–29.

Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., and Clayphan, A. (2015). Latux: An iterative workflow for designing, validating and deploying learning analytics visualisations. *Journal of Learning Analytics*, 2(3):9–39.

Masood, Z., Hoda, R., and Blincoe, K. (2018). Adapting agile practices in university contexts. *Journal of Systems and Software*, 144:501–510.

Mazoué, J. G. (1990). Diagnosis without doctors. *The Journal of medicine and philosophy*, 15(6):559–579.

McCormack, C. (2005). Reconceptualizing student evaluation of teaching: An ethical framework for changing times. *Assessment and Evaluation in Higher Education*, 30(5):463–476.

McKenney, S. and Mor, Y. (2015). Supporting teachers in data - informed educational design. *British journal of educational technology*, 46(2):265–279.

Menchen-Trevino, E. (2016). Web historian: Enabling multi-method and independent research with real-world web browsing history data. *IConference 2016 Proceedings*.

Michos, K. and Hernández Leo, D. (2016). Towards understanding the potential of teaching analytics within educational communities. In *International Workshop on Teaching Analytics. Proceedings of the Fourth International Workshop on Teaching Analytics, in conjunction with EC-TEL*, pages 1–8. CEUR Workshop Proceedings.

Miyares, J. and Catalano, D. (2016). Institutional analytics is hard work: A five-year journey. *EDUCAUSE review*.

Mohanan, K. (2005). The place of student feedback in teaching evaluation.

Molenaar, I. and Knoop-van Campen, C. (2017). Teacher dashboards in practice: Usage and impact. In *European Conference on Technology Enhanced Learning*, pages 125–138. Springer.

Moore, B. L. (2018). *The Role of Data Analytics in Education Possibilities and Limitations*, book section 6. New York, 1st edition.

Moore, R. C. and Lewis, W. (2010). Intelligent selection of language model training data.

Mortelmans, D. and Spooren, P. (2009). A revalidation of the set37 questionnaire for student evaluations of teaching. *Educational Studies*, 35(5):547–552.

Moskal, A. and Cramond, J. (2012). Otago inform: online evaluation request system. In *Proceedings of the 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction*, pages 96–96.

Munezero, M., Montero, C. S., Mozgovoy, M., and Sutinen, E. (2013). Exploiting sentiment analysis to track emotions in students' learning diaries. In *Proceedings of the 13th Koli Calling International Conference on Computing Education Research*, pages 145–152.

Mller, W., Rebholz, S., and Libbrecht, P. (2016). Automatic inspection of e-portfolios for improving formative and summative assessment. In *International Symposium on Emerging Technologies for Education*, pages 480–489. Springer.

Nargesian, F., Samulowitz, H., Khurana, U., Khalil, E. B., and Turaga, D. S. (2017). Learning feature engineering for classification. In *IJCAI*, pages 2529–2535.

Ndukwe, I., Daniel, B., and Butson, R. (2018). Data science approach for simulating educational data: Towards the development of teaching outcome model (tom). *Big Data and Cognitive Computing*, 2(3):24.

Ndukwe, I. G. and Daniel, B. K. (2020). Teaching analytics, value and tools for teacher data literacy: a systematic and tripartite approach. *International Journal of Educational Technology in Higher Education*, 17(1):1–31.

Nelson, L. K., Burk, D., Knudsen, M., and McCall, L. (2018). The future of coding: A comparison of hand-coding and three types of computer-assisted text analysis methods. *Sociological Methods and Research*, page 0049124118769114.

Neumann, R. (2000). Communicating student evaluation of teaching results: rating interpretation guides (rigs). *Assessment and Evaluation in Higher Education*, 25(2):121–134.

Nguyen, A., Gardner, L. A., and Sheridan, D. (2017). A multi-layered taxonomy of learning analytics applications. In *PACIS*, page 54.

Nguyen, P. X., Hong, T. T., Van Nguyen, K., and Nguyen, N. L.-T. (2018). Deep learning versus traditional classifiers on vietnamese students' feedback corpus. In *NAFOSTED Conference on Information and Computer Science (NICS)*, pages 75–80. IEEE.

Nielsen, J. (1993). Usability engineering. *Fremont, California: Morgan*.

Nielsen, J. (2012). Usability 101: Introduction to usability.

Nielsen, J. and Loranger, H. (2006). *Prioritizing web usability*. Pearson Education.

Nitin, G. I., Swapna, G., and Shankararaman, V. (2015). Analyzing educational comments for topics and sentiments: A text analytics approach. In *2015 IEEE frontiers in education conference (FIE)*, pages 1–9. IEEE.

Olson, D. L. and Lauhoff, G. (2019). *Descriptive data mining*, pages 129–130. Springer.

Ong, V. K. (2015). Big data and its research implications for higher education: Cases from uk higher education institutions. In *Advanced Applied Informatics (IIAI-AAI), 2015 IIAI 4th International Congress on*, pages 487–491. IEEE.

Oon, P.-T., Spencer, B., and Kam, C. C. S. (2017). Psychometric quality of a student evaluation of teaching survey in higher education. *Assessment and Evaluation in Higher Education*, 42(5):788–800.

Ordenes, F. V., Theodoulidis, B., Burton, J., Gruber, T., and Zaki, M. (2014). Analyzing customer experience feedback using text mining: A linguistics-based approach. *Journal of Service Research*, 17(3):278–295.

Palmer, S. (2012). Student evaluation of teaching: keeping in touch with reality. *Quality in higher education*, 18(3):297–311.

Palmer, S. and Campbell, M. (2015). Text analytics visualisation of course experience questionnaire student comment data in science and technology. In *Australasian Association for Engineering Education*, pages 1–10. School of Engineering, Deakin University.

Pantazos, K. and Vatrapu, R. (2016). Enhancing the professional vision of teachers: A physiological study of teaching analytics dashboards of students' repertory grid exercises in business education. In *System Sciences (HICSS), 2016 49th Hawaii International Conference on*, pages 41–50. IEEE.

Pantazos, K., Vatrapu, R. K., and Hussain, A. (2013). Visualizing repertory grid data for formative assessment. In *IWTA@ LAK*.

Parkin, H. J., Hepplestone, S., Holden, G., Irwin, B., and Thorpe, L. (2012). A role for technology in enhancing students engagement with feedback. *Assessment and Evaluation in Higher Education*, 37(8):963–973.

Paryudi, I. and Fenz, S. (2014). Evaluation of user interface design of semergy system. In *Proceedings of the 16th International Conference on Information Integration and Web-based Applications and Services*, pages 469–473.

Pennings, H. J., Brekelmans, M., Wubbels, T., van der Want, A. C., Claessens, L. C., and van Tartwijk, J. (2014). A nonlinear dynamical systems approach to real-time teacher behavior: Differences between teachers. *Nonlinear Dynamics, Psychology, and Life Sciences*, 18(1):23–45.

Persico, D. and Pozzi, F. (2015). Informing learning design with learning analytics to improve teacher inquiry. *British Journal of Educational Technology*, 46(2):230–248.

Petersen, R. J. (2012). Policy dimensions of analytics in higher education. *Educause Review*, 47(4):44–46.

Poulos, A. and Mahony, M. J. (2008). Effectiveness of feedback: The students perspective. *Assessment and Evaluation in Higher Education*, 33(2):143–154.

Pozzi, F. and Persico, D. (2013). Sustaining learning design and pedagogical planning in cscl. *Research in Learning Technology*, 21.

Pradini, R. S., Kriswibowo, R., and Ramdani, F. (2019). Usability evaluation on the sipr website uses the system usability scale and net promoter score. In *2019 International Conference on Sustainable Information Engineering and Technology (SIET)*, pages 280–284. IEEE.

Prensky, M. (2009). H. sapiens digital: From digital immigrants and digital natives to digital wisdom. *Innovate: journal of online education*, 5(3).

Price, A., Wieman, C., and Perkins, K. (2019). Teaching with simulations. *The Science Teacher*, 86(7):46–52.

Prieto, L. P., Rodrguez-Triana, M. J., Kusmin, M., and Laanpere, M. (2017). Smart school multimodal dataset and challenges. In *Joint Proceedings of the Sixth Multimodal Learning Analytics (MMLA) Workshop and the Second Cross-LAK Workshop co-located with 7th International Learning Analytics and Knowledge Conference*, volume 1828, pages 53–59. CEUR.

Prieto, L. P., Sharma, K., Dillenbourg, P., and Jess, M. (2016). Teaching analytics: towards automatic extraction of orchestration graphs using wearable sensors. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge*, pages 148–157. ACM.

Prieto, L. P., Sharma, K., Kidzinski, Ł., Rodríguez-Triana, M. J., and Dillenbourg, P. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of computer assisted learning*, 34(2):193–203.

Prinsloo, P. and Slade, S. (2013). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. In *Proceedings of the third international conference on learning analytics and knowledge*, pages 240–244.

Purcell, K., Heaps, A., Buchanan, J., and Friedrich, L. (2013). How teachers are using technology at home and in their classrooms. *Washington, DC: Pew Research Centers Internet and American Life Project*.

Puschmann, C. and Powell, A. (2018). Turning words into consumer preferences: How sentiment analysis is framed in research and the news media. *Social Media+ Society*, 4(3):2056305118797724.

Pyasi, S., Gottipati, S., and Shankararaman, V. (2018). Sufat-an analytics tool for gaining insights from student feedback comments. In *IEEE Frontiers in Education Conference (FIE)*, pages 1–9. IEEE.

Rajaraman, A. and Ullman, J. D. (2011). *Mining of massive datasets*. Cambridge University Press.

Rajput, Q., Haider, S., and Ghani, S. (2016). Lexicon-based sentiment analysis of teachers evaluation. *Applied Computational Intelligence and Soft Computing*, 2016.

Ramsden, P. (2003). Student surveys and quality assurance. In *Proceedings of the Australian Universities Quality Forum*, pages 126–135.

Rashid, A., Asif, S., Butt, N. A., and Ashraf, I. (2013). Feature level opinion mining of educational student feedback data using sequential pattern mining and association rule mining. *International Journal of Computer Applications*, 81(10).

Rathke, D. and Harmon, J. (2011). Purposes of student course evaluations, owens community college.

Ravi, K. and Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-based systems*, 89:14–46.

Roberts, L. D., Chang, V., and Gibson, D. (2017a). *Ethical considerations in adopting a university-and system-wide approach to data and learning analytics*, pages 89–108. Springer.

Roberts, L. D., Howell, J. A., and Seaman, K. (2017b). Give me a customizable dashboard: Personalized learning analytics dashboards in higher education. *Technology, Knowledge and Learning*, 22(3):317–333.

Roschelle, J., Dimitriadis, Y., and Hoppe, U. (2013). Classroom orchestration: synthesis. *Computers and Education*, 69:523–526.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, 110(1):145.

Saar, M., Kusmin, M., Laanpere, M., Prieto, L. P., and Rtmann, T. (2017). Work in progresssemantic annotations and teaching analytics on lecture videos in engineering education. In *Global Engineering Education Conference (EDUCON), 2017 IEEE*, pages 1548–1551. IEEE.

Saar, M., Prieto, L. P., Rodrguez-Triana, M. J., and Kusmin, M. (2018). Personalized, teacher-driven in-action data collection: technology design principles. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, pages 58–62. IEEE.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment and Evaluation in Higher Education*, 35(5):535–550.

Saif, H., He, Y., and Alani, H. (2012). Semantic sentiment analysis of twitter. In *International semantic web conference*, pages 508–524. Springer.

Saifee, V. and Jay, T. (2013). Applications and challenges for sentiment analysis: a survey. *International Journal of Engineering Research and Technology (IJERT)*, 2.

Sampson, D. (2017). Teaching and learning analytics to support teacher inquiry. In *Global Engineering Education Conference (EDUCON), 2017 IEEE*, pages 1881–1882. IEEE.

Santhanam, E., Lynch, B., and Jones, J. (2018). Making sense of student feedback using text analysis–adapting and expanding a common lexicon. *Quality Assurance in Education*.

Sarikaya, A., Correll, M., Bartram, L., Tory, M., and Fisher, D. (2018). What do we talk about when we talk about dashboards? *IEEE transactions on visualization and computer graphics*, 25(1):682–692.

Sayed-Mouchaweh, M. and Lughofer, E. (2012). *Learning in non-stationary environments: methods and applications*. Springer Science and Business Media.

Schempp, P., McCullick, B., Pierre, P. S., Woorons, S., You, J., and Clark, B. (2004). Expert golf instructors' student-teacher interaction patterns. *Research Quarterly for Exercise and Sport*, 75(1):60–70.

Schmidlin, K., Clough-Gorr, K. M., and Spoerri, A. (2015). Privacy preserving probabilistic record linkage (p3rl): a novel method for linking existing health-related data and maintaining participant confidentiality. *BMC medical research methodology*, 15(1):46.

Schneiderman, B. and Plaisant, C. (2005). Designing the user interface: Strategies for effective human-computer interactions. the united states of america.

Schumacher, C. and Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, 78:397–407.

Schwendimann, B., Rodriguez-Triana, M., Vozniuk, A., Prieto, L., Shirvani Boroujeni, M., Holzer, A., Gillet, D., and Dillenbourg, P. (2016). Understanding learning at a glance: A systematic literature review of learning dashboards. In *International Conference on Learning Analytics and Knowledge (LAK)*, volume 10, pages 148–157.

Seldin, P., Miller, J. E., and Seldin, C. A. (2010). *The teaching portfolio: A practical guide to improved performance and promotion/tenure decisions*. John Wiley and Sons.

Sergis, S. and Sampson, D. G. (2017). *Teaching and learning analytics to support teacher inquiry: A systematic literature review*, pages 25–63. Springer.

Shackelford, S. J. and Raymond, A. H. (2014). Building the virtual courthouse: Ethical considerations for design, implementation, and regulation in the world of odr. *Wis. L. Rev.*, page 615.

Shah, M. and Pabel, A. (2019). Making the student voice count: using qualitative student feedback to enhance the student experience. *Journal of Applied Research in Higher Education*.

Sharp, H. (2003). *Interaction design*. John Wiley and Sons.

Shen, J., Chen, H., and Jiang, J. (2018). A research on techniques for data fusion and analysis of cross-platform mooc data. In *2018 17th International Conference on Information Technology Based Higher Education and Training (ITHET)*, pages 1–8. IEEE.

Shneiderman, B. (2004). Designing for fun: how can we design user interfaces to be more fun? *interactions*, 11(5):48–50.

Shneiderman, B. and Plaisant, C. (2010). *Designing the user interface: Strategies for effective human-computer interaction*. Pearson Education India.

Siemens, G. (2012). Learning analytics: envisioning a research discipline and a domain of practice. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 4–8.

Siemens, G. and Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5):30.

Sindhu, I., Daudpota, S. M., Badar, K., Bakhtyar, M., Baber, J., and Nurunnabi, M. (2019). Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation. *IEEE Access*, 7:108729–108741.

Smith, K. and Welicker-Pollak, M. (2008). What can they say about my teaching? teacher educators' attitudes to standardised student evaluation of teaching. *European Journal of Teacher Education*, 31(2):203–214.

Spooren, P., Vandermoere, F., Vanderstraeten, R., and Pepermans, K. (2017). Exploring high impact scholarship in research on student's evaluation of teaching (set). *Educational Research Review*, 22:129–141.

Steadman, I. (2013). Big data and the death of the theorist. *Wired Online*, 25:2013.

Stier, S., Breuer, J., Siegers, P., and Thorson, K. (2019). Integrating survey data and digital trace data: Key issues in developing an emerging field.

Stiggins, R. (2017). *The perfect assessment system*. ASCD.

Stodden, V. (2020). The data science life cycle: a disciplined approach to advancing data science as a science. *Communications of the ACM*, 63(7):58–66.

Stupans, I., McGuren, T., and Babey, A. M. (2016). Student evaluation of teaching: A study exploring student rating instrument free-form text comments. *Innovative Higher Education*, 41(1):33–42.

Subramanya, S. (2014). Toward a more effective and useful end-of-course evaluation scheme. *Journal of Research in Innovative Teaching*, 7(1).

Suehiro, D., Taniguchi, Y., Shimada, A., and Ogata, H. (2017). Face-to-face teaching analytics: Extracting teaching activities from e-book logs via time-series analysis. In *Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference on*, pages 267–268. IEEE.

Taniguchi, Y., Suehiro, D., Shimada, A., and Ogata, H. (2017). Revealing hidden impression topics in students journals based on nonnegative matrix factorization. In *Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference on*, pages 298–300. IEEE.

Tatto, M. T., Schmelkes, S., Guevara, M. D. R., and Tapia, M. (2006). Implementing reform amidst resistance: The regulation of teacher education and work in mexico. *International Journal of Educational Research*, 45(4-5):267–278.

Tavares, D. L., Perkins, K., Kauzmann, M., and Velez, C. A. (2019). Towards a teacher dashboard design for interactive simulations. In *Journal of Physics: Conference Series*, volume 1287, page 012055. IOP Publishing.

Thille, C. and Zimmaro, D. (2017). Incorporating learning analytics in the classroom. *New Directions for Higher Education*, 2017(179):19–31.

Thomas, C. (2018). Multimodal teaching and learning analytics for classroom and online educational settings. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*, pages 542–545. ACM.

Thomas, M. J. (2002). Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning*, 18(3):351–366.

Todd, J. J. and Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, 428(6984):751–754.

Tseng, C.-W., Chou, J.-J., and Tsai, Y.-C. (2018). Text mining analysis of teaching evaluation questionnaires for the selection of outstanding teaching faculty members. *IEEE Access*, 6:72870–72879.

ur Rehman, M. H., Chang, V., Batool, A., and Wah, T. Y. (2016). Big data reduction framework for value creation in sustainable enterprises. *International Journal of Information Management*, 36(6):917–928.

Van Der Aalst, W. (2016). Data science in action. In *Process mining*, pages 3–23. Springer.

Van Harmelen, M. and Workman, D. (2012). Analytics for learning and teaching. *CETIS Analytics Series*, 1(3):1–40.

van Leeuwen, A. and Rummel, N. (2017). Teacher regulation of collaborative learning: research directions for learning analytics dashboards. *Making a Difference: Prioritizing Equity and Access in CSCL*, 2(805-806):1939–1382.

Vanbrabant, L., Martin, N., Ramaekers, K., and Braekers, K. (2019). Quality of input data in emergency department simulations: Framework and assessment techniques. *Simulation Modelling Practice and Theory*, 91:83–101.

Vanhoof, J. and Schildkamp, K. (2014). From professional development for data use to data use for professional development. *Studies in educational evaluation*, (42):1–4.

Vassiliadis, P., Simitsis, A., and Skiadopoulos, S. (2002). Conceptual modeling for etl processes. In *Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP*, pages 14–21.

Vatrapu, R., Reimann, P., Bull, S., and Johnson, M. (2013a). An eye-tracking study of notational, informational, and emotional aspects of learning analytics representations. In *ACM International Conference Proceeding Series*, pages 125–134.

Vatrapu, R., Reimann, P., Hussain, A., and Kocherla, K. (2013b). Towards teaching analytics: Repertory grids for formative assessment (rgfa). In *CSCL 2013 Conference Proceedings*, volume 2, pages 422–426.

Vatrapu, R., Teplovs, C., Fujita, N., and Bull, S. (2011). Towards visual analytics for teachers' dynamic diagnostic pedagogical decision-making. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, pages 93–98. ACM.

Vatrapu, R. K. (2012). Towards semiology of teaching analytics. In *Workshop Towards Theory and Practice of Teaching Analytics, at the European Conference on Technology Enhanced Learning, TAPTA*, volume 12.

Vatrapu, R. K., Kocherla, K., and Pantazos, K. (2013c). iklassroom: Real-time, real-place teaching analytics. In *IWTA@ LAK*.

Vázquez, F. (2018). Ontology and data science. `https://towardsdatascience.com/ontology-and-data-science-45e916288cc5`. Online; accessed 01 October 2019.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., and Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, 18(6):1499–1514.

Vieira, C., Parsons, P., and Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers and Education*, 122:119–135.

Voithofer, R. and Golan, A. M. (2018). *Data Sources for Educators*, book section 5, page 18. 1 edition.

Vozniuk, A., Govaerts, S., and Gillet, D. (2013). Towards portable learning analytics dashboards. In *2013 IEEE 13th international conference on advanced learning technologies*, pages 412–416. IEEE.

Wang, J., Chang, Q., Xiao, G., Wang, N., and Li, S. (2011). Data driven production modeling and simulation of complex automobile general assembly plant. *Computers in industry*, 62(7):765–775.

Wang, W., Chen, L., Thirunarayan, K., and Sheth, A. P. (2012). Harnessing twitter" big data" for automatic emotion identification. In *International Confernece on Social Computing*, pages 587–592. IEEE.

Warman, S. M. (2015). Challenges and issues in the evaluation of teaching quality: How does it affect teachers' professional practice? a uk perspective. *Journal of veterinary medical education*, 42(3):245–251.

Watkins, J., Fabielli, M., and Mahmud, M. (2020). Sense: a student performance quantifier using sentiment analysis. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE.

Wise, A. F. and Jung, Y. (2019). Teaching with analytics: Towards a situated model of instructional decision-making. *Journal of Learning Analytics*, 6(2):53–69.

Wojton, M., Heimlich, J., Burris, A., and Tramby, Z. (2014). Sense-making of big data spring break 2013 visualization recognition and meaning making. *Report, Indiana University, Lifelong Learning Group*.

Wolff, A., Zdrahal, Z., Nikolov, A., and Pantucek, M. (2013). Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In *Proceedings of the third international conference on learning analytics and knowledge*, pages 145–149.

Worsham, N. (2018). Natural language processing text analytics splunk app. `https://github.com/geekusa/nlp-text-analytics/blob/master/PROJECT_FILES/blog.md`. Online; accessed 10 November 2019.

Wu, C., Buyya, R., and Ramamohanarao, K. (2016). Big data analytics= machine learning+ cloud computing. *arXiv preprint arXiv:1601.03115*.

Wyeld, T., Jiranantanagorn, P., Shen, H., Liao, K., and Bednarz, T. (2021). Understanding the effects of real-time sentiment analysis and morale visualisation in backchannel systems: A case study. *International Journal of Human-Computer Studies*, 145:102524.

Xing, W., Guo, R., Petakovic, E., and Goggins, S. (2015). Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, 47:168–181.

Xiong, G. and Li, Z. (2007). Role-based access control in educational administration system. *Network and Computer Security*, 7.

Xu, B. and Recker, M. (2012). Teaching analytics: A clustering and triangulation study of digital library user data. *Educational Technology and Society*, 15(3):103–115.

Yigitbasioglu, O. M. and Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, 13(1):41–59.

Zhang, M., Wang, J., and Zhou, R. (2019). Entropy value-based pursuit projection cluster for the teaching quality evaluation with interval number. *Entropy*, 21(2):203.

Zhang, Y., Jin, R., and Zhou, Z.-H. (2010). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1-4):43–52.

LIST OF PUBLICATIONS

International peer-reviewed journals and conferences:

1. **Ndukwe, I.G.** and Daniel, B.K. (2020) Teaching analytics, value and tools for teacher data literacy: a systematic and tripartite approach. Int J Educ Technol High Educ 17, 22. https://doi.org/10.1186/s41239-020-00201-6

2. **Ndukwe, I. G.**, Amadi, C. E., Nkomo, L. M., and Daniel, B. K. (2020). Automatic Grading System using Sentence-BERT Network. The 21^{st} International Conference on Artificial Intelligence in Education.

3. Nkomo, L. M., **Ndukwe, I. G.**, Daniel B. K. (2020). Social Network and Sentiment Analysis: Investigation of Students Perspectives on Lecture Recording. 19th European Conference on e-Learning.

4. **Ndukwe, I. G.**, Daniel, B. K., and Amadi, C. E. (2019). A Machine Learning Grading System Using Chatbots. The 20^{th} International Conference on Artificial Intelligence in Education.

5. **Ndukwe, I. G.**, Daniel, B. k., and Butson, R. (2018). Data Science Approach for Simulating Educational Data: Towards the Development of Teaching Outcome Model (TOM). Big Data and Cognitive Computing, 2(3), 24.

Oral presentations:

1. **Ndukwe, I. G.** and Daniel, B. K. (2019). Conceptualisation of Teaching Analytics, Value and Tools: A Systematic and Tripartite Literature Review, workshop held at Higher Education Development Centre (HEDC), University of Otago, on the 6 August, 2018.

2. **Ndukwe, I. G.** and Daniel, B. K. (2019). Using chatbots for interactive assessments; the big idea, workshop held at Higher Education Development Centre (HEDC), University of Otago, on the 4 June, 2018.

3. **Ndukwe, I. G.** and Daniel, B. K. (2018). The Development of a System for Analytics: Big Data in Higher Education, workshop held at Higher Education Development Centre (HEDC), University of Otago, on the 17 May, 2018.

4. **Ndukwe, I. G.** and Daniel, B. K. (2018). Data Science Approach in Education: Simulating Student Teaching Evaluation Data, workshop held at Higher Education Development Centre (HEDC), University of Otago, on the 14 August, 2018.

Awards during PhD study:

1. Full doctoral scholarship from 2017 to 2020 by University of Otago, New Zealand.

2. Conference scholarship for poster presentation by the 20^{th} international conference on artificial intelligence in education, Chicago, Illinois.

Appendices

Appendix A

LIST OF PUBLICATIONS

SN	Title	Authors	Year	Scopus Source title
TA_01	An eye-tracking study of notational, informational, and emotional aspects of learning analytics representations	Vatrapu, R., Reimann, P., Bull, S., Johnson, M.	2013	ACM International Conference Proceeding Series
TA_02	Analysis of human-to-human tutorial dialogues: Insights for teaching analytics	Chounta, I.-A., McLaren, B.M., Albacete, P., Jordan, P., Katz, S.	2016	CEUR Workshop Proceedings
TA_03	Connecting performance to social structure and pedagogy as a pathway to scaling learning analytics in MOOCs: An exploratory study	Goggins, S.P., Galyen, K.D., Petakovic, E., Laffey, J.M.	2016	Journal of Computer Assisted Learning
TA_04	Enhancing the professional vision of teachers: A physiological study of teaching analytics dashboards of students' repertory grid exercises in business education	Pantazos, K., Vatrapu, R.	2016	Proceedings of the Annual Hawaii International Conference on System Sciences
TA_05	Face-to-Face Teaching Analytics: Extracting Teaching Activities from E-Book Logs via Time-Series Analysis	Suehiro, D., Taniguchi, Y., Shimada, A., Ogata, H.	2017	Proceedings - IEEE 17th International Conference on Advanced Learning Technologies, ICALT 2017
TA_06	Group informatics: A multi-domain perspective on the development of teaching analytics	Goggins, S.P.	2012	CEUR Workshop Proceedings
TA_07	Helping teachers to help students by using an open learner model	Ginon, B., Johnson, M.D., Turker, A., Kickmeier-Rust, M.	2016	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)
TA_08	IKlassroom: Real-time, real-place teaching analytics	Vatrapu, R.K., Kocherla, K., Pantazos, K.	2013	CEUR Workshop Proceedings
TA_09	Learning analytics for the academic: An action perspective	Dix, A., Leavesley, J.	2015	Journal of Universal Computer Science
TA_10	Providing real-time feedback for student teachers in a virtual rehearsal environment	Barmaki, R., Hughes, C.E.	2015	ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction
TA_11	Revealing Hidden Impression Topics in Students' Journals Based on Nonnegative Matrix Factorization	Taniguchi, Y., Suehiro, D., Shimada, A., Ogata, H.	2017	Proceedings - IEEE 17th International Conference on Advanced Learning Technologies, ICALT 2017
TA_12	Smart school multimodal dataset and challenges	Prieto, L.P., Rodrguez-Triana, M.J., Kusmin, M., Laanpere, M.	2017	CEUR Workshop Proceedings
TA_13	Tales of a companion teacher analytics	Libbrecht, P., Kortenkamp, U., Rebholz, S., Mller, W.	2013	CEUR Workshop Proceedings
TA_14	Teaching analytics: A clustering and triangulation study of digital library user data	Xu, B., Recker, M.	2012	Educational Technology and Society
TA_15	Teaching analytics: Towards automatic extraction of orchestration graphs using wearable sensors	Prieto, L.P., Sharma, K., Dillenbourg, P., Jess, M.	2016	ACM International Conference Proceeding Series
TA_16	Teaching and learning analytics to support teacher inquiry	Sampson, D.	2017	IEEE Global Engineering Education Conference, EDUCON
TA_17	Towards a teaching analytics tool for supporting reflective educational (re)design in Inquiry-based STEM Education	Sergis, S., Sampson, D.G.	2016	Proceedings - IEEE 16th International Conference on Advanced Learning Technologies, ICALT 2016
TA_18	Towards semiology of teaching analytics	Vatrapu, R.K.	2012	CEUR Workshop Proceedings
TA_19	Towards teaching analytics: Communication and negotiation tool (CoNeTo)	Vatrapu, R., Tanveer, U., Hussain, A.	2012	NordiCHI 2012: Making Sense Through Design - Proceedings of the 7th Nordic Conference on Human-Computer Interaction
TA_20	Towards teaching analytics: Repertory grids for formative assessment (RGFA)	Vatrapu, R., Reimann, P., Hussain, A., Kocherla, K.	2013	Computer-Supported Collaborative Learning Conference, CSCL
TA_21	Towards understanding the potential of teaching analytics within educational communities	Michos, K., Hernndez-Leo, D.	2016	CEUR Workshop Proceedings
TA_22	Using teaching analytics to inform assessment practices in technology mediated problem solving tasks	Gauthier, G.	2013	CEUR Workshop Proceedings
TA_23	Visualizing repertory grid data for formative assessment	Pantazos, K., Vatrapu, R., Hussain, A.	2013	CEUR Workshop Proceedings

Continued...

SN	Title	Authors	Year	Scopus Source title
TA_24	What information do teachers demand from a computerized classroom? An exploratory analysis	Martnez-Mons, A., Reffay, C., Lcuyer-Cabioch, G., Luengo, V.	2017	CEUR Workshop Proceedings
TA_25	Work in progress - Semantic annotations and teaching analytics on lecture videos in engineering education	Saar, M., Kusmin, M., Laanpere, M., Prieto, L.P., Ruutmann, T.	2017	IEEE Global Engineering Education Conference, EDUCON
TA_26	Automatic inspection of E-portfolios for improving formative and summative assessment	Mller, W., Rebholz, S., Libbrecht, P.	2017	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)
TA_27	Sokrates Teaching Analytics System (STAS): An automatic teaching behavior analysis system for facilitating teacher professional development	Ku, O., Liang, J.-K., Chang, S.-B., Wu, M.	2018	ICCE 2018 - 26th International Conference on Computers in Education, Main Conference Proceedings
TA_28	Multimodal teaching and learning analytics for classroom and online educational settings	Thomas, C.	2018	ICMI 2018 - Proceedings of the 2018 International Conference on Multimodal Interaction
TA_29	Personalized, teacher-driven in-Action data collection: Technology design principles	Saar, M., Prieto, L.P., Rodriguez-Triana, M.J., Kusmin, M.	2018	Proceedings - IEEE 18th International Conference on Advanced Learning Technologies, ICALT 2018
TA_30	A research on techniques for data fusion and analysis of cross-platform MOOC data	Shen, J., Chen, H., Jiang, J.	2018	2018 17th International Conference on Information Technology Based Higher Education and Training, ITHET 2018
TA_31	Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data	Prieto, L.P., Sharma, K., Kidzinski, ., Rodrguez-Triana, M.J., Dillenbourg, P.	2018	Journal of Computer Assisted Learning
TA_32	Latent dirichlet allocation for textual student feedback analysis	Gottipati, S., Shankararaman, V., Lin, J.	2018	International Conference on Computers in Education
TA_33	Analyzing educational comments for topics and sentiments: A text analytics approach	Nitin, G.I., Swapna, G., Shankararaman, V.	2015	IEEE frontiers in education conference
TA_34	Learning sentiment from students feedback for real-time interventions in classrooms	Altrabsheh, N., Cocea, M., Fallahkhair, S.	2014	International conference on adaptive and intelligent systems
TA_35	Analysis of factors that affect the students academic performance-Data Mining Approach	Hajizadeh, N., Ahmadzadeh, M.	2014	International Journal of advanced studies in Computer Science and Engineering (IJASCSE)
TA_36	Feature level opinion mining of educational student feedback data using sequential pattern mining and association rule mining	Rashid, A., Asif, S., Butt, N.A., Ashraf, I.	2013	International Journal of Computer Applications
TA_37	Visualizing Student Opinion Through Text Analysis	Cunningham-Nelson, S., Baktashmotlagh, M., Boles, W.	2019	IEEE Transactions on Education
TA_38	Text analytics visualisation of Course Experience Questionnaire student comment data in science and technology	Palmer, S., Campbell, M.	2015	Australasian Association for Engineering Education
TA_39	An Improved Phrase-Based Approach to Annotating and Summarizing Student Course Responses	Luo, W., Liu, F., Litman, D.	2018	arXiv preprint arXiv:1805.10396
TA_40	A novel ILP framework for summarizing content with high lexical variety Student Course Responses	Luo, W., Liu, F., Liu, Z., Litman, D.	2018	Natural Language Engineering
TA_41	Making sense of student feedback using text analysis–adapting and expanding a common lexicon	Santhanam, E., Lynch, B., Jones, J.	2018	Quality Assurance in Education
TA_42	Making the student voice count: using qualitative student feedback to enhance the student experience	Shah, M., Pabel, A.	2019	Journal of Applied Research in Higher Education
TA_43	All work and no play: A text analysis	Downer, K., Wells, C., Crichton, C.	2019	International Journal of Market Research
TA_44	A Correlation Analysis of the Sentiment Analysis Scores and Numerical Ratings of the Students in the Faculty Evaluation	Lalata, J.P., Gerardo, B., Medina, R.	2019	International Conference on Artificial Intelligence and Pattern Recognition
TA_45	Using data mining to extract knowledge from student evaluation comments in undergraduate courses	Koufakou, A., Gosselin, J., Guo, D.	2016	International Joint Conference on Neural Networks (IJCNN)
TA_46	Beyond Average: Contemporary statistical techniques for analysing student evaluations of teaching	Kitto, K., Williams, C., Alderman, L.	2019	Assessment and Evaluation in Higher Education
TA_47	Text Mining Analysis of Teaching Evaluation Questionnaires for the Selection of Outstanding Teaching Faculty Members	Tseng, C., Chou, J., Tsai, Y.	2018	IEEE Access
TA_48	Text analytics approach to extract course improvement suggestions from students' feedback	Gottipati, S., Shankararaman, V., Lin, J. R.	2018	Research and Practice in Technology Enhanced Learning
TA_49	Using Opinion Mining in Student Assessments to Improve Teaching Quality in Universities	de Oliveira, A.B., Alves, A.L.F., de Souza Baptista, C.	2019	International Conference on Intelligent Systems Design and Applications
TA_50	Sentiment analysis in teaching evaluations using sentiment phrase pattern matching (SPPM) based on association mining	Pong-Inwong, C., Songpan, W.	2019	International Journal of Machine Learning and Cybernetics
TA_51	SUFAT-An Analytics Tool for Gaining Insights from Student Feedback Comments	Pyasi, S., Gottipati, S., Shankararaman, V.	2018	IEEE Frontiers in Education Conference (FIE)
TA_52	Deep learning versus traditional classifiers on vietnamese students' feedback corpus	Nguyen, P.X., Hong, T.T., Van Nguyen, K., Nguyen, N.L.	2018	NAFOSTED Conference on Information and Computer Science (NICS)
TA_53	Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation	Sindhu, I., Daudpota, S.M., Badar, K., Bakhtyar, M., Baber, J., Nurunnabi, M.	2019	IEEE Access

Continued...

SN	Title	Authors	Year	Scopus Source title
TA_54	Annotating opinions and opinion targets in student course feedback	Chathuranga, J., Ediriweera, S., Hasantha, R., Munasinghe, P., Ranathunga, S.	2018	Language Resources and Evaluation (LREC)
TA_55	Translating learning into numbers: A generic framework for learning analytics	Greller, W., Drachsler, H.	2012	Journal of Educational Technology and Society
TA_56	Learning dashboards: an overview and future research opportunities	Verbert, K., Govaerts, S., Duval, E., Santos, J.L., and Van Assche, F., Parra, G., Klerkx, J.	2014	Personal and Ubiquitous Computing
TA_57	Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms	Holstein, K., McLaren, B.M., Aleven, V.	2018	International conference on artificial intelligence in education
TA_58	Understanding learning at a glance: A systematic literature review of learning dashboards	Schwendimann, B., Rodriguez-Triana, M., Vozniuk, A., Prieto, L., Shirvani Boroujeni, M., Holzer, A., Gillet, D., Dillenbourg, P.	2016	International Conference on Learning Analytics and Knowledge (LAK)
TA_59	Intelligent Tutors as Teachers' Aides: Exploring Teacher needs for Real-time Analytics in Blended Classrooms	Holstein, K., McLaren, B.M., Aleven, V.	2017	International Learning Analytics and Knowledge Conference
TA_60	Towards a teacher dashboard design for interactive simulations	Tavares, D.L., Perkins, K., Kauzmann, M., Velez, C.A.	2019	Journal of Physics: Conference Series
TA_61	Cross-disciplinary Participatory and Contextual Design Research: Creating a Teacher Dashboard Application	Abel, T.D., Evans, M.A.	2013	Interaction Design and Architecture(s) Journal
TA_62	Attention please! Learning analytics for visualization and recommendation	Duval, E.	2011	International Conference on Learning Analytics and Knowledge
TA_63	A Qualitative Evaluation of Evolution of a Learning Analytics Tool	Ali, L., Hatala, M., Gašević, D., Jovanović, J.	2012	Computers and Education

End of Table

Appendix B

QUESTIONNAIRE ONE

Presentation of teaching evaluation data

Dear participant (potential),

I am conducting a study to provide more effective ways to present the results of students teaching evaluation to improve how the teachers could use that information.

This project has ethical approval from the University of Otago with reference number 19/097.

You are invited to spare 20-30 minutes of your valuable time to help with this study and contribute to how the presentation of students teaching evaluation data is carried out. You are eligible to participate if you are an academic.

Providing information through this online survey is taken as an indication of voluntary consent to participate.

1. Do you collect teaching evaluation data?

- ◯ Yes
- ◯ No

Presentation of teaching evaluation data

2. What type of teaching evaluation data are you currently collecting (choose all the apply)?

- ☐ Student teaching evaluation data
- ☐ Peer observation
- ☐ Informal discussions with colleagues
- ☐ (Other) ____

Briefly describe the reasons for your choice:

____

3. In what format(s) do you get your teaching evaluation data (choose all that apply)?

- ☐ In raw data
- ☐ Summary statistics (tables, graphs, proportions, etc.)
- ☐ Textual form
- ☐ (Other) ____

4. How do you rate student evaluation data as a means of improving your teaching?

Very high	high	Neutral	Low	Very Low
○	○	○	○	○

5. How often do you evaluate your teaching?

○ Always

○ Often

○ Sometimes

○ Rarely

○ Never

Briefly describe the reasons for your choice:

____

6. List up to five digital tools you use, if any, when interacting with teaching evaluation data.

	Check
Words	☐
PDF	☐
Excel	☐
Dashboards	☐
SPSS	☐
Weka	☐
Tableau	☐
Power BI	☐
Splunk	☐
SAP	☐
SQL	☐
Python	☐

Using teaching evaluation data

7. How often do you use the results of teaching evaluation to inform your teaching?

- ○ Always
- ○ Often
- ○ Sometimes
- ○ Rarely
- ○ Never

Briefly describe the reasons for your choice:

____

8. When was the last time you used teaching evaluation data to improve your teaching?

- ○ Last week
- ○ Last month
- ○ Last semester
- ○ Last year
- ○ Never

Briefly describe the reasons for your choice:

____

9. What do you think is the most important reasons for using teaching evaluation?

- for improving teaching outcomes
- for promotion
- for learning about teaching
- (Other) _______________

Briefly describe the reasons for your ranking:

____

10. How confident are you in interpreting teaching evaluation data?

Very confident	Confident	Neutral	Slightly confident	Not confident
○	○	○	○	○

11. How likely would you require the support of others to help with the interpretation of teaching evaluation data?

Very Likely	Likely	Neutral	Unlikely	Very Unlikely
○	○	○	○	○

Prior Knowledge about teaching evaluation dashboards

12. Which of the following best describes how much you know about the use of dashboard for presenting data for teachers?

○ I don't know anything about teacher evaluation dashboard

○ I know a little about teacher evaluation dashboard, but I could learn more

○ I am an expert in the use of teacher evaluation dashboard

Briefly describe the reasons for your choice:

____

13. How likely would you use teacher evaluation dashboard?

Very Likely	Likely	Neutral	Unlikely	Very Unlikely
○	○	○	○	○

Briefly describe the reasons for your choice:

____

Demographic Information

14. Rank

○ Professor

○ Associate Professor

○ Senior Lecturer

○ Lecturer

○ Senior Teaching Fellow

○ Teaching Fellow

○ (Other) ________________________________

15. Division

○ Commerce

○ Health Sciences

○ Humanities

○ Sciences

○ (Other) __________________________________

16. Age

- ○ Under 25 years
- ○ 26 - 30 years
- ○ 31 - 40 years
- ○ 41 - 50 years
- ○ 51 - 65 years
- ○ Above 66 years

17. How many years have you been teaching altogether (full or part-time, counting this year)?

- ○ 0 - 5 years
- ○ 5 - 10 years
- ○ 10 - 15 years
- ○ 15 - 20 years
- ○ More than 20 years

18. Do you have any other comments related to the interpretation of teaching evaluation data?

____

Thank You! You have reached the end of the questionnaire!

Please leave your ***Email Address*** *in the text box below, if you would be interested in participating in evaluating software that incorporates a dashboard for presenting teaching evaluation data.*

____

Appendix C

ETHICS ACKNOWLEDGEMENT LETTER

UNIVERSITY of OTAGO
Te Whare Wānanga o Otāgo
NEW ZEALAND

Academic Services
Manager, Academic Committees, Mr Gary Witte

19/097

16 August 2019

Assoc. Prof. B Daniel
Higher Education Development Centre

Dear Assoc. Prof. Daniel,

I am writing to let you know that, at its recent meeting, the Ethics Committee considered your proposal entitled “**Developing Teacher Analytics and Teacher Dashboard to Support EvidenceBased Teaching**”.

As a result of that consideration, the current status of your proposal is:- **Approved**

For your future reference, the Ethics Committee’s reference code for this project is:- **19/097**.

The comments and views expressed by the Ethics Committee concerning your proposal are as follows:-

While approving the application, the Committee would be grateful if you would respond to the following:

Data storage

Please note that it is the responsibility of the Principal Investigator, rather than the student, to destroy the data.

Information Sheet

The Committee suggests, for clarity, that the reference to the evaluation in paragraph 2 is reworded to “teaching evaluation by students”, to avoid any potential confusion.

Information Sheet and Consent Form

Reference is made to an open-questioning technique on both the Information Sheet and Consent Form. Please clarify whether this is an error. If not, please provide the Committee with more detail on the types of questions to be asked.

Please provide the Committee with copies of the updated documents, if changes have been necessary.

Approval period: Approval is for up to three years from the date of this letter. If this project has not been completed within three years from the date of this letter, re-approval must be requested. If the nature, consent, location, procedures or personnel of your approved application change, please advise me in writing.

Conditions of approval: Upon approval, it is expected that all members of the research team are made aware of what the standard conditions of ethical approval covers. This includes the date ethical approval expires, as well as the process regarding applying for amendments to the research.

Final Report: The Human Ethics Committee asks for a Final Report to be provided upon completion of the study. The Final Report template can be found on the Human Ethics Web Page

http://www.otago.ac.nz/council/committees/committees/HumanEthicsCommittees.html

Locality authorisation: Studies requiring locality authorisation, i.e. permission from the organisations at which the study is taking place or from which participants are being accessed, must be confirmed before the study commences.

Yours sincerely,

Gary Witte

Mr Gary Witte
Manager, Academic Committees
Tel: 479 8256
Email: gary.witte@otago.ac.nz

c.c. Assoc. Prof. B Daniel Higher Education Development Centre

Appendix D

NON-UNIFORM RANDOM PROBABILITY DISTRIBUTION TECHNIQUE

Table D.1 Non-uniform random distributions for points on specific programs and papers.

	PAPERS	CSC 101	MTH 101	CHM101	BIO101
PROGRAMS					
Biochemistry	Points:	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
	Seed:	50, 35, 13, 1, 1	30, 45, 13, 4, 8	1, 1, 13, 35, 50	1, 1, 13, 40, 45
Computer Science	Points:	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
	Seed:	1, 1, 13, 40, 45	4, 8, 13, 45, 30	5, 20, 40, 20, 15	30, 45, 13, 8, 4
Medicine	Points:	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
	Seed:	30, 40, 15, 10, 5	4, 8, 13, 45, 30	5, 20, 40, 20, 15	1, 1, 13, 35, 50
Information Science	Points:	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
	Seed:	1, 1, 13, 40, 45	30, 45, 13, 8, 4	1, 1, 13, 40, 45	5, 20, 40, 20, 15
Pharmacy	Points:	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
	Seed:	30, 45, 13, 8, 4	30, 45, 13, 8, 4	4, 8, 13, 45, 30	4, 8, 13, 45, 30
Accounting	Points:	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
	Seed:	null	null	null	null
Dental Technology	Points:	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
	Seed:	null	null	null	null
Education	Points:	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
	Seed:	null	null	null	null
Genetics	Points:	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
	Seed:	null	null	null	null
Health Informatics	Points:	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
	Seed:	null	null	null	null
Physics	Points:	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5	1, 2, 3, 4, 5
	Seed	null	null	null	null

Table D.1 illustrates the various seeds that were used to generate points for various programs, including Biochemistry, Computer Science, Medicine, Information Science, Pharmacy, and for several papers, such as CSC101, MTH101, CHM101 and BIO101. It includes other programmes, such as Accounting, Dental Technology, Education, Genomics, Health Information, Physics that were generated randomly without any seed or specific order.

Table D.2 Non-uniform random distributions for points on several programs and papers for each question.

PROGRAMS	PAPERS	CSC101	QUESTIONS
Accounting:	Point:	1, 2, 3, 4, 5	Q1
	Seed:	30, 45, 13, 8, 4	
Dental Technology:	Point:	1, 2, 3, 4, 5	Q2
	Seed:	45, 40, 13, 1, 1	
Education:	Point:	1, 2, 3, 4, 5	Q3
	Seed:	40, 35, 13, 6, 6	
Genetics:	Point:	1, 2, 3, 4, 5	Q4
	Seed:	30, 30, 20, 10, 10	
Health Informatics:	Point:	1, 2, 3, 4, 5	Q5
	Seed:	30, 40, 20, 5, 5	

Table D.2 illustrates sample seeds used to generate the points for Q1 to Q5 for CSC101 paper for several programs, including Accounting, Dental Technology, Education, Genetics and Health Informatics.

Table D.3 Standard Evaluation Questions used in the University

ID	QUESTIONS
Q1	How organized have you found Dr Spock's contribution to this course?
Q2	How would you rate Dr Spock's ability to communicate ideas and information?
Q3	How much has *Dr Spock* stimulated your interest in the subject?
Q4	How would you describe Dr Spock's attitude toward students in this course?
Q5	Overall, how effective have you found Dr Spock in teaching this course?

Table D.3 illustrates the five standard questions ranging from Q1 to Q5 that are used by the University of Otago to measure teaching quality for CSC101 paper considering several program.

Appendix E

THE PROCESS MODE FOR SET DATA SIMULATION

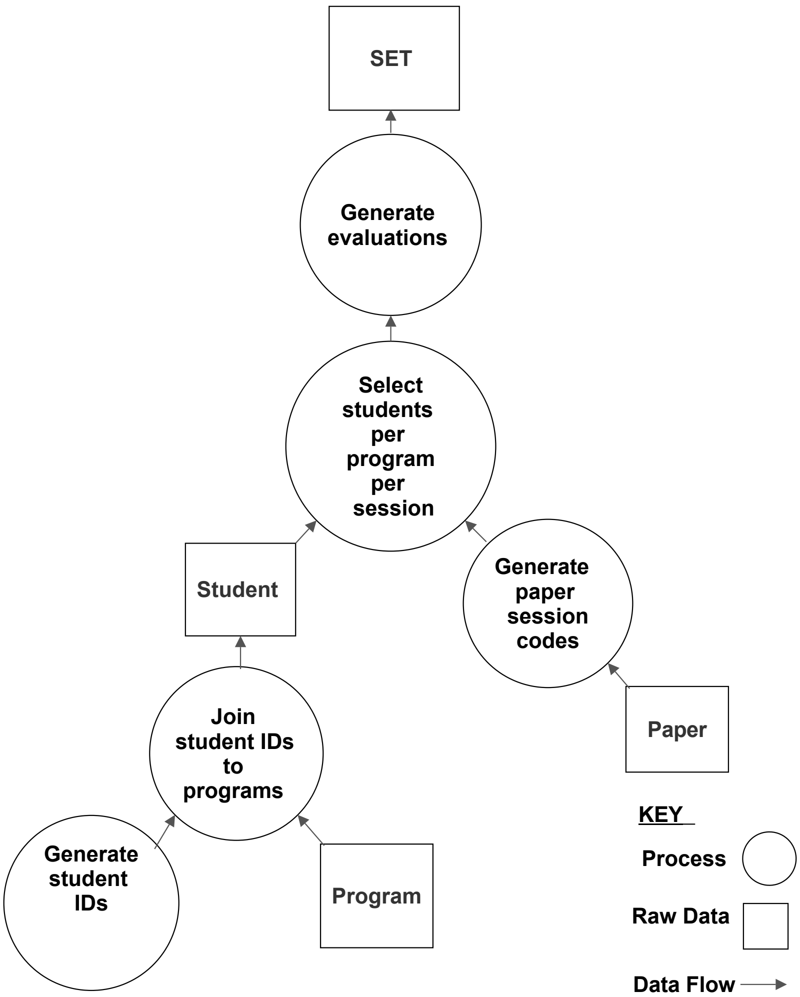


Fig. E.1 Process Diagram for generation of SET data

Figure E.1 illustrates the process model of how the simulated SET data was generated. The *Generate student IDs* process randomly generates student Identities (IDs). For this study, 1000 random student IDs per year, starting from the year 2013 to 2017 were generated. The *programs* data are the various programs run in the institution. The eleven programs used in this simulation include Biochemistry, Computer Science, Medicine, Information Science, Pharmacy, Accounting, Dental Technology, Education, Genomics, Health Information and Physics.

Next, the *Join Student IDs to Programs* process tries to map the student IDs with the various programs to generate *student* data. This mapping method is random and in no particular order. The *paper* data represents the course codes; six papers, including CSC101, MTH101, CHM 101, BIO101, GEO101 and PHY101 were used for this simulation.

The *Generate Papers Session Code* process creates a unique code for each paper. For example, the unique paper code OU_CSC101_2013_2014 indicates that the paper CSC101 was taught in a 2013/2014 session at the University of Otago. For this simulation, six papers per session produced a total of thirty papers for five sessions, starting from 2013 to 2017.

Then, *Select students per program per session* process randomly selects student IDs per program per paper. The selection process is such that an upper bound and a lower bound can be set, and the number of students that can be selected will not exceed the set boundary limits per program per session. For our simulated data, the boundary limits were set with 3000 as the lower bound and 5000 as the upper bound.

Finally, *Generate Student Evaluations* process forms the SET data. This process generates points ranging from 1 to 5 using a non-uniform random probability sampling technique, such that the rate at which each point can occur can be determined by setting the seed. This seed represents the percentage at which each point should occur per program per paper per session for each question. Also, some assumptions were made with regards to the timestamp generated for this data simulation. The first assumption is that evaluation is carried out once, and the second assumption is that the evaluation will occur only within the month of September. Hence, the timestamps that were generated for our dataset were timestamps randomly generated within the month of September for each year. In general, a total of 224,881 SET data records was generated for this experiment.

Appendix F

UML MODEL FOR GENERATING THE SIMULATED SET DATA

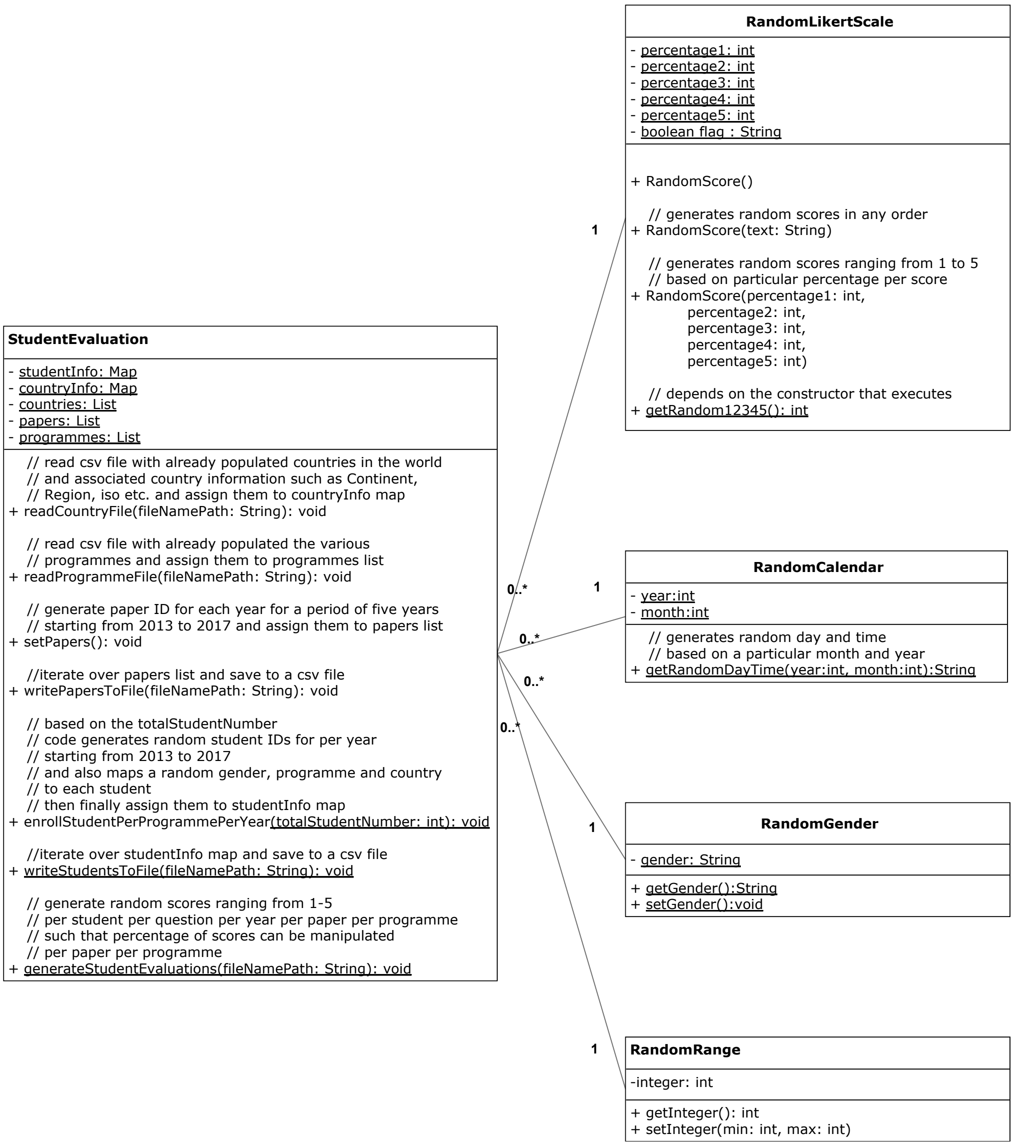


Fig. F.1 UML Diagram of Classes that Simulated Evaluation Data

Figure F.1 is a Unified Modeling Language (UML) diagram that describes the Java programme used to generate the SET dataset. It consists of five major classes, as follows; *StudentEvaluation, RandomScore, RandomCalendar, RandomGender, RandomRange*. In this section, the functionalities of the two classes *RandomLikertScale* and *StudentEvaluation* were explained. Other classes are self-explanatory; as their meanings can be derived from class names.

The *RandomLikertScale* class functions as a seed and generates points ranging from 1 to 5. This class is quite flexible, and the percentage probability at which a particular value occurs can be determined. The *RandomScore()* is an overloaded constructor, and the number of parameters the constructor holds during instantiation determines which of the overloaded constructor executes. The one parameter overloaded constructor, *RandomScore(text: String)*, generates random points ranging from 1 to 5, and the constructor determines the rate of occurrence. In contrast, the five parameters overloaded constructor, *RandomScore(percentage1: int, percentage2: int, percentage3: int, percentage4: int, and percentage5: int)* are passed five different integer values (ranging from 1 to 100) as arguments during instantiation. Each of these values maps to a particular point and determines the percentage at which that point will occur.

Secondly, the *StudentEvaluation* class encapsulates several variables, including *studentInfo, countryInfo, countries, papers and programs*. It also holds several methods, such as *readCountryFile(), readProgrammeFile(), setPapers(), writePapersToFile(), enrollStudentPerProgrammePerYear(), writeStudentsToFile() and generateStudentEvaluations()*.

The *readCountryFile()* and *readProgrammesFile()* are functions that read countries and programs from separate external Excel files. Both methods have the *fileNamePath* parameter that holds file paths as the argument. The file paths point to the files where the various countries the programs would be read from into the method.

The *enrollStudentPerProgramPerYear()* function generates student IDs. This method has only one parameter, *totalStudentNumber*, and is passed the total number of students as an argument. It has a principal function that estimates and generates random values of student IDs per program per year.

The *setPaper()* method has two parameters: *paperids* and *years*. This method performs an element-wise concatenation on the two list objects that are passed to it as arguments, which then returns a new array list with unique paper IDs.

The *writePapersToFile()* has one parameter named *fileNamePath*, which accepts the file path as an argument. This argument points to the Excel file where the distinct generated paper ids will be written. Also, the *writeStudentsToFile()* saves into an Excel file individual student information generated as a result of the *enrollStudentPerProgramPerYear()* method call. A student information record contains a randomly generated unique student ID, along with other information such as a randomly assigned timestamp, gender, program, and country. This method also has one parameter, *fileNamePath*. When this method is invoked, the file path, where the student information will be written, is passed to it as an argument.

The generateStudentEvaluations() function generates a random point between 1 and 5. These Likert scales are repeatedly generated for each question (Q1, Q2, Q3, Q4, and Q5) per student. This method has only one parameter, 'fileNamePath'; the file path is passed as an argument to this method during the execution of the programme.

Appendix G

PROTOCOL USED FOR THE USABILITY STUDY

The goal of this study is to determine the usability and usefulness of a Teacher's Evaluation Dashboard (TED). The intent is to help determine how valuable this dashboard is to the teachers. Hence, this usability study aims to answer the design-related questions: Is the content presented in a way that is easy to understand? Is the dashboard useful?

To provide us with a more accessible design insight and opportunities to probe and ask for clarifications, as well as get open-ended comments from participants, we used an In-person/Moderated Remote Usability Testing approach.

During the usability session, a short tutorial will be given to the participant on how to use the quantitative part of the dashboard. Then the participant will be allowed to perform three main tasks on the dashboard. First, is the aggregate task: the participant will be allowed to perform and recognise different aggregates presented on the Number-ratings dashboard. Second is the comparative task: the participant will be asked to perform and recognise some comparisons tasks on the Number-ratings dashboard. The third is a collaborative task: the researcher and participants will collectively explore four different open-ended comments dashboards. The whole usability session will be recorded using Zoom.

Task 1: Aggregate

CSC101 is a large class that attracts students from diverse programmes. You happen to be the instructor that handled this class from 2013 up to 2017.

Task 1a: how many students from Computer Science responded to CSC101 paper, and how did they rate the teacher's performance?

Task 1b: how many students from Health Informatics responded to CSC101 paper, and how did they rate the teacher's performance?

Task 1c: how many students from Health Informatics responded to Q1 of CSC101 paper, how did they rate the teacher's performance?

Task 1d: how many students from Health Informatics responded to Q1 of CSC101 paper in the year 2013, and how did they rate the teacher's performance?

Task 2: Comparison

How did the students from the Health Informatics program rate the teaching performance, compared with Students from Computer Science program, in CSC101 paper, in Q5 from the year 2013 through 2017?

Task 3: Collaborative Task

Can we explore text analytics of the open-ended comment dashboard to see the various words and phrases, named entities, clusters and sentiments?

A short interview will be performed after every main task to know what the participants believe about the usefulness of the dashboard. Two fundamental questions will be asked, as follows: 1. Do you think this dashboard will be useful to you as a teacher? 2. How do you think this dashboard could be further improved?

Finally, the participant will be asked to fill a ten minutes exit usability questionnaire to determine how easy the dashboard is to use.

Appendix H

THE ORIGINAL AND MODIFIED SUS

No.	Original Question	Modified Question
1	I think that I would like to use this system.	I think that I would like to use this dashboard frequently.
2	I feel this feature is too complicated even though it can be made simpler.	I found the dashboard unnecessarily complex.
3	I think this feature is easy to use.	I thought the dashboard was easy to use.
4	I think I need help from a technical person to be able to use this feature.	I think that I would need the support of a technical person to be able to use this dashboard.
5	I find that there are various kinds of features that are well integrated into the system.	I found the various functions in this dashboard were well integrated.
6	I think a lot of inconsistencies are found in this feature.	I thought there was too much inconsistency in this dashboard.
7	I think the majority of users will be able to learn this feature quickly.	I would imagine that most academics would learn to use this dashboard very quickly.
8	I find that this feature is very impractical when used.	I found the dashboard very cumbersome to use.
9	I am very sure I can use this feature.	I felt very confident using the dashboard.
10	I have to learn many things first before I can use this feature.	I needed to learn a lot of things before I could get going with this dashboard.

Appendix I

THEMATIC ANALYSIS OF THE FREQUENCY OF PERFORMING TEACHING EVALUATION

Theme	Responses	Interpretations	Example of Quotations	Partici-pants
University Requirement and Standard Practice	Always(p1, p5, p34, p38, p39, p51, p48); Rarely(p42, p45)	Making SET a compulsory requirement and practice could lead to a situation of academics preempting what the outcome of the ratings will be. The negative effect of this kind of prediction may not be good enough to motivate change in instructional improvement, both for lowrated and high-rated instructors.	it is part of the programme in my department.	p1
			Done automatically where I work. Not my choice.	p5
			The student teacher evaluations are compulsory so not really a choice. i ask for anonymous feedback (paper and pencil) from the students who attend lectures throughout the semester.	p34
			Obligation.	p38
			In terms of the formal evaluations, i collect those only when i must (as required by the university).	p39
			I do it because i have to, i already have a good idea of what they outcome will be.	p42
			I have been told not to ”over sample” the students, so not to obtain ratings every semester.	p45
			My hod requires me to get student evaluations.	p48
			I use student evaluations when required by the university's system of review.	p51
Frequency of Evaluation	Always(p6, p20, p21, p28, p37, p41, p57); Rarely(p32, 49)	Participants that frequently perform evaluations have developed their own culture and frequency patterns that work for them. While participants also expressed awareness that suggests that survey fatigue, could suppress response rates. One respondent that did not agree with frequent evaluations argued evaluating other teaching components other than teaching practice.	I feel that this is appropriate; however, the type of evaluation I do varies in frequency.	p6
			I evaluate at the end of the semester. students already have 2 class rep meetings during the semester where they can provide feedback, so doing it more often seems like overkill.	p20
			I often experiment with new ideas in the classroom so have i times chosen not to evaluate, but i almost always do.	p21
			always try to do different things, so always need new evaluations.	p28
			I try to evaluate innovations frequently and established courses at least every 3 years.	p32
			I try to evaluate 1 to 2 blocks of teaching each year, bearing in mind that students are overwhelmed with all of their lecturers seeking same feedback.	p37
			Good to do it every year so don't get complacent.	p41
			Each paper has to be evaluated at least once every third time it is taught. I don't do it more often than necessary because I don't think the information is very valuable.	p49

Continued...

Theme	Responses	Interpretations	Example of Quotations	Partici-pants
			I evaluate my teaching every time I teach. I use different forms of data on different occasions and at different frequencies.	p57
Students Feedback	Always(p8, p19, p25, p41, p47, p50, p52, p53, p58)	It is important to note that a number of participants that perform frequent evaluations have expressed that they have many opportunities to gather informal feedback on the learning of their students, and therefore formal evaluation data have been viewed as less important. This sort of informal feedback from students has also contributed immensely to course content development.	I regularly evaluate my teaching to ensure i get feedback on the learning experience in order to act on any changes that might be needed, or to explain to students why we run sessions in a particular way (that might have been criticised). it is important for me to let students know their feedback matters and is considered [p8].	p58
			There is always room for improvements things to learn from the students. a teaching event may have been successful in the past but for a variety of reasons may no longer to effective.	p19
			I believe it is important to be in touch with the student experience. however, I also relie on informal conversations with students, and a journal assessment to monitor how things are going.	p25
			Good to do it every year so don't get complacent. also i ask questions (eg which topic was your favourite?) that help me optimise the syllabus for the following year.	p41
			Feedback is useful.	p47
			Important to get student feedback.	p50
			I respect the student voice and the multiple forms of formal and informal exchanges that take place.	p52
			Variety of reasons giving students opportunities to express views/feedback,to be aware of any reactions to changes in course content, assessment or teaching approaches, to take into account differences in class profiles (e.g., change in mix of domestic and international students; background of students (i am in education so can be very diverse).	p53
			It's important to gather student feedback regularly.	p58
Promotion and Progression	Always(p10, p12, p13, p33, p43, p54)	Participants have expressed promotion and progression as one of the main reasons why they perform evaluations.	this data is needed for promotion.	p10
			I am required to do so for confirmation.	p12
			Evidence for continued employment.	p13
			We are required to do so for confirmation, progress and promotion.	p43
			We have to evaluate it at the end of each teaching period. i then look at it every 5 years or so when i apply for promotion etc.	p54
Improve Teaching	Always(p3, p9, p16, p31, p51, p56); Rarely(p23)	Considering that SET results are used to judge academics on their performance and have a direct impact on their careers. Participants also recognised that carrying out frequent SET contributes to improving teaching practice.	it is good to have feedback for improvement.	p3
			I am constantly tweaking my workshops and student evaluation data is one way of getting some feedback to see if these changes are having the desired effect.	p9
			I view it as unethical to not constantly find ways to improve at everything i do, given that this is my paid profession and students (and taxpayers) have a great deal at stake in the system. Again, we demand endless research review (which contributes little to society), but we place very little emphasis on teaching, beyond the nefarious system that punishes divisions and programmes for "excessively low" pass rates.	p16
			I am constantly making changes to my teaching. sometimes it is in response to prior evaluations and other times it is due to something that I may have read in the education literature. I like to have feedback on whether the changes i have made have had the desired/anticipated effect on students.	p31
			It's onerous and I don't value the information gleaned from them. pass rates are a better guide.	p23
			Quality of the evaluation questions is very poor.	p33

Continued...

Theme	Responses	Interpretations	Example of Quotations	Partici-pants
			Due to learning and teaching can be different in each class experience. i need to continually reflect on what has occurred and evaluate if learning has occurred or not and act or react accordingly.	p51
			I believe that the qualitative comments students provide, give a great insight, and points for improvement in learning and teaching.	p56
Poor Response	Always(p48); Rarely(p44, p49)	Respondents who did not agree with frequent evaluations believe that this issue of low response rate could affect teachers into giving high grades and less course work to win the approval of the few students that respond to the survey.	Response rates are low, the surveys are somewhat limited and what you can find out, and there is more to good teaching than good evaluations so other things need to be considered.	p44
			Student rates of filling out feedback forms are very low and they often give contradictory feedback.	p48
			Many students don't complete them, my results are also quite consistent, so they tell me even less.	p49

End of Table

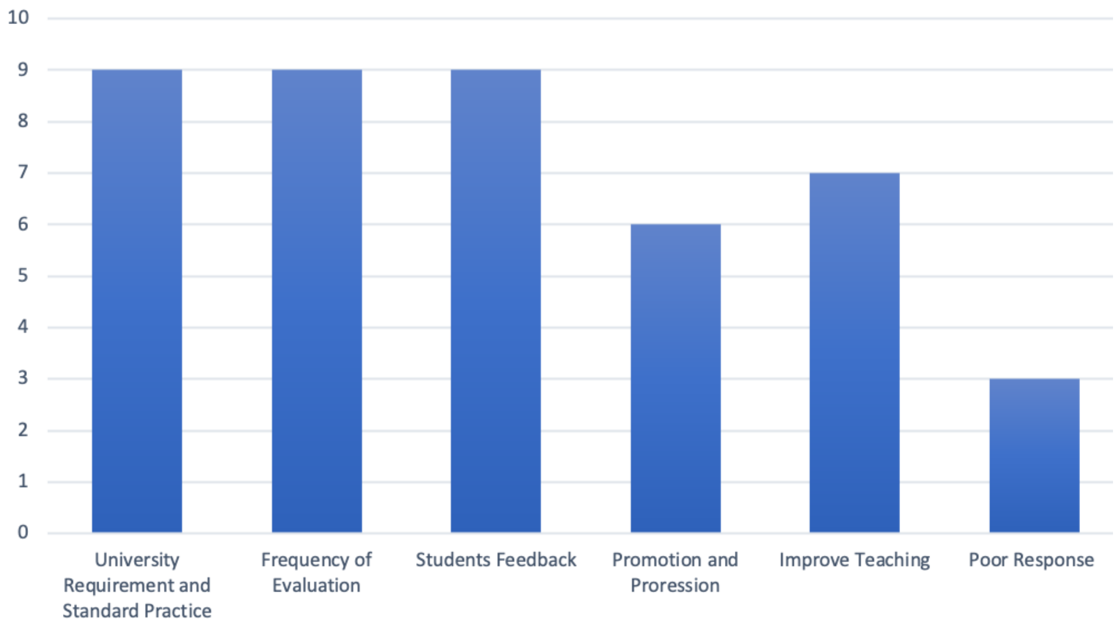


Fig. I.1 Frequency Table of themes identified from how often participants perform teaching evaluation data.

Appendix J

THEMATIC ANALYSIS OF THE USE OF TEACHING EVALUATION TO IMPROVE TEACHING

Theme	Responses	Interpretations	Example of Quotations	Partici-pant
University Requirement and Standard Practice	Always(p16, p50); Never(p15)	There is a widespread view that academics are hostile to evaluations, despite several studies researchers have conducted over the last few decades, demonstrating their validity and reliability.	To be honest, in the online form I now view the student evaluations as more akin to online trolling than anything else. Although we are required to do them, a couple of years ago I made an active decision never ever to look at them. This decision has made me happier, and far more confident in my teaching. I seek personal, face to face feedback, feedback through class reps, and assess my teaching success through students' results all these are useful.	p15
			Why collect it if I'm not going to use it?	p16
			No use gathering it if I don't use it!	p50
Students Feedback	Always(p1, p19, p20, p31, p41, p58); Rarely(p5, p18, p37, p43, p46); Never(p54)	Positive views about students ability to make judgements about teaching were revealed severally, although some negative views were also expressed especially with regards to students open responses.	I always take on board student feedback that can improve the delivery of content in my teaching [p1].	p1
			I can only learn from qualitative comments and students don't always make them.	p5
			Constructive criticism or suggestions are rare.	p18
			There is always something in the feedback I can learn from. Even if it simply explaining to students more clearly why things are done a certain why.	p19
			The feedback provided by students is not detailed enough to inform all of my teaching practices, but I use it where I can.	p20
			As indicated previously, I use evaluations to gauge student opinion about my teaching. Based on this I adjust what I do in the classroom. It is a continuous and iterative process.	p31
			Sometimes students note useful things that help me reflect on my teaching and how to improve.	p37
			prevents complacency and keeps me teaching topics that students are interested in.	p41
			If just one person makes a comment, I don't feel I need to change, but if I get multiple people giving the same comment, I think about how to deal with it.	p43
			Occasionally students will provide meaningful qualitative responses and I incorporate in the next iteration of the course.	p46
			Because it is either a too personal reflection from students therefore useless to me OR stuff I already know and are working on improving or changing.	p54
			Every student cohort giveS useful feedback.	p58
Improve Teaching	Always(p1, p3, p8, p9, p10, p13, p26, p28, p32, p33, p34, p44, p51, p57); Rarely(p12, p39); Never(p23)	Despite demonstrating the validity and reliability of the frequent use SET data in areas such as course content development.	I also rely on other feedback to continually enhance my teaching practices.	p1
			The data help me to revise the course for the next run.	p3
			As a reflective practitioner, gaining feedback (and responding to it) is a core part of my teaching practice.	p8
			After running a workshop a number of times and I know that it is working well, I might not evaluate it again. I am mindful not to ask students to evaluate everything just for the sake of It.	p9

Continued...

Theme	Responses	Interpretations	Example of Quotations	Partici-pant
			I don't see any point in evaluating, if it doesn't inform what I do next!	p10
			The data is usually useless. Occasionally there is something I can use, but this is rare.	p12
			Particularly the comments are useful for alerting me to elements of the paper that are not working as well as they could (particularly applies to distance students).	p13
			as above, they provide little of use, have had a tendency to become niggardly since going on line and do not reflect actual achievements in terms of passing the papers.	p23
			I value improvement in teaching ability.	p26
			what's the point of evaluation, if not to inform your teaching practices. Problem is, most teachers use just for promotion.	p28
			One of the most important things in teaching is the quality improvement cycle.	p32
			Explained in previous question. Little option of alternative means. About to develop my own.	p33
			Not always because feedback is often contradictory. e.g. some list labs as their favourite part of the paper, others list them as their least favourite.	p34
			I always read the results and consider them, but it is relatively rare that they provide something useful that I can implement to improve my teaching.	p39
			Where I get helpful feedback I try to incorporate it. I often try and collect informal feedback from classes although this tends to be more about content than the quality of my teaching.	p44
			The results allows me to determine if my current actions are creating an environment of significant learning that will meet the intended learning outcomes. The results will alter my actions accordingly to meet this outcome.	p51
			The fundamental purpose of me evaluating my teaching is so that I can improve. I always evaluate my teaching in one way or another.	p57
Poor Response	Always(p17); Rarely(p21, p37)	Student poor response rate may result to low-quality data; this is still an issue that discourages participants from frequently engaging with SET data. Data fusion to link students anonymously to their attendance, grades, activities and SET is a possible solution that could be implemented to assist the teacher in knowing if students that filled the evaluation attended the classes as well their activities and performances outcome.	With the low participation numbers in recent years the value has been reduced for the HEDC formal surveys and so I have now more emphasis on informal surveys of students.	p17
			It depends on what type of evaluation. I consider all the feedback, but some of it I find more useful than others. I do not trust the student evaluations as much as other forms of evaluations. This is because there is often such a low response rate, especially after going to online only versions. The low response rate seems to polarise the evals and there is no way to know whether students filling them out have actually attended class.	p21
			However low response rates mean that you don't get a good idea of what students as a whole find helpful this way.	p37
Data Quality	Always(p49); Rarely(p45)	However, strong reservations were also voiced concerning the quality of student feedback, with a number of participants questioning the quality of student judgements at the Institution.	It depends on the quality of the data. If they just say that they liked/did not like the course, there is no real feedback to use to implement changes. They need to provide some constructive feedback, such as "I particularly liked x topic/exercise" or "I struggled with z" so then I know which aspects of the course work/do not work.	p45
			I always consider the responses, but comments such as 'this paper has too much reading' are not going to compel me to change the reading load. I teach 300 level ENGL majors and they are not working more than 13 hrs/wk for me even with the current reading load.	p49

End of Table

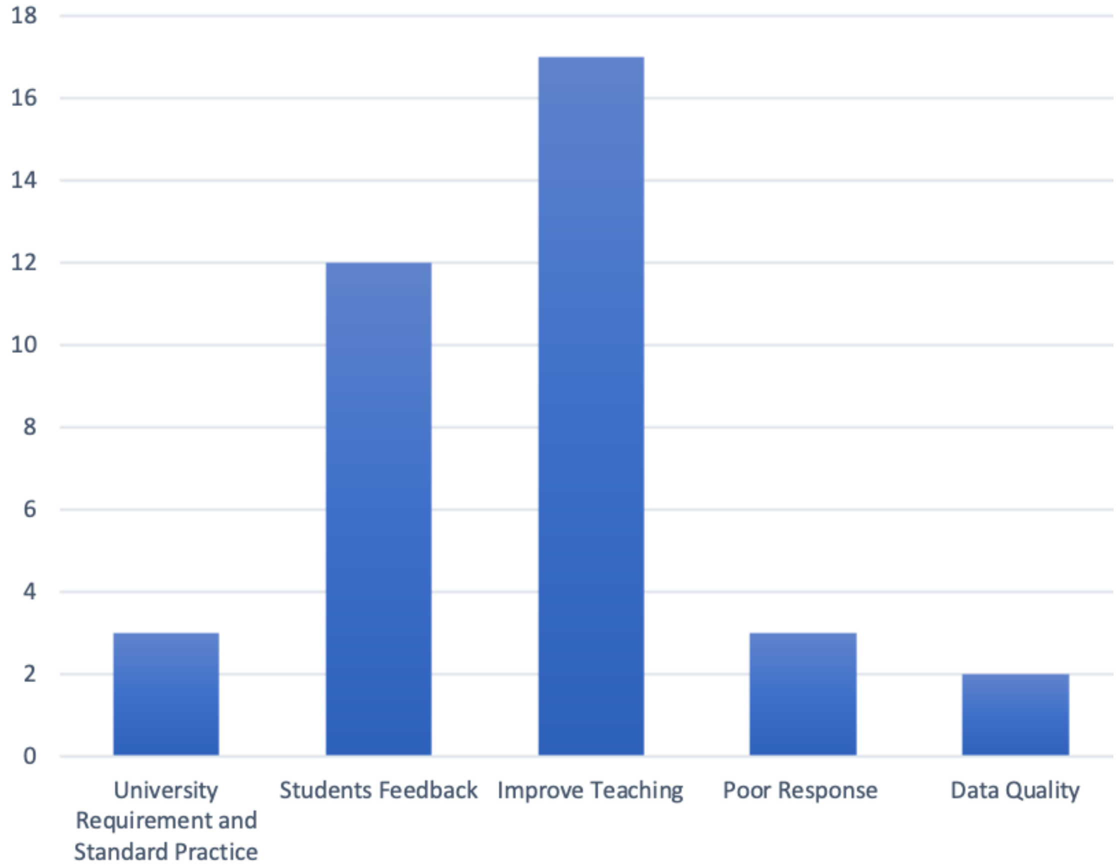


Fig. J.1 Frequency Table of themes identified from how often participants use the results of teaching evaluation to inform teaching.

Appendix K

THEMATIC ANALYSIS OF THE LAST TIME TEACHING EVALUATION WAS USED TO IMPROVE TEACHING

Theme	Responses	Interpretations	Example of Quotations	Partici-pants
Students Feedback	Recently (p5, p34, p32, p42, p47); Never (p15, p18, p54);	In order to mitigate this challenge, some participants carry out various forms of informal feedback or other forms of evaluations such as POB or IDC, rather than the formal evaluation that is performed less frequently.	There were some things I could easily change based on feedback from students.	p5
			The opinions expressed through the evaluation forms are extremely variable and contradictory so all that you find out is whatever you do, you can't please everyone.	p15
			Can't really think of any feedback that was helpful.	p18
			Teaching is a reflexive processes that is alwayschanging and responsive. change doesn't always happen as directly as implied by the question.	p21
			I just finished a block of teaching in a 300 level science course. This year, I made some changes based on prior evaluations, so I wanted to see whether and how the students responded to these changes.	p32
			Very little discussion in class, so I asked for anonymous feedback about how the paper is going.	p34
			I received feedback which wasn't great as I had adapted some things in the class, so I made further changes.	p42
			Had just heard from class reps.	p47
			Because it is either a too personal reflection from students therefore useless to me OR stuff I already know and are working on improving or changing.	p54
Improving Teaching	last year (p43); Recently (p8, p16, p19, p32, p51)	However, participants that rarely perform evaluations pointed out concerns with regards to the time SET is usually carried out; this happens to be towards the end of the course work. Hence this constraint makes it difficult to report SET results back to current students. This tradition of disconnected relationship with SET data is evident, and a strong reason not to feedback to students.	As noted above, evaluation data is integral to my teaching.	p8
			My teaching and learning circle met last week to discuss our teaching observations and collect suggestions for improvement. I have a whole list of things I'll be picking up next week when classes resume.	p16
			Currently reviewing feedback to help improvement the second half of the semester.	p19
			We did a focus group in June, i got the results back last week, I am using them to plan next years teaching.	p32
			Most of our classes are only taught yearly, so there isn't a chance to implement them until the next teaching period.	p43
			It informs my practice and its decisions.	p51

Continued...

Theme	Responses	Interpretations	Example of Quotations	Partici-pants
Course Content Development	Recently (p37, p41, p44, p55)	Participants that recently used evaluations to inform teaching claim to be interested in checking their teaching and courses experiences with students. Most of them were interested in ideas to fine-tune their courses towards improving the experiences of students, p44.	I went back to previous years evaluations to inform changes to lectures for this year [p37].	p37
			Eval data from 2018 led to a syllabus change in July 2019.	p41
			I think about a course each time I teach it and wonder consider way to improve the content, make it more relevant, and think about better ways of communicating what I would like students to think about (and to find out their expectations of the course).	p44
			I modify all of my lectures and classes before I deliver them each semester.	p55

End of Table

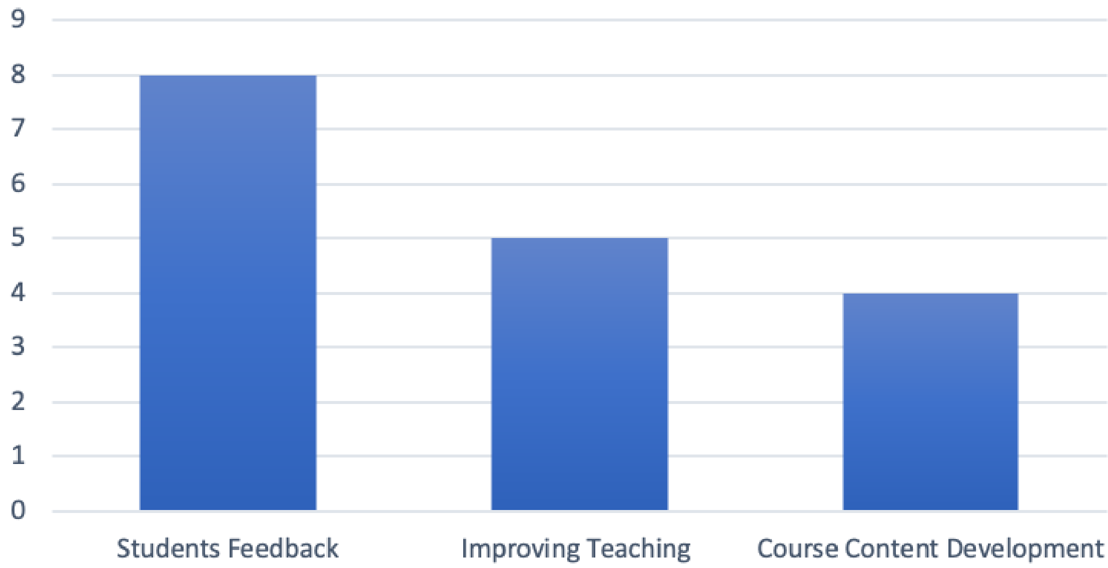


Fig. K.1 Frequency Table of themes identified from the last time participants used teaching evaluation data to improve teaching.

Appendix L

THEMATIC ANALYSIS OF THE TYPES OF TEACHING EVALUATION DATA COLLECTED

Theme	Responses	Interpretation	Example of Quotations	Partici-pants
University Requirement and Standard Practice	SET(p5, p34, p38, p39, p45, p54, p58); SET+IDC(p15, p23, p36, p43, p44, p48, p56); All (p12, p55)	University requirement and standard practices towards externally controlled SET in institutions may affect the attitudes, engagement and responses of academics to evaluations.	I don't need to ask for it. All my teaching get evaluated, though I probably could ask not to evaluate all the time. My teaching is not conventional, i.e. not university papers.	p5
			I am required to collect student data.	p12
			Student teaching evaluation data is required by the university.	p15
			Evaluations mandatory but largely useless.	p23
			Required by University.	p34
			required to collect student evaluation data.	p36
			we are obliged to.	p38
			I do not find the current student teaching evaluations useful, but am forced to collect them.	p39
			We are required to get student teaching evaluation of much of our teaching.	p43
			What is required and what is available.	p44
			Student data is the easiest to obtain because there is an established procedure for it, and also we have guidelines for how frequently we should obtain it and what scores we should aim for.	p45
			It is mandatory for me to collect student teaching evaluations. Informal discussions with my colleagues often provide useful information.	p48
			It is collected by the university and automatically sent to us. We have to read and respond to it online.	p54
			SET is compulsory by the university the others we do because we are interested in our teaching and students learning.	p55
			Student evaluation data is collected in a standardised measure.	p56
			Required by my university.	p58
Students Feedback	SET(p31, p41); SET+IDC(p19, p33, p49);	However, most teachers endorse the formative application of student ratings, where evaluations are used as a method for gathering useful feedback from students	Collecting acting on feedback during the semester shows students the feedback directly impacts them. Having multiple surveys, students comment on teaching that occurred closest to the survey which may be forgotten by the end of the semester. Five instructors are involved in delivering the labs. The instructors discuss improvements informal feedback they received from students during after labs. This is then written on the lab instructor notes as post-lab notes improvements for next time.	p19
			I would like to know the opinion of my audience in regards to the usefulness and effectiveness of my teaching.	p31
			To be as responsive to students' learning needs and also ensure content and assignments remain applicable to their work settings and roles.	p33
			Canvases my direct audience (students); 100+ pieces of feedback versus just 1 from a peer observer.	p41
			Have to collect the XXXX forms. I co-teach with some colleagues, and we discuss teaching and I ask students what they like and don't like about classes, information not elicited by the XXXX forms .	p49

Continued...

Theme	Responses	Interpretation	Example of Quotations	Partici-pants
Promotion (Promotion and Progression)	SET(p3, p17, p29); SET+POB(p26); SET+IDC(p25, p27, p37, p42, p46); All (p7, p21); All+Others(p47)	Most participants require SET for promotion, progression and confirmation.	This is a needed documentation for the teaching portfolio [p3].	p3
			compliance and for promotion!	p7
			Requirements for promotion (XXX forms) To guide my teaching and get feedback (informal surveys).	p17
			we have been advised that having a range of teaching evaluation data is helpful for promotion and progression.	p21
			Promotion and feedback.	p25
			I am on the Confirmation Path and need this documentation to be confirmed [p26].	p26
			Formal teaching evaluations are required for progression and promotion.	p27
			required for promotion.	p29
			Need student evaluations for end of year review, but don't find it particularly useful.	p37
			Have to do the teaching evaluation for promotion purposes.	p42
			I collect student teaching evaluation data because it is a requirement for confirmation and promotion.	p46
			Needed for promotion, improving teaching.	p47
Improve Teaching Practice	SET(p18, p20); SET+POB(p26); SET+IDC(p8, p10, p21, p27, p37, p42, p46); All+Others (p16)	There are also several other reasons the Universities require teachers to evaluate teaching, such as quality of teaching.	The university survey data also gives a broader perspective on our activities and what is working well (or not) for students.	p8
			Efficiency.	p10
			I believe that teaching is the most valuable thing that the academic staff of our University do, even if the irrational demand for constant research is both monetised and weaponized against the need for constant improvement in our teaching and approaches.	p16
			least hassle.	p18
			Feedback from the student perspective is important for improving my teaching practices.	p20
			I do find informal discussions with colleagues about teaching to be fruitful and helpful.	p21
			I also value the data to help me improve my teaching generally.	p26
			Informally discussions with colleagues are the most useful form of evaluation of my teaching.	p27
			Also talk to colleagues for their perspectives.	p37
			much prefer learning and discussing ideas with colleagues informally.	p42
			I have discussions with colleagues to support my development.	p46
Diverse feedback	SET(p39); SET+POB(p9); SET+IDC(p8, p15, p19, p23, p43, p48, p53); All (p1, p6, p12, p28); All+Others (p32,p51, p52, p57)	However, academics use different types of evaluation data, and the reasons are associated with getting a diverse range of feedback from teaching practice	I believe it is important to gather a diverse range of feedback on my teaching.	p1
			Better to get different perspectives, Helps with reflection.	p6
			it is important To gather as many forms of data as possible. student evaluation is easy To administer, and a lot of my teaching is done in teams so discussing these sessions with teaching colleagues is valuable.	p8
			It is necessary for my to collect SED, but it is also becoming required to include Peer Observation. I co-teach some workshops so informal discussion about teaching is natural for me.	p9
			I value the observations of my colleagues, both informal and formal peer observation.	p12
			Informal discussion with colleague is an indispensable part of teaching development.	p15
			In the skills papers:- We collect feedback via a paper survey 2-3 times during the semester. As the papers are new, multiple points of feedback are helpful in improving the paper. Five instructors are involved in delivering the labs. The instructors discuss improvements informal feedback they received from students during after labs. This is then written on the lab instructor notes as post-lab notes improvements for next time. Post semester meeting of all the instructors to review the feedback discuss improvements.	p19

Continued...

Theme	Responses	Interpretation	Example of Quotations	Partici-pants
			Discussions with colleagues keeps abreast of aberrations in cohorts from one year to the next.	p23
			formal evaluation not always useful; need other sources, qual and quan.	p28
			We collect as many data points as possible to triangulate.	p32
			As a senior academic, I provide peer observation for others, and discuss / mentor their teaching; I'm fairly well set in my own style.	p39
			I haven't organised for a formal peer observation recently but plan to do so. I talk with colleagues with whom I co-teach a paper about that paper, and in more general ways with other teaching staff.	p34
			Informal discussions with my colleagues often provide useful information.	p48
			The wide range of data allows me to understand the students current (as well as past and future) status, understandings, concerns and expectations about the learning experience. Multiple sources helps me triangulate and get views from a wide range of students.	p51
			Need many forms of evidence.	p52
			Discussions colleagues and with own evaluations - to help work out what makes a difference to student learning and to explore differences in students' engagement and participation.	p53
			It is helpful to have a diverse data set to contribute to my evaluation of my teaching.	p57
System Problem	SET(p39); SET+POB(p9); SET+IDC(p8, p15, p19, p23, p43, p48, p53); All (p1, p6, p12, p28); All+Others (p32,p51, p52, p57)	Although the benefit inherent in the online evaluation was highlighted, some interest also was expressed towards giving preference towards traditional evaluation, which suggested that the traditional method of evaluation produce a higher rate of return.	In the skills papers:- We collect feedback via a paper survey 2-3 times during the semester. Paper surveys in lab increases response rates. In the didactic course with multiple instructors —used the university online end of paper evaluation due to ease of use in evaluating a paper with many instructors with a diversity of topics. Students completed at the beginning of the last mandatory teaching session.	p19

End of Table

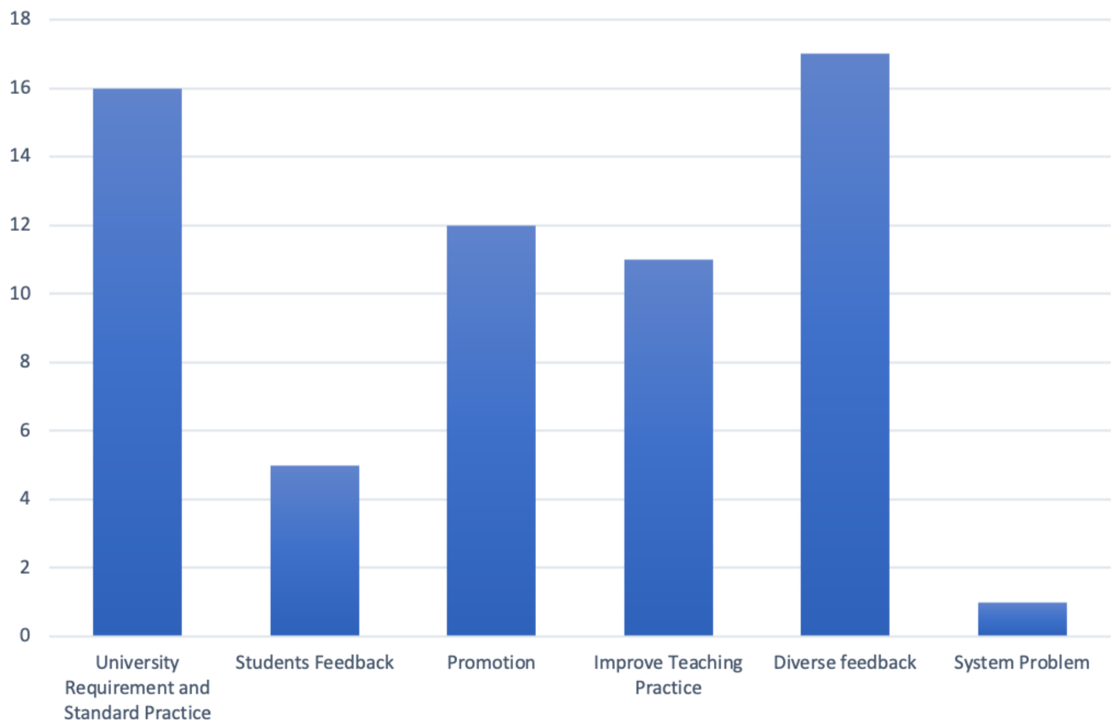


Fig. L.1 Frequency Table of themes identified from why participants collect types of teaching evaluation data.

Appendix M

THEMATIC ANALYSIS OF THE IMPORTANT REASON FOR USING TEACHING EVALUATION DATA

Theme	Responses	Interpretations	Example of Quotations	Partici-pant
University Requirement and Standard Practice	Promotion(p12); Others(p39, p53)	Other issues mentioned were the related unease about the institutional use of SET data, including concerns about data quality, low responses and institutional dependence on one source of measurement to base assumptions and decisions.	Teaching evaluations are a poor method of assessing what is happening in class. We are however required to go through the process, so we do.	p12
			Mandated by the university.	p39
			We have an online system with mainly generic items that are not useful to me as a teacher, the repetitive nature of these bore students, and this combined with low response rates means that despite being an advocate for gathering student feedback and being interested in student feedback I do not think the current system works to provide quality feedback.	p53
Students Feedback	ImproveTeachOut-come(p20, p45, p58); Others(p54)	Participants questioned the student's ability to judge teaching, and a concern was expressed that students could be biased due to many reasons, such as accessible courses and likeable teachers. A view expressed by participant.	This is the reason we require feedback from students, it's for their benefit not our own.	p20
			I find it most useful when students give open ended comments about what does or does not work for supporting their learning. The numerical metrics used in promotion are not particularly meaningful as it depends on how many people respond, and who responds usually it is the people who particularly loved or hated their experience. I would find it more useful if we could capture the average students in the middle who just thought it was "kinda OK".	p45
			it would be useful if students put their name to comments so that they comments were more constructive feedback when you get groups of students saying the same thing out of anger or frustration it isn't helpful.	p54
			Feedback as part of the teaching learning communications loop.	p58
Promotion	ImproveTeachOut-come(p21, p28, p34, p43, p44); Promotion(p23, p25, p46); Others(p54)	Several participants that indicated improving teaching is the most important reason for performing teaching evaluation, also believed SET is set up in a manner that makes promotion the focus.	I think this is the most important reason, but also think that this is impeded by the poor quality of data in teaching evaluations. So it turns into a process for promotion rather than about teaching quality.	p21
			Promotion is the only use for them.	p23
			I use informal evaluation to really change my teaching. The current ones are just good for promotion.	p25
			I would have said both learning outcomes and learning about teaching. Using for promotion doesn't improve either of these.	p28
			I must use the evaluations for promotion, but in general student feedback is important for improving outcomes.	p34
			If I can discover ways that work or don't work at helping my students to learn, that is the optimal outcome. Unfortunately, in practice I think they are more used for promotion.	p43

Continued...

Theme	Responses	Interpretations	Example of Quotations	Partici-pant
			That's what we should use it for: sadly requirements for promotion seems more concerned with blunt tools for assessing easily measured aspects of performance.	
			Given the low response rates and students' general lack of interest in honestly completing evaluations, their only real purpose is for promotion and confirmation.	p46
Improving Teaching	ImproveTeachOut-come(p3, p10, p19, p49); Promotion(p37); Others(p39)	The overall goal of SET is to improve teaching outcome.	Evaluations help me improve the course so that it benefits students better in terms of what they are looking for when they attend the course.	p3
			because evaluation should inform teaching.	p10
			The goal of teaching is helping students to learn hence improving teaching outcomes (i.e. students' learning) is most important. As part of this you learn about teaching.	p19
			to monitor quality I think it should be for improving teaching but most people use it as a check box for promotion, so long as they're not doing terribly, then it's considered okay.	p37
			The formal evaluations are useless for improving teaching, and I am not eligible for promotion. Students are able to express themselves far better in conversation than on written evaluations.	p39
			The results don't matter that much for promotion unless your results are dodgy, and I'm only really interested in learning about teaching my own classes, not more generally.	p49
Improving Learning	ImproveTeachOut-come(p8); learningAbt-Teach(p32); Others(p5, p9, p51)	Participants did not just see teaching evaluations improving themselves only, but also viewed it as part of a shared teacher-learner relationship in which both parties have a significant stake.	It's mostly a management tool to ensure a certain level of student satisfaction from teaching.	p5
			With a very student centred philosophy, it is all about improving the learning experience.	p8
			Ultimately, teaching should improve learning.	p9
			Learning about teaching covers learning about your teaching as well as learning about teaching to contribute to the scholarship around teaching.	p32
			My role is about creating a space for significant learning to happen that meets the intended learning outcomes. the data collected via teaching evaluations helps me to determine am I doing that, what I am doing wrong, and what I could do differently. For improving my facilitation of an environment that creates significant learning moments in the students.	p51
Quality of Data			We have an online system with mainly generic items that are not useful to me as a teacher, the repetitive nature of these bore students, and this combined with low response rates means that despite being an advocate for gathering student feedback and being interested in student feedback I do not think the current system works to provide quality feedback.	p53

End of Table

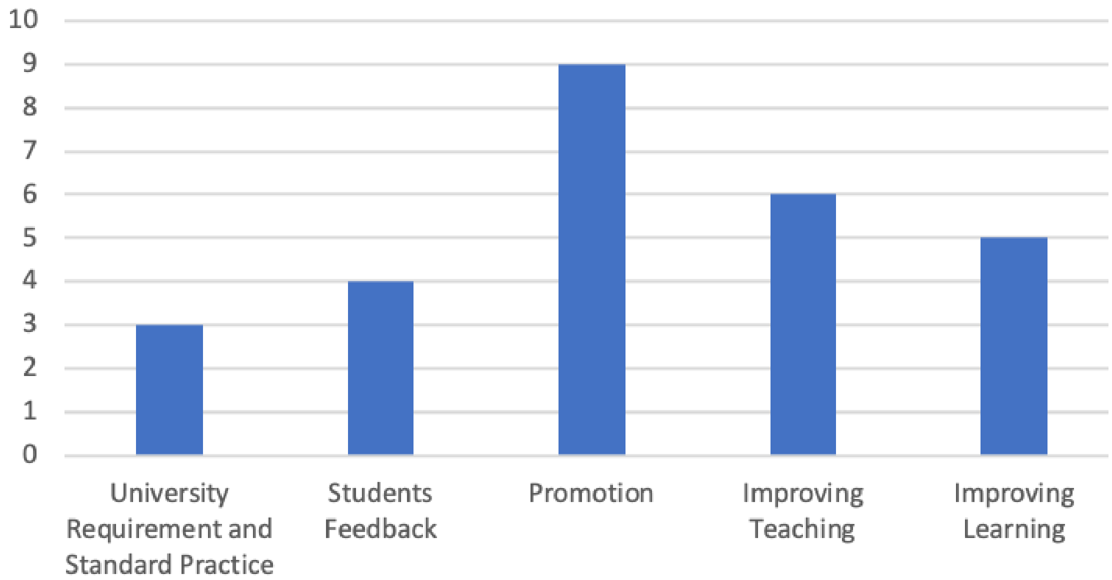


Fig. M.1 Frequency Table of themes identified from the last time participants used teaching evaluation data to improve teaching.

Appendix N

THEMATIC ANALYSIS OF THE AWARENESS OF DASHBOARDS

Theme	Responses	Interpretations	Example of Quotations	Partici-pant
Evaluation System	Unaware(p3, p36, p54, p55, p58); Aware(p33)	Few participants were aware of dashboards, but maintained that they had not experienced it in the Institution, p33. However, several participants re-echoed that they were not aware of dashboards. Additionally, some of them seemed not to be satisfied in the way SET results are presented in the Institution, and chanted that they had limited access to SET data p36	There aren't any dashboards provided for teachers at XXXX so there is no example that I could refer to.	p3
			I have not experienced the use of dashboards at XXXX.	p33
			never heard of it. We usually just get emailed the student evaluation summaries in PDF form.	p6
			I don't collect or prepare and present the actual data. I just read the graphs and results summaries and comments and then fill out an online feedback form.	p54
			At our university we only get access to the summarised data. Our use of the system is limited by the administrators.	p55
			We don't have teaching evaluation dashboards at my university. We get a PDF report sent to us.	p58
Not aware of Dashboards	Unaware(p8, p13, p31, p32, p36, p39, p45, p46, p47, p51, p52, p54); Aware(p4)	Many of the participants do not have experience of dashboards	Cars have dashboards.	p4
			Have not heard about these.	p8
			I'm not familiar with Dashboard, what is it?	p13
			I do not know what this dashboard is.	p31
			I have never heard of it.	p32
			I do not know what that term means.	p39
			not sure what else do say.	p45
			I am not familiar with dashboard.	p46
			Haven't used it.	p47
			I never had the opportunity to try.	p51
			Because I don't know.	p52
Aware of Dashboards	Unaware(p28, p43); Aware(p23); Expert(p56)	Only very few participants might have experienced dashboards	I've looked at it once I think.	p23
			Self-evident!	p28
			Obvious	p43
			I currently work in the learning and teaching group, and am involved heavily with use of the evaluation data	p56

Key: XXXX represents name of a department, faculty or institution.

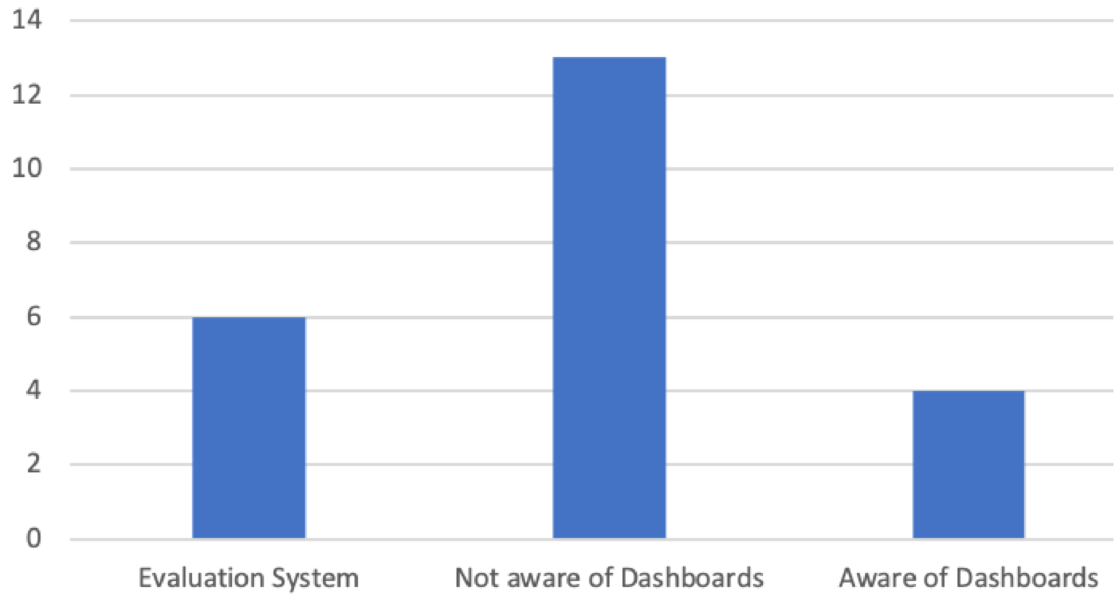


Fig. N.1 Frequency Table of themes identified from the last time participants used teaching evaluation data to improve teaching.

Appendix O

THEMATIC ANALYSIS OF THE UTILISATION OF DASHBOARDS

Theme	Responses	Interpretations	Example of Quotations	Partici-pant
Promotion	Likely(p7, p44)	Some participants remained sceptical and feared that TADs could be another form of Institutional auditing tool to monitor performance rather than improve teaching performance.	for fun to see if I can make stronger claims about my competence.	p7
			I would like to know more. It would be good if it was helpful in assessing my teaching but there is a risk that it is just another way of collating information for promotion and progression.	p44
Evaluation System Functionality	Likely(p3, p34, p56); Neutral(p31, p49, p51, p53)	A few others indicated conditional acceptance of TAD based on perceived usefulness.	I will surely use it if it is available and can provide better statistics than what is provided through the normal evaluation report.	p3
			Whether or not I would use it would depend on what sort of functionality it offers.	p31
			I'm not sure if 'dashboard' is a new or proposed method for dealing with Student Teacher Evaluations data, or is this just describing the current reporting format as a dashboard?	p34
			If it were more informative than the current statistical summaries, I'd defiantly use it.Otherwise, I'd engage with it as required and ask my students questions in class to find out things I really care about.	p49
			It depends how user friendly it is, and can it collate and interpret the sources of evaluation I have just described before for me.	p51
			if I am already using under another name, and would need to try it to see if it provided useful information before I could say if likely or unlikely to use.	p53
			To have a faster and consistent way of analysing the data is always very beneficial.	p56

Continued...

Theme	Responses	Interpretations	Example of Quotations	Partici-pant
DB Usefulness	Likely(p8); Neutral(p16, p46, p19); UnLikely(p15, p23, p45, p52)	TAD supporters claimed that qualitative input would bring significance to quantitative data issues that arise, and incorporating qualitative student feedback data into TADs to determine the level of student opinion positivity and negativity, could enable academics to analyse discrepancies with quantitative outcomes, or even make it possible for academics to compare qualitative scores with the quantitative scores. For example P8. Participants that had a neutral opinion about TAD expressed concerns that students could be easily swayed by accessible courses and likeable teachers, therefore affecting the data quality and visual representation of SET via TAD, a view was expressed in a statement by participant p19. Others argued that SET data was simple enough and did not require an extra tool to interpret. For example, participant p16.	Sounds a good idea, and I am guessing presents a useful summary of the data. However, it is still the comments that are the most informative.	p8
			Very unlikely, because I plan never to look at the evaluation system again.	p15
			The data I collect from the InForm system are sufficiently straightforward that I can use the statistics and the freeform comments for the ends that I need to meet. If the evaluations were more complex, or if my data collection methods were limited to the numbers/bubbles, or if my classes were so large as to make the collection of those statistics meaningless (because above a certain size, classroom learning has not been proven more useful than no learning at all), then I suppose I could use more robust analytical tools. I think the Clocktower could use more robust tools for this sort of thing, the 5 mandatory questions charade is an insult to people who take teaching seriously.	p16
			Don't know anything about them. Concerned that teaching evaluations from students alone sometimes can be like popularity contests. Students sometimes do not distinguish the likability of a teacher from the effective teaching and learning.	p19
			As above, I value other metrics.	p23
			whether it would be useful, but I'm not keen on another platform to have to login to and monitor.	p45
			I am not sure if it will be useful.	p46
			What I do works well, why would I change?	p52
No clue	Likely(p13); Neutral(p5, p14, p16, p28, p31, p32, p36, p39, p46, p47, p50); UnLikely(p4, p6, p18, p45, p53, p57, p58)	Respondents that indicated that they were not interested in TAD were either not keen to migrate to other platforms or valued other metrics other than SET, and therefore questioned the usefulness of TAD p46	Why would I talk about teaching evaluation to my car.	p4
			How can I answer this question if I don't know what the dashboard will contain?	p5
			The system works now and I dont know what a dashboard is.	p6
			If I knew what it was! I use the standard questionnaires supplied by XXXX. I'm not aware of other means of assessment.	p13
			I do not know much about the dashboard.	p14
			I have no idea what it is.	p16
			Don't know anything about it.	p18
			don't know what it is, so can't say one way or the other.	p28
			I do not know anything about it, so I cannot say whether I would be likely to use it or not.	p31
			I have no idea what it does.	p32
			I don't know what it is.	p36
			No clue what it is. I cannot judge its utility without knowing what it can do.	p39
			No idea what it is so can't answer.	p42
			No idea what it is.	p43

Continued...

Theme	Responses	Interpretations	Example of Quotations	Partici-pant
			I have no idea what it is.	p45
			As I am unfamiliar with it.	p46
			Haven't heard of it.	p47
			not heard of it before.	p50
			Dont know what it is.	p53
			As I have no idea what it is, this is a silly question.	p57
			It is not available at my university.	p58

End of Table

Key: XXXX represents name of a department, faculty or institution.

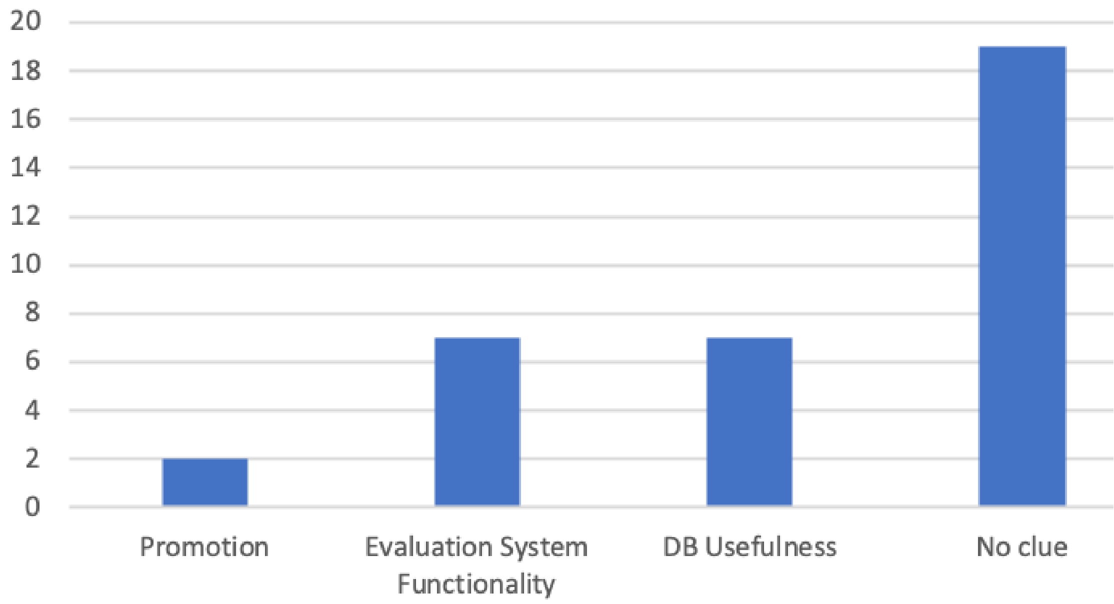


Fig. O.1 Frequency Table of themes identified from the last time participants used teaching evaluation data to improve teaching.

Appendix P

CHI-SQUARE TEST AND FISHER'S EXACT TEST

			Most important reason for SET			
			learning about teaching	promotion	improving teaching outcome	**Total**
Evaluation Type	SET	Count	0	2	15	17
		Expected Count	1.1	4.3	11.7	17.0
	SET POB IDC	Count	1	9	10	20
		Expected Count	1.3	5.0	13.8	20.0
	ALL	Count	2	1	8	11
		Expected Count	0.7	2.8	7.6	11.0
Total		Count	3	12	33	48
		Expected Count	3.0	12.0	33.0	48.0

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	11.110^{a}	4	0.025	0.22		
Likelihood Ratio	11.384	4	0.023	0.028		
Fisher's Exact Test	9.592			0.022		
Linear-by-Linear Association	3.497^{b}	1	0.061	0.068	0.032	0.004
N of Valid Cases	48					

a. 5 cells (55.6%) have expected count less than 5. The minimum expected count is .69.

b. The standardized statistic is -1.870.

Symmetric Measures

		Value	Approximate Significance	Exact Significance
Nominal by Nominal	Phi	0.481	0.025	0.022
	Cramer's V	0.340	0.025	0.022
N of Valid Cases		48		

Appendix Q

SPEARMAN'S RHO CORRELATION DISTRIBUTION TABLE

Spearman's Rho Correlation Distribution Table

ID	ITEMS		(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)	(L)	(M)	(N)
(A)	What type of teaching evaluation data are you currently collecting	Correlation Coefficient	1.000													
		Sig. (2-tailed)														
		N	58													
(B)	In what format(s) do you get your teaching evaluation data	Correlation Coefficient	.437**	1.000												
		Sig. (2-tailed)	0.001													
		N	58	58												
(C)	How do you rate student evaluation data as a means of improving your teaching	Correlation Coefficient	0.003	0.172	1.000											
		Sig. (2-tailed)	0.983	0.196												
		N	58	58	58											
(D)	How often do you evaluate your teaching?	Correlation Coefficient	0.045	0.066	0.240	1.000										
		Sig. (2-tailed)	0.737	0.623	0.070											
		N	58	58	58	58										
(E)	How often do you use the results of teaching evaluation to inform your teaching	Correlation Coefficient	-.270*	-0.090	.570**	0.163	1.000									
		Sig. (2-tailed)	0.040	0.503	0.000	0.222										
		N	58	58	58	58	58									
(F)	What do you think is the most important reasons for using teaching evaluation	Correlation Coefficient	-0.159	0.042	-0.101	-0.158	-0.115	1.000								
		Sig. (2-tailed)	0.233	0.752	0.450	0.235	0.392									
		N	58	58	58	58	58	58								
(G)	How confident are you in interpreting teaching evaluation data	Correlation Coefficient	0.046	-0.145	-0.041	0.067	0.162	0.099	1.000							
		Sig. (2-tailed)	0.734	0.278	0.761	0.617	0.224	0.461								
		N	58	58	58	58	58	58	58							
(H)	How likely would you require the support of others to help with the interpretation of teaching evaluation data	Correlation Coefficient	-.387**	-0.032	0.104	0.028	.265*	0.116	-.328*	1.000						
		Sig. (2-tailed)	0.003	0.813	0.438	0.834	0.044	0.386	0.012							

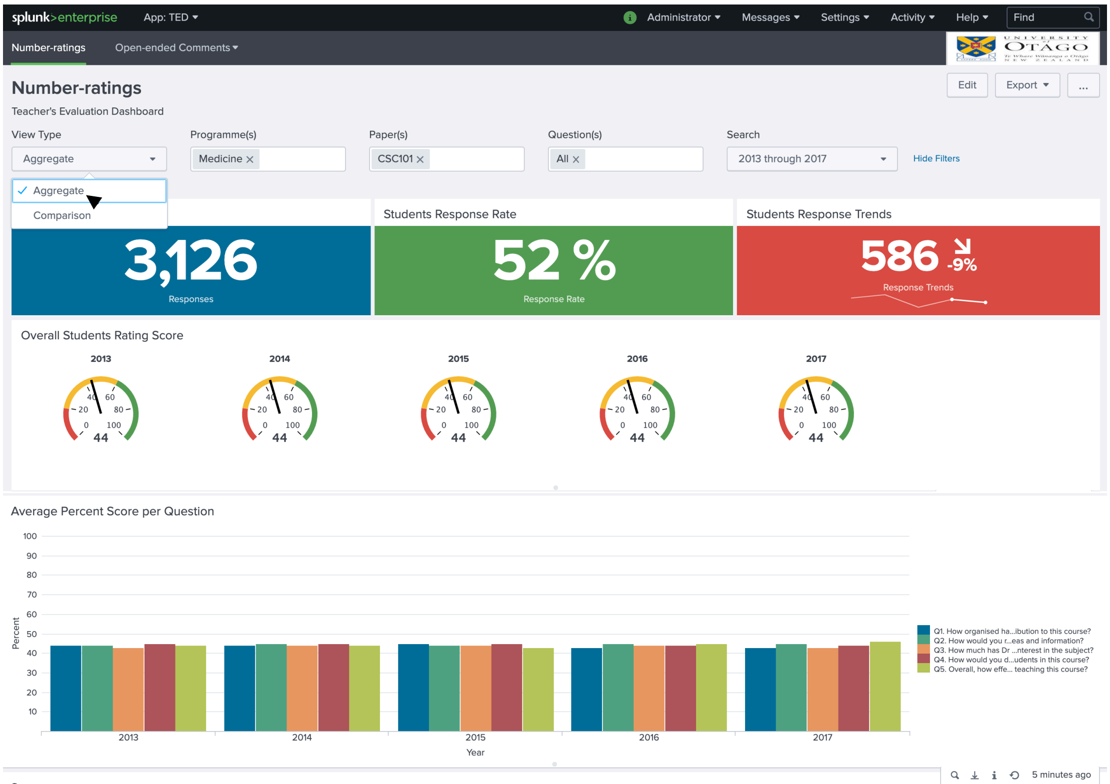
		N	58	58	58	58	58	58	58	58						
(I)	Which of the following best describes how much you know about the use of dashboard for presenting data for teachers	Correlation Coefficient	0.008	0.096	0.112	0.011	-0.099	.314*	-0.013	-0.190	1.000					
		Sig. (2-tailed)	0.955	0.481	0.411	0.937	0.469	0.018	0.925	0.161						
		N	56	56	56	56	56	56	56	56	56					
(J)	How likely would you use teacher evaluation dashboard	Correlation Coefficient	0.176	.441**	0.035	0.043	0.139	0.072	-0.085	0.139	0.089	1.000				
		Sig. (2-tailed)	0.186	0.001	0.795	0.750	0.297	0.589	0.525	0.297	0.516					
		N	58	58	58	58	58	58	58	58	56	58				
(K)	Rank	Correlation Coefficient	-0.105	-0.033	0.013	-0.016	0.082	0.173	.269*	0.021	-0.022	0.117	1.000			
		Sig. (2-tailed)	0.434	0.803	0.925	0.908	0.541	0.195	0.041	0.875	0.872	0.382				
		N	58	58	58	58	58	58	58	58	56	58	58			
(L)	Division	Correlation Coefficient	0.113	0.033	-0.034	0.028	-0.222	.304*	-0.218	0.195	0.147	-0.034	0.046	1.000		
		Sig. (2-tailed)	0.399	0.803	0.801	0.837	0.094	0.020	0.100	0.143	0.279	0.801	0.731			
		N	58	58	58	58	58	58	58	58	56	58	58	58		
(M)	Age	Correlation Coefficient	0.022	0.082	0.241	0.032	0.127	0.043	0.012	-0.108	.270*	0.036	-0.079	.280*	1.000	
		Sig. (2-tailed)	0.870	0.540	0.069	0.813	0.344	0.748	0.929	0.420	0.044	0.786	0.554	0.033		
		N	58	58	58	58	58	58	58	58	56	58	58	58	58	
(N)	Teaching Experience	Correlation Coefficient	.336*	0.252	0.126	0.198	0.010	-0.131	-0.105	-0.118	.295*	0.259	-.298*	0.183	.410**	1.000
		Sig. (2-tailed)	0.013	0.066	0.365	0.152	0.943	0.343	0.450	0.397	0.032	0.059	0.029	0.186	0.002	
		N	54	54	54	54	54	54	54	54	53	54	54	54	54	54

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

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

Appendix R

NUMBER-RATINGS AGGREGATE PAGE OF TED

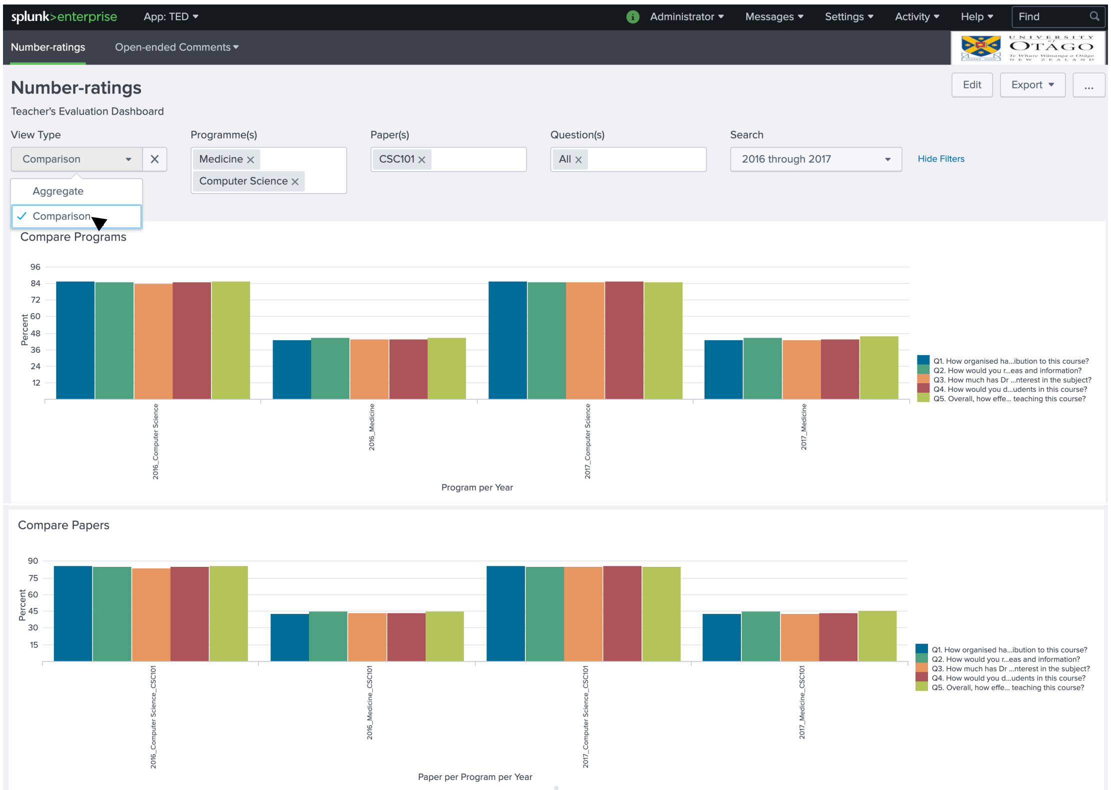


Summary

PAPER CODE	SESSION	TOTAL
CSC101	2013_2014	669
CSC101	2014_2015	732
CSC101	2015_2016	496
CSC101	2016_2017	643
CSC101	2017_2018	586

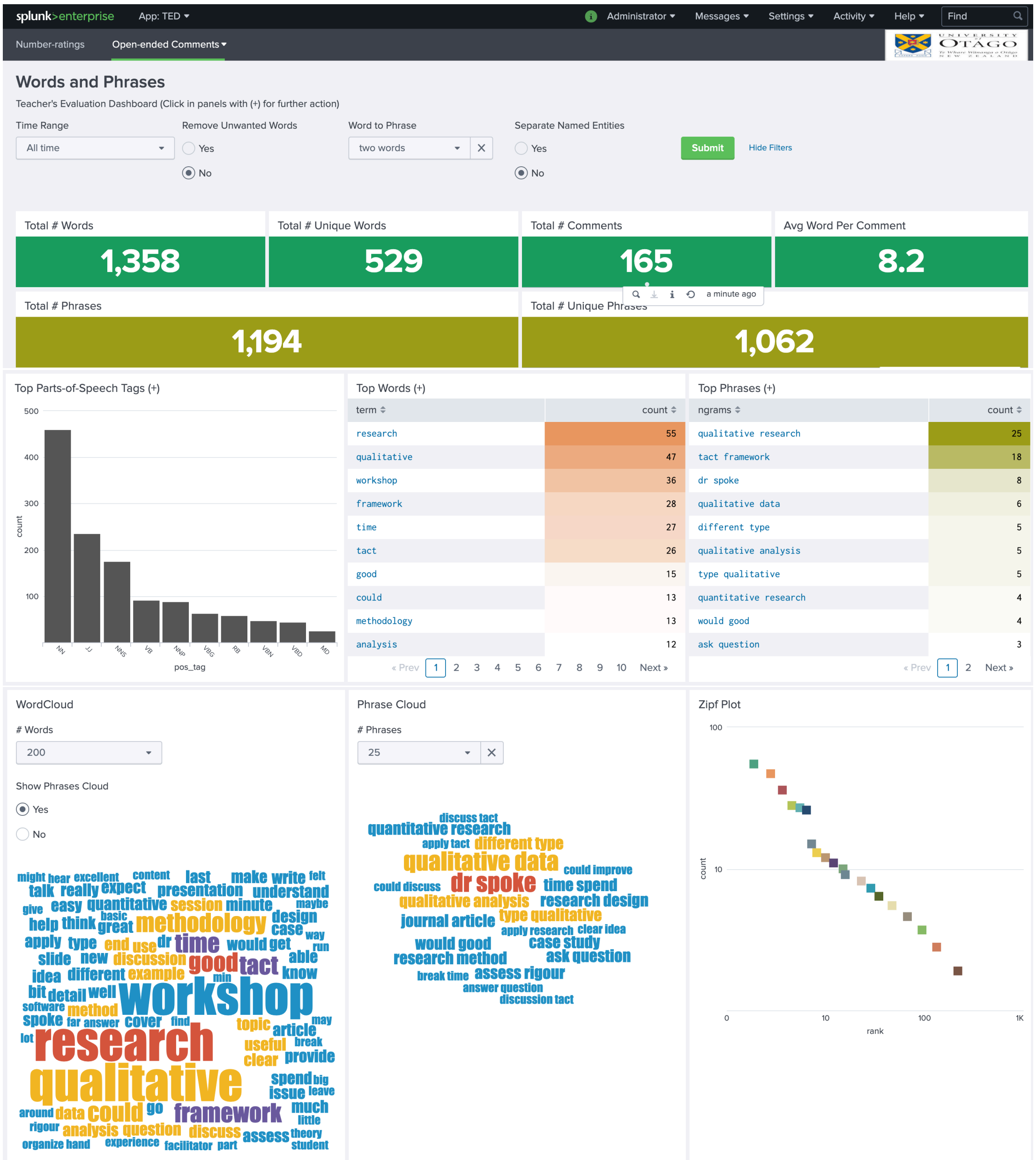
Appendix S

NUMBER-RATINGS COMPARISON PAGE OF TED



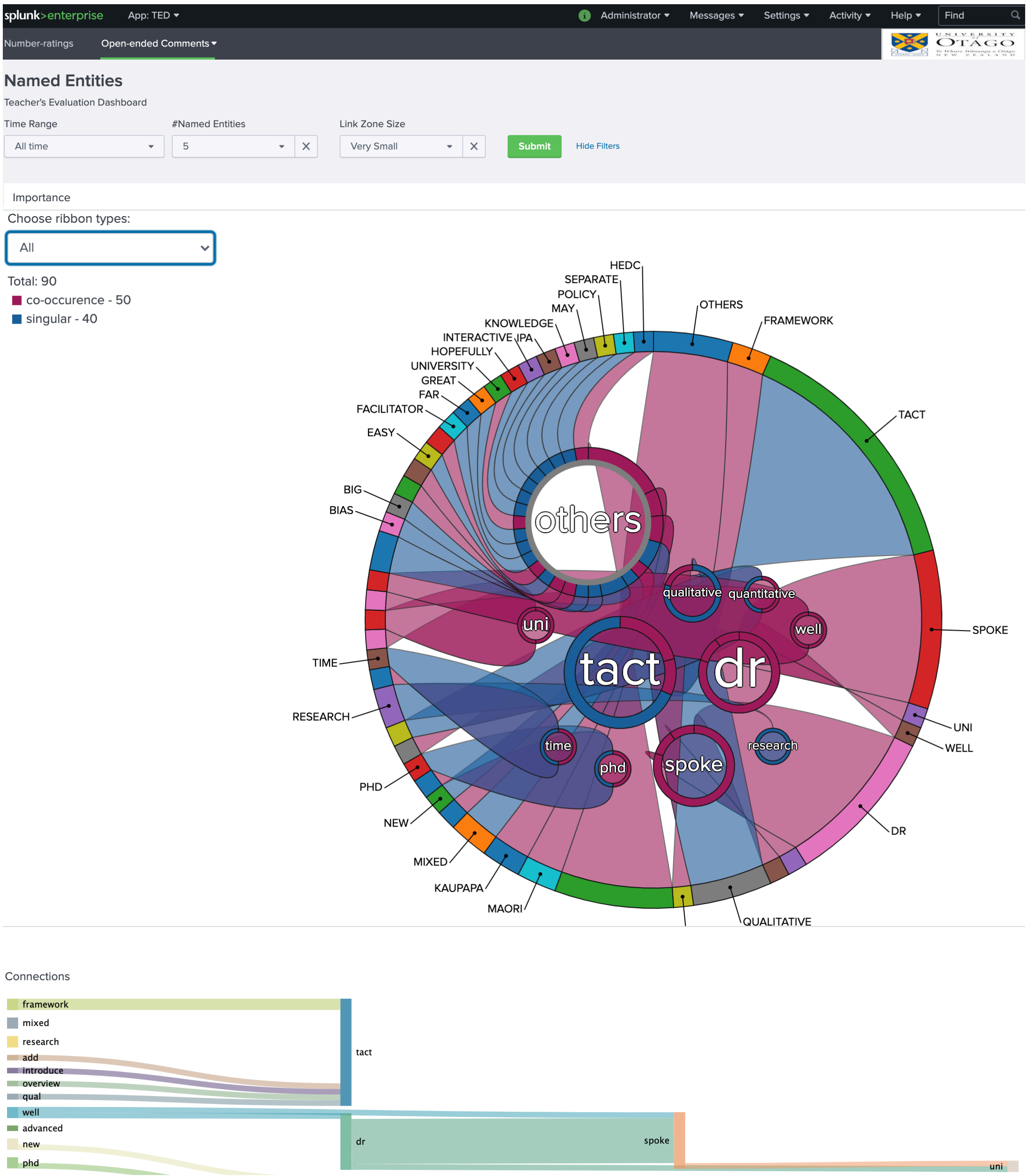
Appendix T

OPEN-ENDED COMMENTS WORDS AND PHRASES PAGE OF TED



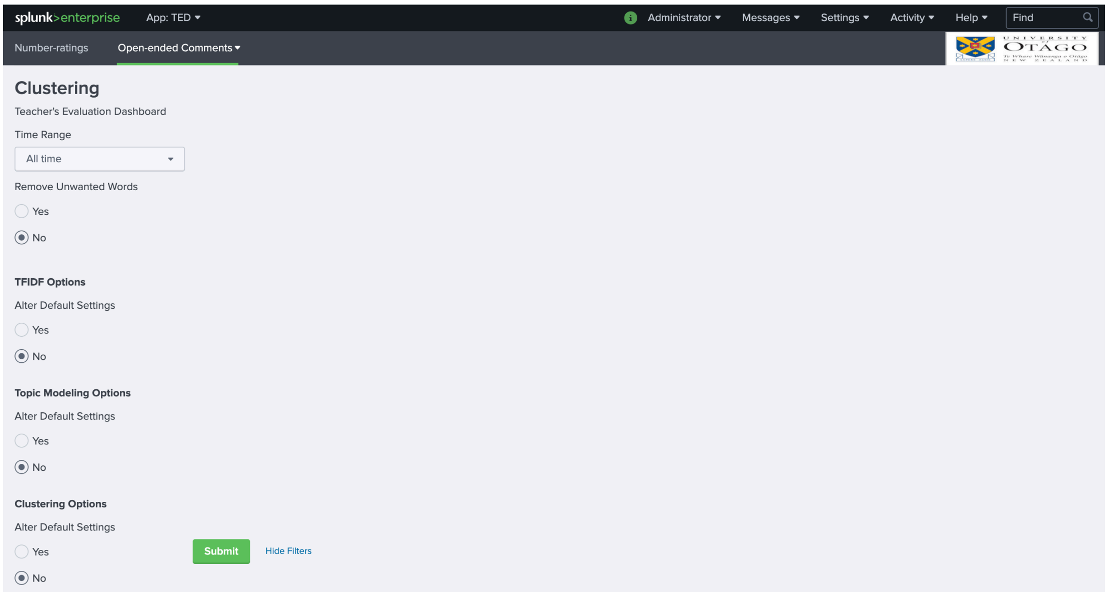
Appendix U

OPEN-ENDED COMMENTS NAMED ENTITY PAGE OF TED



Appendix V

OPEN-ENDED COMMENTS CLUSTER PAGE OF TED WITH FILTER OPTIONS LEFT AS DEFAULT



Top Terms Per Cluster

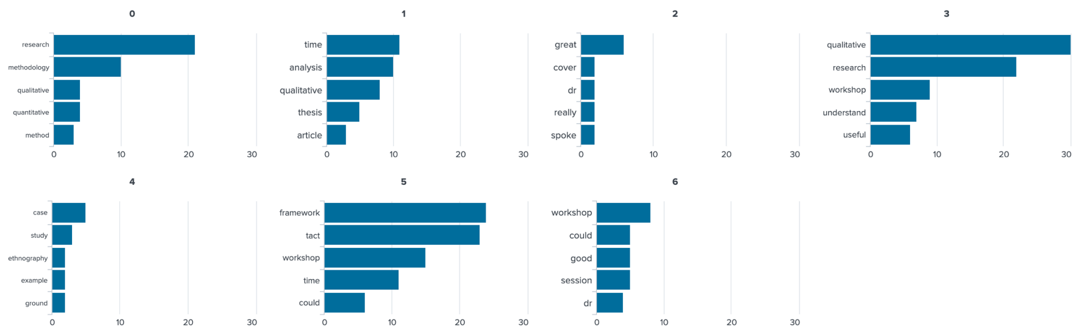
Term Count

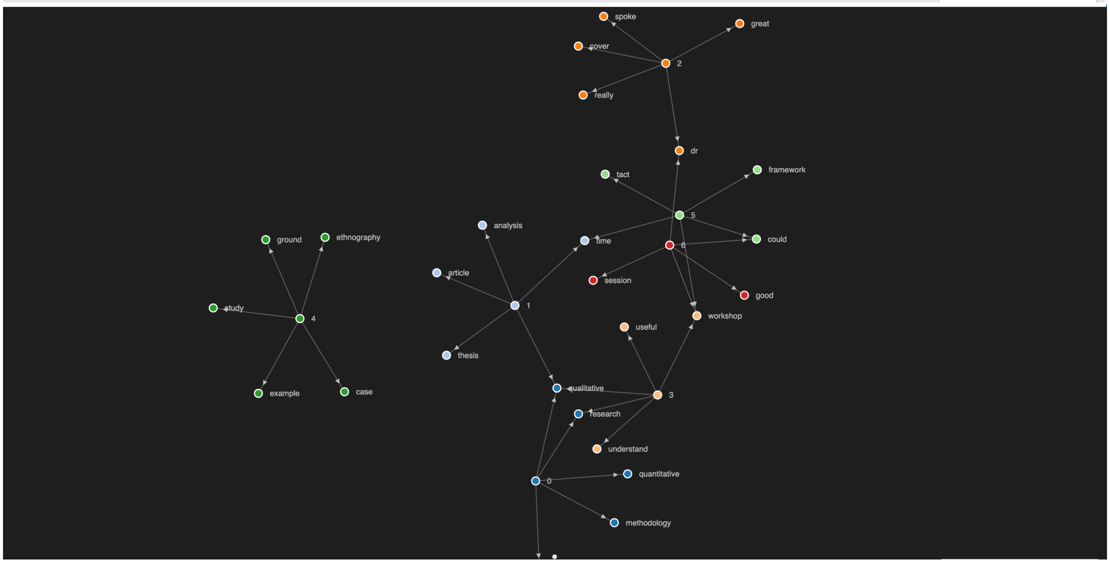
5

terms	cluster	documentProportion	documentCount
method methodology qualitative quantitative research	0	14.337349397590362	119
analysis article qualitative thesis time	1	13.012048192771083	108
cover dr great really spoke	2	3.734939759036145	31
qualitative research understand useful workshop	3	17.349397590361445	144
case ethnography example ground study	4	4.819277108433735	40
could framework tact time workshop	5	13.373493975903614	111
could dr good session workshop	6	33.373493975903614	277

Trellis Size

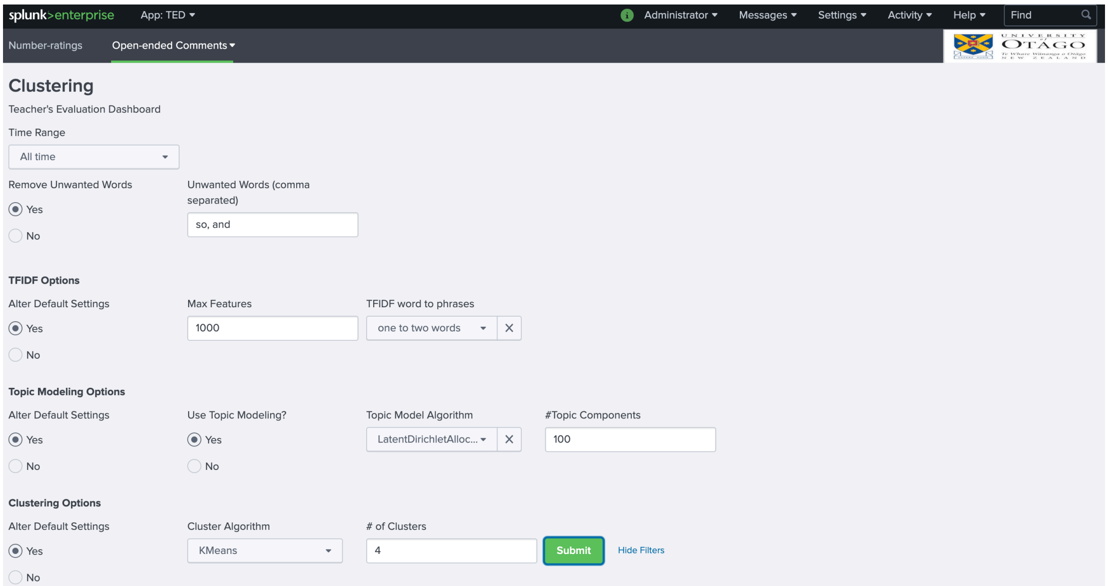
Medium





Appendix W

OPEN-ENDED COMMENTS CLUSTER PAGE OF TED WITH SOME FILTER OPTIONS SELECTED



Top Terms Per Cluster

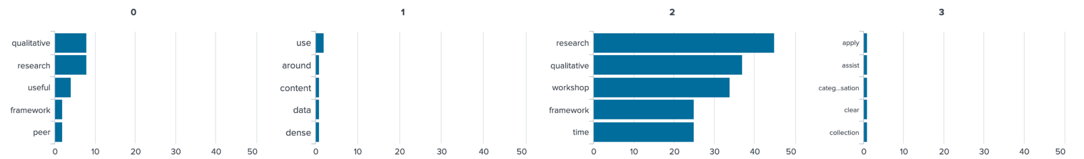
Term Count

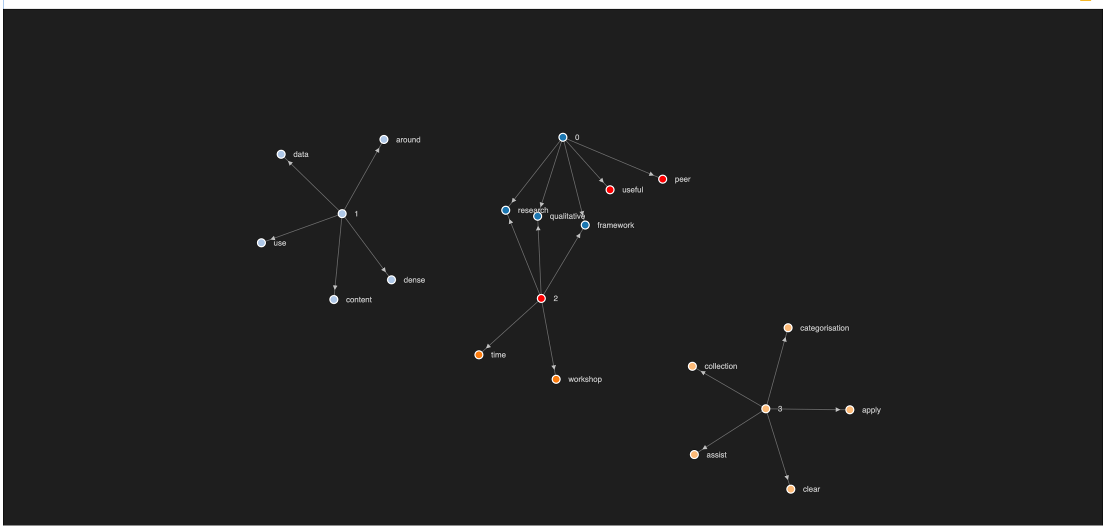
5

terms	cluster	documentProportion	documentCount
framework peer qualitative research useful	0	10.909090909090908	66
around content data dense use	1	3.4710743801652892	21
framework qualitative research time workshop	2	81.81818181818183	495
apply assist categorisation clear collection	3	3.8016528925619832	23

Trellis Size

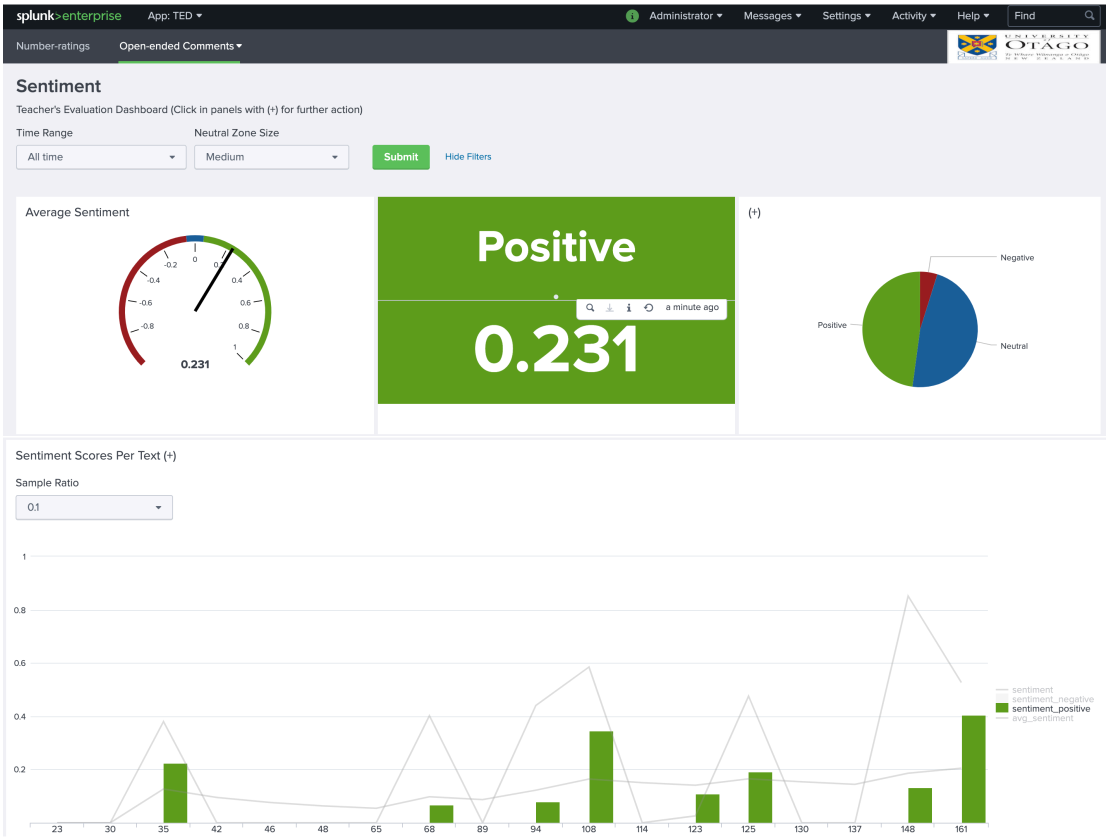
Medium





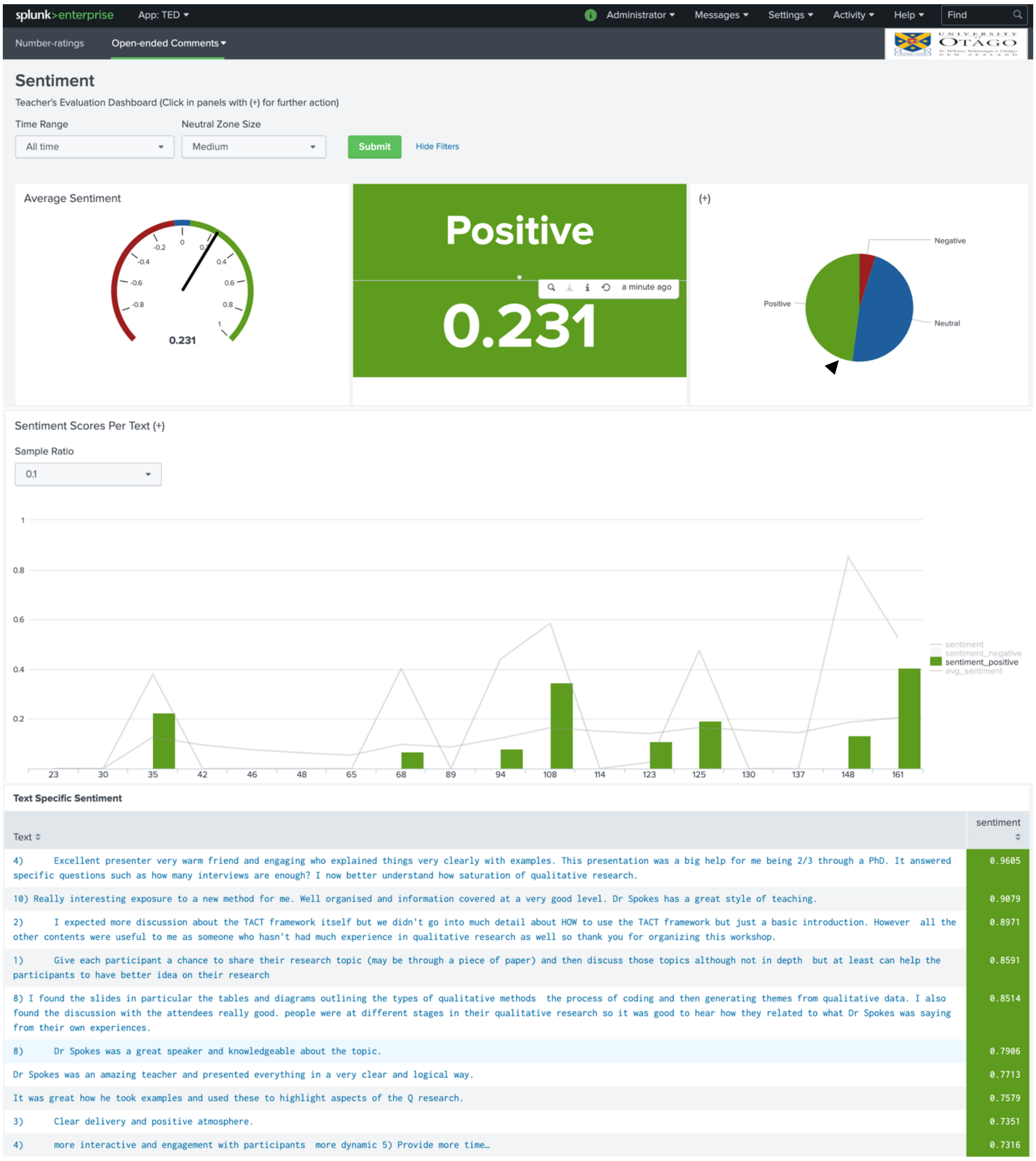
Appendix X

OPEN-ENDED COMMENTS SENTIMENT PAGE OF TED



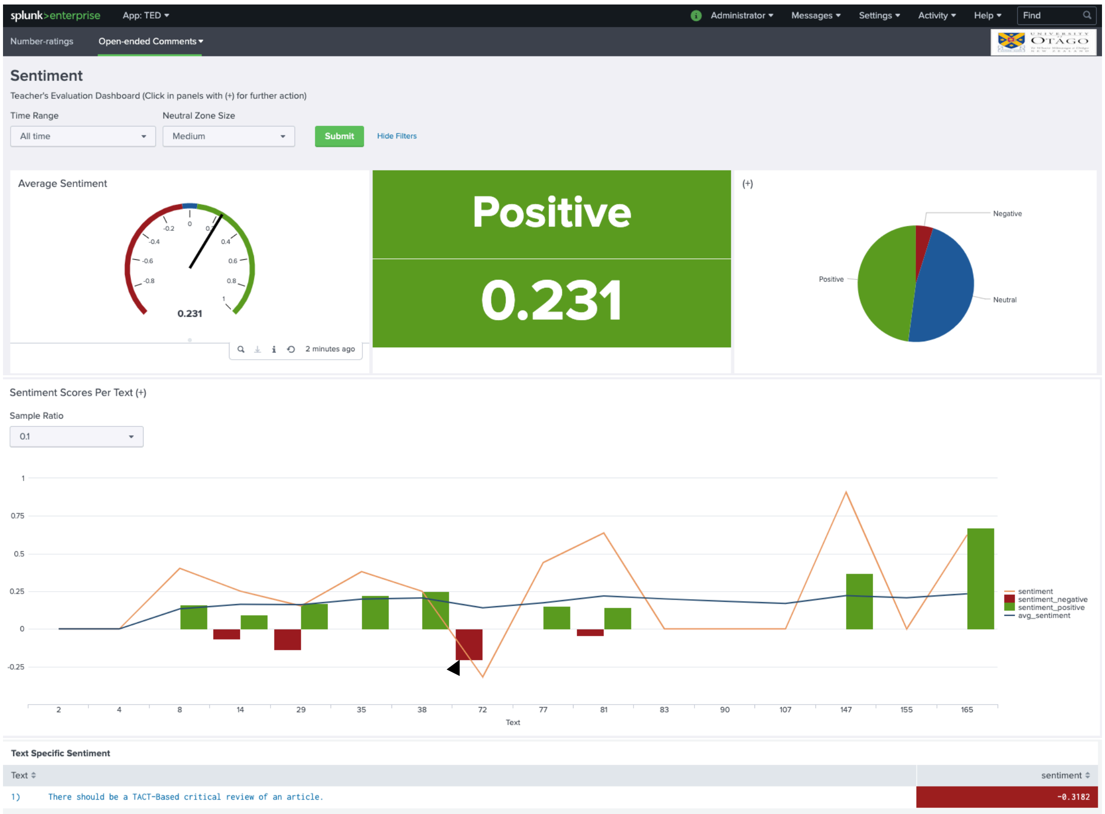
Appendix Y

OPEN-ENDED COMMENTS SENTIMENT PAGE OF TED WITH POSITIVE PORTION OF PIE CHART CLICKED



Appendix Z

OPEN-ENDED COMMENTS SENTIMENT PAGE OF TED WITH ONE NEGATIVE BAR CHART CLICKED



Appendix AA

THEMATIC ANALYSIS FOR THE USABILITY OF THE NAMED ENTITIES DASHBOARD

Theme	Response	Participant	Iteration
Satisfaction	Looks pretty and looks like something I could hang on the wall.	p3	1st
	Seems interesting, the visualisation is good	p5	2nd
	I really like it. Yeah. Really interesting.	p6	2nd
	This looks beautiful	p9	2nd
	It seems very pretty	p13	1st
	It seems very pretty and it looks very fancy.	p14	2nd
	Wow! this is fantastic.	p16	2nd
	It looks pretty though, I like it.	p17	1st
Hard to interpret	Not able to interpret the meaning of this dashboard	p5	2nd
	need some instructions on how to use it	p8	1st
	it is a little bit confusing.	p10	2nd
Information Overload	Lots of information, I might take a wee while to get used to it	p9	2nd
	It's obviously given a lot of information	p10	2nd
	I am trying to wrap my head around that information. I probably need to go over it to see if it's something that would be helpful.	p11	2nd
	Obviously, a lot of information.	p12	1st
	I think this dashboard, in particular, is really detailed to be able to draw conclusions.	p18_19	1st
	The information is dense. This dashboard would require a little bit more time to play around with the content. So we would need to, I suppose, inform ourselves of the types of keywords that we would be looking at ourselves.	p20	2nd
	It takes a little while to get used to. One might need some kind of training or something that would come with it.	p22_23	2nd
Interactivity with the Chord Visualisation	Can the Chord visualisation that presented the relationships between the Named Entities and their co-occurrences be clickable to get the words with the habits for the individual response?	p7	2nd

Appendix AB

THEMATIC ANALYSIS FOR THE USABILITY OF THE WORDS AND PHRASES DASHBOARD

Theme	Response	Participant	Iteration
Satisfaction	This dashboard looks great	p9	2nd
	Interesting dashboard	p10	2nd
	I also like the word cloud.	p11	2nd
	its looks really good.	p15	1st
	This dashboard is really good.	p16	2nd
	This dashboard is really great.	p20	2nd
	looks really good	p22_23	2nd
Labels	Change the label titles to something that will be more preferable e.g. change the n-grams to 'Number of Words or Phrases'	p2	1st
	Filters might need more explanation on how to use	p4	1st
	some names have to be rewarded e.g. n-grams can be renamed as Words to Phrases, stop words renamed to unwanted words.	p8	1st
	Terms Could be changed to words, change n-grams to number of words that occur together	p12	1st
	Some of the terminology used on labels may be quite confusing, for example, rename n-grams to say; how many words, and the options to say; one, two, three, four. You may even want to default with the two n-grams option.	p15	1st
Single Value Counts	I prefer the Average Words Per Comment rather than the Total # Word Count and Total # Comments	p4	1st
	Total Word Count, Total Comments, Unique Word Count, Phrases, Unique Phrases may not make so much sense to a teacher.	p3	1st
Phrases	I like the idea of the top phrases	p4	1st
	This dashboard helps isolate what the students are saying and create opportunities for the teacher to easily collect the most common words or phrases used by the students, and to know how many times they were used	p10	2nd

Appendix AC

THEMATIC ANALYSIS FOR THE USABILITY OF THE COMPARISON DASHBOARD

Theme	Response	Participant	Iteration
Satisfaction	Dashboard is pretty good	p5	2nd
	It is good.	p10	2nd
	The dashboard is pretty good	p12	1st
	It's quite good. This dashboard makes it very easy to make two comparisons.	p15	1st
	That's good, quite clear dashboard.	p18_19	1st
	Looks good to me.	p22_23	2nd
Transfer *Average Percent Score per Question* visualisation from *Comparison* Dashboard to the *Aggregate* Dashboard	Bars do not really compare anything.	P1	1st
	You should transfer the Average Column Chart from this dashboard to the Aggregate dashboard	p2	1st
	Average Column Chart should be taken to the Aggregate section	p12	1st
Replace Programs Comparison Visualisation from line chart to bar chart.	The line chart is not to easy to read, would prefer bars instead, for each year to have bars, one for each programme, next to each other.	P1	1st
	Line charts is not a sensible way of illustrating this kind of data, the bar charts would be a preferable way to represent this kind of information.	p2	1st
	Histogram may be more ideal than a Line graph	p3	1st
	I prefer the bars rather than line graphs representation.	p4	1st
	Column Charts will be a better visualisation than the Line Chart	p8	1st
	Column Charts is a better visualisation than the Line Chart	p12	1st
	Change line to bars	p13	1st
	Bar chart rather than using a line chart	p14	1st
Rename Relational to Comparison	Change the Relational to Comparison	P8	1st
	Change Relational to Comparison.	p12	1st
Paper Comparison visualisation	Can you compare Papers as in addition to comparing programs. I think it would also be useful to be able to compare the responses from the same student programme for two different courses. For example, I might want to compare how Ecology students evaluated my teaching in both PAPER 101 and PAPER 202.	p17	1st

Appendix AD

THEMATIC ANALYSIS FOR THE USABILITY OF THE CLUSTER DASHBOARD

Theme	Response	Participant	Iteration
Satisfaction	I think this is pretty cool.	p6	2nd
	This is brilliant	p7	2nd
	It is brilliant	p8	1st
	I definitely like that bit.	p10	2nd
	It is beautiful	p12	1st
	Really impressive	p16	2nd
	Nice graphic.	p20	2nd
Hide Input Fields	Some of the fields could be hidden.	p2	1st
	Many input fields, hid the details e.g Cluster Algorithm, TFIDF	p8	1st
Automate Number of Clusters Input Text	Apart from being able to enter the number of clusters, it should be able to determine the optimal number of clusters based on the data, so that it could be automated.	p5	2nd
Column Chart Visualisation	Want some column chart cluster to be clickable to find out what each cluster is saying	p10	2nd

Appendix AE

THEMATIC ANALYSIS FOR THE USABILITY OF THE SENTIMENT DASHBOARD

Theme	Response	Participant	Iteration
Satisfaction	it is a good idea	p5	2nd
	I think this is really great.	p6	2nd
	The interface looks cool as well. The dashboard looks cool.	p9	2nd
	I like this one already.	p11	2nd
Average Sentiment Visualisation	I fancy the way the Average Sentiment visualisation simplifies the summary of students comments.	p3	1st
Pie Chart Interactivity	I really liked how the software was able to sort written comments into positive, negative and neutral categories.	p6	2nd
	I like this, the fact that you can click on a portion of either positive, negative or neutral to read the actual comment	p10	2nd
	if you can access the actual comments, it might be a lot more useful	p12	1st
	Clicking on the pie chart should take you to the comments	p17	1st
	Generally speaking, people interpret pie charts terribly. So people have a lot of difficulty with this sort of proportional thinking when they when they are faced with a pie chart, you would probably be better with a little bar chart here because then people are able to see the relative difference more easily for the percentages in the pie chart, bar charts may be preferable.	p20	2nd
Information Overload	I think one of the things that I find at this point is that there is a lot of information here that might be over and above what in most circumstances would really be required, or what I actually use the feedback to do. I would probably need to have a play with the interface.	p16	2nd
Weighted Comments	However, the information could be problematic, for instance, it does not take account of very strong few negative comments compared to many but light positives comments	p1	1st
Time Range	In future, implementing the time range feature will be useful for monitoring improvement in certain parts of teaching over time. For example, knowing that you speak too fast in lectures are looking to get feedback speaking fast over time.	p20	2nd

Appendix AF

THEMATIC ANALYSIS FOR THE USABILITY OF THE AGGREGATE DASHBOARD

Theme	Response	Participant	Iteration
Response Rate visualisation	So it will be helpful to have some kind of metric to know what the response rates are. For instance, a 25% class response rate may not represent the complete information, considering some factors such as the representative sample of the class, the selected sample that like and hate the course, and the method of selection; a 25% response rate that is randomly selected may represent a true reflection of the class more than a non-random selection of 25% response rate. However, I still think it will be good to know the proportion of the class that responded.	p21	1st
Confidence interval	It will also be a good idea to incorporate some sort of confidence intervals. If confidence intervals are wide and there are differences between questions, that may not mean anything. But if the confidence interval is narrow and there are differences between questions, then that might be an indication that the lecturer needs to do something about. One thing that is of concern is a tool providing bad data, points users effort in the wrong direction may be making things worse. Consequently, costing users to invest time and resources that could have been spent somewhere else.	p8	1st
	I would also want to know what the differences are in terms of significant testing.	p22_23	2nd
Color of Radial Gauge Vs Single Value Visualisation	The radial gauge should be changed to one colour rather than the traffic light colours used to represent the information	p1	1st
	the red colour of the Student Responses single value visualisation was conflicting with one of the traffic light colours on the radial gauge, and could be misrepresenting.	p4	1st
	I like the traffic light colours of the radial gauge and the interactivity going on there.	p9	2nd
Time Picker	Search time picker may be a little tricky to know which academic year to select. For instance, use the time picker to select 2013/2014 academic year, how do I know which of the year options to choose from; either 2013 or 2014.	p4	1st

Appendix AG

THE SUS FINAL RESULT FOR THE TEACHER'S EVALUATION DASHBOARD

Participant	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS Raw Score	SUS Final Score	Adjective Ratings
P1(1st)	2	2	4	2	3	2	4	2	3	2	26	65	OK/ fair
P2(1st)	4	4	3	5	3	3	3	3	3	5	16	40	Poor
P3(1st)	1	5	4	3	4	2	2	3	4	3	19	47.5	Poor
P4(1st)	4	2	4	2	2	4	4	2	4	2	26	65	OK/ fair
P5(1st)	2	1	5	2	4	2	4	2	4	2	30	75	Good
P6(1st)	4	5	2	4	4	3	2	4	2	4	14	35	Worst imaginable
P7(1st)	4	2	5	2	4	2	4	2	5	5	29	72.5	Good
P8(1st)	5	2	5	2	4	2	4	2	4	2	32	80	Good
P9(1st)	4	4	4	3	4	3	4	3	4	3	24	60	OK/ fair
P10(1st)	4	2	4	2	4	2	5	2	4	3	30	75	Good
P11(1st)	3	1	4	2	3	2	3	2	4	1	29	72.5	Good
										Average SUS Score	**25**	**62.5**	**OK/ fair**

FIRST ITERATION

Participant	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS Raw Score	SUS Final Score	Adjective Ratings
P1(2nd)	2	1	4	1	4	2	4	2	4	3	29	72.5	Good
P2(2nd)	3	2	5	1	4	2	5	1	5	1	35	87.5	Excellent
P3(2nd)	3	2	4	1	4	2	4	2	4	2	30	75	Good
P4(2nd)	4	2	5	2	4	1	4	1	5	1	35	87.5	Excellent
P5(2nd)	5	2	4	2	5	1	4	1	4	2	34	85	Excellent
P6(2nd)	4	2	4	2	4	2	4	2	4	3	29	72.5	Good
P7(2nd)	3	1	4	1	5	2	5	2	4	1	34	85	Excellent
P8(2nd)	5	1	5	1	5	1	5	1	5	1	40	100	Best imaginable
P9(2nd)	3	3	4	2	4	2	4	2	3	2	27	67.5	OK/ fair
										Average SUS Score	**32.56**	**81.39**	**Good**

SECOND ITERATION

SUS Score	Adjective Ratings
92	Best imaginable
85	Excellent
72	Good
52	OK / fair
38	Poor
25	Worst imaginable

GRADING SUS KEY (Bangor et al., 2009)

Appendix AH

THEMATIC ANALYSIS FOR THE USEFULNESS OF THE NUMBER-RATINGS DASHBOARD

Theme	Response	Participant	Iteration
Flexibility of Questions	The flexibility of the questions is a good idea, rather than restricting the questions to the five standard questions used in the Institution. This question flexibility helps the teacher's dashboard address both compliance and learning. For example, a teacher might want to investigate a question about critical thinking; and that may be the focus of his teaching. The questions filter indicates to the teacher what specific questions are saying and where to improve. For example, a teacher may get 80% in Q1, 45% in Q2, 85% in Q3, 90% in Q4, and 70% in Q5, although the overall score was 74%, this kind of representation point that Q2 is where the teacher needs to improve.	p3	1st
	I like the fact that the dashboard is flexible and has the ability to add other questions apart from the university standard questions	p5	2nd
Dashboard presentation of SET result is an improvement	This is a huge improvement in reading this information on a PDF. Usually, we just see these numbers in written format; we don't usually see any charts or graphs to help with the visual representation of the data.	p6	2nd
	Because we got to the stage now where we are trying to do our own ratings or get email sent out to students to ask them about teaching. I think it's very difficult to get that out and it's difficult to see what the matrices are. So this will be something that will definitely help, because every once in a while, you want to get significant feedback of what you're teaching on how you teach, and that's what basically drives you to put more stuff online or change your teaching or, you know, change the speed of the way you talk or change the way you present the information.	p10	2nd
	I think this is a better system than what we already have; the static print outs of PDFs and sending it out via emails, thereby making it quite impossible to compare early years to know the trends over time. So this is a real advantage where you can select two time periods and look at trends through time. -At the moment as we are tied on to a particular system that people are used to seeing, so it will require some education of for that change, but this is probably more accurate than what we are already using.	p11	2nd
	One of the things that this dashboard does seem to demonstrate quite well is to see the trends or changes in previous years, which is very hard to get sort of a sense of the data in the current system that presents just a year's worth of data. This trend is good for alerting the teacher when something is going wrong.	p13	1st
	I think this dashboard provides a much more useful summary than the summary that I currently receive. The summary I received currently gives me those statistics; however, it is quite a complicated document and difficult to compare across the years. One has to go to different resources to find the previous statistics to try and compare across the years.	p20	2nd
Usefulness of filters	The Program(s) filter is very innovative. However, the program filter could have some implications, such as isolating and excluding a particular group of students that always rate instructor low, and make a case to the University to justify why that is the case. On the other hand, it may also assist the instructor in cherry-picking only the groups with a high rating and present it to the University for promotion.	p3	1st
	Program (s) filter is really handy; to be able to sort students by their programmes particularly helpful for me, because I teach a first-year course, which is quite large, and students are coming from a wide variety of programmes.	p6	2nd
	I like the fact that one can either hit aggregate or compare papers And programs.	p11	2nd
	Filtering by programs is very innovative; however, it may create room for manipulation. For instance, a teacher can identify those programs that rated high and those that rated low, to get a hint of those to exclude from his workshops.	p16	2nd
	I can see the application for this, especially when you do have students from different programmes taking the same paper or course.	p17	1st
	I think if you are lecturer teaching more than one paper, or taking a paper that has students across two or more different programmes, it will be very useful to see what you're doing well, and what you are not doing well.	p18_19	1st

Continued...

Theme	Response	Participant	Iteration
	From the point of individual teachers time, this dashboard could be useful forever; I mean for individual categories of teachers that want to filter down to programmes. However, for privacy concerns, the programmes filter should have strict rules around not drilling down to groups that are too small. Hence, it will be a good idea to create a rule to turn off the programmes filter features when the groups that are too small, this would prevent easy identification of individual students in smaller groups.	p22_23	2nd
Usefulness of visualisation	very intuitive	p2	1st
	The dashboard is useful.	p4	1st
	I think it would be very useful	p7	1st
	The dashboard is useful, the dashboard is great.	p9	2nd
	I like the idea of being able to show trends over an extended period of time how things are progressing.	p13	1st
	I Like it's very intuitive.	p15	1st
	So I think it is useful.	p16	2nd
	I can see this really is useful to see trends over the years.	p17	1st

End of Table

Appendix AI

THEMATIC ANALYSIS FOR THE USEFULNESS OF THE WORDS AND PHRASES DASHBOARD

Theme	Response	Participant	Iteration
Evaluative Comments	It will be more useful for quality of teaching if the dashboard could count only the top evaluative words like good, not good, different, useful, in order to track the evaluative comments. I do not find it useful to make a case for promotions; most of the information has to do with the class's content rather than promotion.	p1	1st
	I may be interested in the verbs and adjectives from the Parts-of-Speech visualisation because those are words that might be related to evaluations.	p4	1st
	It's a lot of information, it gives me information about the student's perspectives, but it does not give me any information about evaluative comments, so I do not find this dashboard particularly useful. It would be great for the program to be able to sort out evaluative comments, like what they thought of the course. And there is also a risk of taking words out of context and assume things that are not intended.	p12	1st
Useful for a Large Class	This will be useful for large classes rather than small classes	p2	1st
	This would be helpful for the teacher, but it will work well for programs that have larger students	p3	1st
	it will be very very valuable for lecturers that handle many students, however, when there are few students coupled with a low response rate of about 20%, then there is no need for this kind of analysis.	p9	2nd
	It will be good for massive number of students.	p11	2nd
	If you were trying to analyse a very large class over a period, I guess, the word clouds will be helpful to see how things changed. And especially if you were looking at the content of what you were doing, then you could see whether they are being reflected. For example, if the XY framework is really important in what you are doing and it shows up in your dashboard, then that is really helpful.	p21	1st
Useful visualisation	This is a good way of representing the words that occurred in the students' comments to find out what possible themes appeared most on students comments.	p5	2nd
	The idea of bringing all similar comments together on a comparative basis would be very useful	p7	2nd
	I think this is interesting and useful for searching for particular terms. This is a powerful tool.	p8	1st
	I like the fact that you can bring up the comments on the entirety to see what they're talking about; this enables one to drill down.	p11	2nd
	I am definitely impressed. I have always wanted to make sense of my text data. This dashboard will make my life as a researcher a lot easier.	p16	2nd
	I have always loved to see how to make sense of text data in my own course, so I can see some trends in my text data	p14	2nd
	It brings out students' voices.	p18_19	1st

Appendix AJ

THEMATIC ANALYSIS FOR THE USEFULNESS OF THE NAMED ENTITIES DASHBOARD

Theme	Response	Participant	Iteration
Useful	Might be useful for some teachers who want to carry out analysis in linguistics or a language teacher.	p3	1st
	I am not so familiar with this visual way of representing connections between words, I have never seen anything like this before, but it does make sense to me even though this is the first time I have seen this kind of thing, its quite intuitive.	p6	2nd
	It is powerful	p8	1st
	The idea is very valuable,	p7	2nd
	I can see the combinations as more useful, if you are looking at a lot of years, that is quite encapsulating.	p12	1st
	I think it's useful. I see it as it keeps a pictorial view of things that occur together. that is useful.	p14	2nd
	This dashboard could be useful.	p17	1st
	the kind of comments one gets, will determine what can be drawn out. For instance, if a lecture was a big bubble in the diagram, you might think that you either are doing lectures very well or not well at all.	p15	1st
	it may bring out peoples habits, for example, if the lecturer comes late to lectures every time, then Dr Spoke may correlate with late.	p18_19	1st
	So this will be very, very useful, but also gain something that I would need to play around with to really be able to know exactly how useful this interface would be.	p20	2nd
	With sufficient explanation, I can see it could be a useful tool. It will be particularly useful for people that deal with a very large volume of comments.	p22_23	2nd
Not Useful	Did not find it useful.	p2	1st
	I am not sure I will use it.	p12	1st
	If this was my subject, and I was looking at this, it might be quite difficult to figure out what actions I can take from the singular word popping up in that bubble.	p15	1st
	I don't understand the concept of this, but it's very impressive.	p16	2nd
Not Applicable	You may need to provide some examples on how this dashboard could be useful.	p2	1st
	Do not understand how it would contribute towards helping the teacher performance evaluation and improving teaching.	p4	1st
	can not find how it can be applied to improve teaching.	p5	2nd
	But I am not sure you can use a programme to analyse the intricacies of a language, but I do see the value if we can actually do that.	p7	2nd
	I will consider using it, but I would have to understand how it can be applied to improve my teaching.	p10	2nd
	I find it difficult to figure out how it can help the teacher get valuable information. For example, if it could help teachers identify different concepts related to the topic of discussion.	p17	1st
	It would take a while to think about the nature of the comments that would be intuitive to understand own specific subject area, as to the types of things that are useful to my teaching. We were interested in this particular type of data because we want to be making changes to our teaching, either to how we are performing, ourselves or to the material that we are presenting to the students.	p20	2nd

Appendix AK

THEMATIC ANALYSIS FOR THE USEFULNESS OF THE CLUSTER DASHBOARD

Theme	Response	Participant	Iteration
Not useful	Not much useful, but could enable one make claims about what part of the teaching needs to be improved and what part of the teaching is good or bad	p1	1st
	May not be too useful	p3	1st
	I am trying to think of a scenario where I would like to cluster students comments into groups.	p6	2nd
	I do not quite understand its use	p9	2nd
	it does not really tell me anything about what the students experienced in the course.	p12	1st
Useful	Found the clustering analysis useful, a teacher could be able to use this kind of analysis to make meaning from the qualitative comments a large number of people are making.	p2	1st
	This is quite interesting, and it looks like it could be valuable.	p4	1st
	This is useful. It is a unique idea	p5	2nd
	I think this looks really useful. I have never considered doing this before, probably because it has never been easier possible for me to do.	p6	2nd
	This could be a different way to think about the comments	p11	2nd
	It not clear to me how I would use this data	p13	1st
	I think this dashboard is useful. However, students that do not follow the rules of grammar, how does it affect your programme.	p14	2nd
	Again, very, very useful. This dashboard would be the sort of thing that would be useful to have on something like a discussion forum. So in my teaching, we are starting to look at using discussion forums with remote teaching. Using cluster visualisation would be quite a good way to work out common questions amongst students; if a teacher has got much feedback and has no staff to assist in dealing with reading through the comments and trying to work out what are the common themes to come up with answers that can address the majority of people's question. However, one of the difficulties is that the students often have difficulty structuring their questions, so they don't know what they don't know. Having this sort of interface; to have keywords to cluster decision options students have for certain types of problems will be useful. It would enable the teachers' workout which students have managed to understand the words that we regularly use during the course and those who do not understand how to structure their questions. Therefore, separating those top students who know what they are doing from those struggling with structuring. This separation will enable the teacher to know how to address their questions, such as answering top students differently, compared to those struggling with structuring the question itself and will need a lot more description of any solution teacher can offer.	p20	2nd
	I think the idea is really useful; this dashboard provides a nice way of exploring the evaluation questions and provides a useful way to analyse the data, especially for your courses.	p21	1st
Data Related Issues	The choice of cluster terms could make this useful or less useful. And I guess you will have to look at comment data over a long period of time to figure out how to set up the clustering algorithm to become more useful, but I can see that clusters and connections could be useful, but it would not show the words it is not set to show.	p12	1st
	I guess, depending on the clusters and the amount of data, one might be able to find something useful from it.	p15	1st

Appendix AL

THEMATIC ANALYSIS FOR THE USEFULNESS OF THE SENTIMENT DASHBOARD

Theme	Response	Participant	Iteration
Useful	More useful than the other dashboards	p1	1st
	I find it useful for teachers.	p3	1st
	This dashboard is the most valuable.	p4	1st
	This is the most useful. This dashboard could help the teacher know his strengths and weaknesses. What would be most valuable for me is identifying the problems; a teacher always wants to see what aspects to improve. This dashboard will be most valuable to identify what kinds of problems exist. It will also be a good idea to have a dashboard that combines the clustering Algorithm with the Sentiment Algorithm to tell me what areas I need to improve in groups/clusters). For example, it should take the negative comments and cluster them into themes or take the positives and cluster them into themes of goods or jobs well done. This dashboard can further issue advise to the teacher and highlight areas of good performance, and areas of improvement; to tell the teacher the group of things they might need to improve upon from the negative comments as well as the group of strengths from the positive comments; to show what themes have emerged, to know what needs to be improved, such as the structure of the course or if it is the articulations that need to be improved. As a lecturer, I look at how I can deliver value-added services in the form of things that my students say I need to improve upon.	p5	2nd
	I think this is a really powerful tool.	p6	2nd
	This could be very useful. It will be a good idea if this dashboard is able to pickup or detect personal comments	p7	2nd
	So this would be really helpful, particularly to get an overall feeling for how much the class was positive about your course.	p11	2nd
	Seeing the positive and negative comments will help in enhancing ones teaching.	p13	1st
	This dashboard is really valuable. All of the quantitative dashboards you have are valuable, but I think this one appears to be the most useful.	p15	1st
	This would be useful.	p17	
	I think that is very useful. The negative feedback can affect some aspects of change in some areas of teaching, say lecture delivery. This dashboard can also help to know if students have been consistently saying a particular thing. For instance, two or more people saying one negative thing about the paper on one occasion, or consistently saying the same negative point every time, or saying different negative things.	p18_19	1st
	This dashboard provides the most helpful information. This information gives you a clue as to what you could do next year differently than this year and what you might need to drop or what you might need to add	p21	1st
Diversify Attention	I think that's very human nature to focus in on the negatives (because when I read my comments, I tend to focus more on the negatives leaving the positives out). But having a great tool like this can help one balance out the positives and negatives. It gives you a feel of the overall sentiment, rather than just focusing on the negatives.	p11	2nd
	I really sort of applaud, the direction you're going in this because I think people, probably don't use evaluation data enough to improve their teaching, or if they do, it's very easy to get blindsided by one or two particular comments without having a sense overall of what the data is actually showing.	p13	1st
	Lectures often tend to focus on the negative comments despite the positive ones. This visualisation has a way to make teachers look at the entirety rather than just focusing on the negatives.	p15	1st
	I think the nice thing about this dashboard, is when reading through the comments, teachers can be discouraged by the first few comments they see. Moreover, I remember one year the first comment was negative, and it just made me feel bad immediately. We know that these things can not be taken too personally, of course, not with a large class. However, regardless, the first comment was negative. Whereas if I can filter, I can look at what went well first, giving me much more resilience to take on board the things that did not go so well and to be more constructive about responding to those things.	p20	2nd

Continued...

Theme	Response	Participant	Iteration
Algorithim Trust and Validation Issues	I find it useful if it has a lot of data and if the algorithm behind it cab be trusted	p2	1st
	It will be interesting to drill down into the positive and negative comments; it will enable me to judge for myself whether the system is interpreting the comments correctly or not.	p4	1st
	I would like to be aware of how accurately the software can do this (a percentage of uncertainty would be useful to know).	p6	2nd
	Teachers can use the sentiment data to justify so many things, but it will be good to test the algorithm that performs the sentiment.	p8	1st
Promotions	I will be wary of using the sentiment index for promotion. For instance, teacher 'A' has a sentiment index of 0.7; it should get a promotion and teacher 'B' has a sentiment index of -0.3; it should not get promoted. I think technology probably is not quite at the level we should use sentiment for promotion purposes.	p9	2nd
	I can sort of say that this would be quite useful, particularly to serve as evidence of your teaching for promotion purposes.	p13	1st
	To help in situations where there are too many negatives, how would this dashboard distinguish between negatives that are not the teacher's fault and those that are? It does not affect promotions, for example, too many negatives that were not the teacher's fault.	p14	2nd
	So for those people who are on perhaps confirmation pathway who are needing very specific things, such as promotions, would find this dashboard very, very useful for them because they need to be able to really break down their feedback far more than perhaps somebody who is really just looking at it for refining their teaching or feedback on content.	p20	2nd
Large Number	This will be helpful particularly for people that take large classes. How many people would read through hundreds of comments?	p11	2nd
	I think this would be an extremely useful tool because I tend to find that I do not often have much time to go through all the comments from my evaluations. Occasionally, I scan through and highlight a couple of positive and negative comments to attend to later to make some changes to my course. However, I have a massive class of about 500 to 800 students, and time is never on my side. Dashboards like this would be absolutely magic for identifying particular problems that I know come up on the course, such as finding keywords very easily in the comments, seeing how people respond to particular problems already identified, or particular areas that may have changed. So I might specifically be asking for comments in that area.	p20	2nd

End of Table

Appendix AM

THEMATIC ANALYSIS FOR THE USEFULNESS OF THE TEACHER'S EVALUATION DASHBOARD

Theme	Respnse	Participant	Iteration
Useful	Generally, there are some useful parts and some not useful part. - There are parts I could see myself using to interpret my data and be a better teacher, and there are also parts that I am not sure if they are relevant.	p4	1st
	I think it would be very useful	p7	2nd
	it is a valuable approach, and the use of machine learning makes perfect sense. We can use it for knowing the kind of questions that we really do not know, such as, did the students feel like they had sufficient background? Was I teaching to their background? Was I teaching it to the right level? Was it too much work? Do they feel that they had to do a lot more work for my paper than somebody else's paper? Or did they feel that it was easier or harder than somebody else's paper? Those kinds of relatively valuable feedback a teacher could pull in using this tool.	p9	2nd
	I think that could be very useful, and particularly professional programmes where it's really important that students have a comprehensive introduction to course content, which in itself can really be a problem in terms of accrediting them or assuring that they're qualified on graduation. So this could be really useful for drilling down and figuring out where there are holes or issues.	p12	1st
	I think it's useful.	p17	1st
	I think it is really good and I can see this dashboard looks really promising.	p18_19	1st
Improvement compared to current System	I am experienced, and I have been evaluating my teaching for a long time now using what is already existing (I can just read all the comments because they are not much). So I am set in my ways, and I do not need a dashboard to do that.	p3	1st
	I think this is a huge improvement over the way we currently view students evaluation results. Being able to compare student responses from different student programmes easily is very useful, especially for classes which have students from a wide variety of programmes (such as the courses I teach).	p6	2nd
	I think it's useful. I have been co-teaching on the teacher training with XXX, and each time we start the programme, we have to go through the troublesome process of scroll through previous evaluations to pull them together. There is no way of actually putting them all together without going through the whole laborious task of reading every comment and making notes and then connecting it. However, with a user-friendly tool like this, it would be so much easier for a teacher to plan their next session without having to scroll through and finding loads of previous PDF files that have been sent and having to find them. Having ones SET data sitting in one place and then getting access to data using dashboards that can perform other functions like comparing SET scores over the years makes a big difference.	p17	1st
	So initially, when I first started teaching, sometime in 2012. I got the raw forms returned to me, and it was not typed up in any way. Currently in is typed out and sent out in a PDF or CSV format and I have to read them. So, at least, they're easier to read off than the handwritten ones. I would rate this system exponentially higher than the current evaluation system.	p20	2nd
Number Rating Vs Comments	I find the qualitative comments more valuable than numbers.	p7	2nd
	I think most lecturers will go for the Open-ended dashboard more than the Student Rating dashboard because it seems a bit more intuitive and it's also harder and time-consuming to perform analysis with text data. -Also being that our institution already performs some sort of quantitative analysis on the numbers, so more teachers will be interested in the text data.	p8	1st
	I can see this will be very useful for those that are taking papers that require lots of students and also for those that oversee programs.	p12	1st
	This dashboard is what speaks to me; I think the qualitative comments are more valuable than just the number ratings. I find that even now, when we get student evaluations, like people have clicked on certain things, but then you go and read the comments and you realise that they don't always match up to the ratings. When you read the comments, it is quite clear, what actually was the issue.	p17	1st
	Oh, look at this, particularly for the comments section on the types of evaluations that I perform. This kind of dashboard is just this is straight ahead of what I need.	p20	2nd

Continued...

Theme	Respnse	Participant	Iteration
	Improving teaching with the std five questions by getting good scores is not helping; it is the text that gives one something about the lecturer's presentation. Even during peak lectures, some students still find it boring or think it was a poor lecture. However, some other students may find it fantastic, really good. The difference is how they see it, and there is not a lot the teacher can do about that. Nevertheless, if many students said that the lecture was disorganised and did not understand what the teacher was talking about, that would be helpful.	p21	1st
Prompt students to write	This dashboard would also work well if a teacher gives students a list of words or list of phrases that they can incorporate in their comment, so that they would actually produce something that will be consistent across the comments, even though the danger of doing that is that we might be limiting the range of responses.	p7	2nd
	And the other thing about knowing that I have got something like this to use, it would allow me to prompt the students to use certain words when making comments, and that way, the system would find those comments very easily, such that I would then be able to search. For instance, a teacher who wants the students to provide feedback on how you felt about the changes to' comparison mean' could prompt them to use the word' comparison mean' in their comments. Making it easy for the teacher to use the word, 'comparison mean' as the search term to find all the comments associated with the phrase 'comparison mean', would make it very useful. Particularly, if I was to add that sort of collaborative notes to inform the students about how I make the best use of their comments.	p20	2nd
	This could be helpful if you are able to persuade students to write. You ask them to put some really comprehensive discussion, and that would be a useful thing.	p21	1st
Age	This dashboard necessitates a certain degree of Computer competence, for someone of my generation, we are far more used to a paper project. I think this is a future-oriented thing, and probably, it will become more valuable as the years go by, and people become more and more attuned to it, in that respect, I think it is valuable. Older lecturers may struggle to get a hold of it.	p7	2nd
Gaming the System	There are two uses of course evaluations, and we only see one of them applied effectively in the institution, which is using it to keep the staff in check. Course evaluations are like the grade-books for teachers, and it is common to think about course evaluations as something that will positively contribute to teaching. I think it makes perfect sense for us (the teachers) to figure out how to improve things. In my view, to make this dashboard more useful as a personal tool for the teachers; then teachers will have to perform this analysis independently of the department's head. This separation will also avoid teachers who did not get a perfect rating, trying to game the system.	p5	2nd
Data Access	it will be good for lecturers to access their own data, it is not easy to answer if this dashboard is useful on someone else's teaching data.	p1	1st
	testing this dashboard with real data will be a great idea	p3	1st
	And I would be very happy to have access to this kind of software to look at my student evaluations.	p6	2nd
	These dashboards might not fit the kind of feedback I have looked at in the past, and how I have learned from that feedback. I can see the potential, but I guess I would love to be able to play with my own data, to be able to say how useful.	p13	1st
	So this dashboard is something that I would be interested in using to visualise how students rate my teaching performance in my course	p14	2nd
	It will be quite interesting to get these aggregate data to explore my workshop. I am very curious to use this dashboard to explore my own data	p16	2nd
Data Fusion	There are possibilities of comparing the quantitative with the qualitative.	p3	1st
	it will be a good idea to be able to do a validation of the comments versus the grades.	p7	2nd
	It will be good to compare the students' number ratings against students pass/fail rates. For instance, A students compared to B students, compared to C students and those that failed. (It will be useful to filter out to see how those two are associated). Because it is perceived that teachers that teach a course too hard might get low ratings compared to those that teach a course to easy.	p9	2nd
	From my perspective, if this dashboard could provide a means to anonymously match examples of people who perform very well and examples of people that perform poorly. This kind of information would be helpful for teachers to compare anonymous performances and trends. For instance, to examine what the third-year papers look like compared to second-year papers to see if they are the same patterns in all of them or if there are differences. as well as how the students responded.	p21	1st

End of Table

Appendix AN

THEMATIC ANALYSIS FOR THE USABILITY OF THE TEACHER'S EVALUATION DASHBOARD

Theme	Response	Participant	Iteration
Satisfaction	The variety of visualisation options for the data was amazing. I have never seen some of the types of graphs/diagrams that were used to display the data, but nevertheless, they were intuitive and easy to understand.	p6	2nd
	The interface design is okay. It's very clean	p8	1st
	This looks great.	p11	2nd
Renaming Titles and Labels	Put some headings to make the visualisations self explanatory.	p1	1st
	I think it probably might be just a few clearer signposts, that need to be renamed.	p4	1st
PDF	It will be a good idea to save file in pdf format	p5	2nd
	Can you get it into a pdf format?	p22_23	2nd

Appendix AO

QUESTIONNAIRE TWO

Teacher Analytics Dashboard Usability Study

Dear participant,

I am carrying out a usability study to determine the usefulness of the teacher analytics dashboard TADB.

This project has ethical approval from the University of Otago with reference number 19/097.

You are invited to spare 10 minutes of your valuable time to help with this study and contribute to how the students teaching evaluation data is presented at the University of Otago.

I think that I would like to use this dashboard frequently.

- ○ Strongly agree
- ○ agree
- ○ Neutral
- ○ disagree
- ○ Strongly disagree

I found the dashboard unnecessarily complex.

- ○ Strongly agree
- ○ agree
- ○ Neutral
- ○ disagree
- ○ Strongly disagree

I thought the dashboard was easy to use.

- ○ Strongly agree
- ○ agree
- ○ Neutral
- ○ disagree
- ○ Strongly disagree

I think that I would need the support of a technical person to be able to use this dashboard.

- ○ Strongly agree
- ○ agree
- ○ Neutral
- ○ disagree
- ○ Strongly disagree

I found the various functions in this dashboard were well integrated.

- ○ Strongly agree
- ○ agree
- ○ Neutral
- ○ disagree
- ○ Strongly disagree

I thought there was too much inconsistency in this dashboard.

- ○ Strongly agree
- ○ agree
- ○ Neutral
- ○ disagree
- ○ Strongly disagree

I would imagine that most academics would learn to use this dashboard very quickly.

- ○ Strongly agree
- ○ agree
- ○ Neutral
- ○ disagree
- ○ Strongly disagree

I found the dashboard very cumbersome to use.

- ○ Strongly agree
- ○ agree
- ○ Neutral
- ○ disagree
- ○ Strongly disagree

I felt very confident using the dashboard.

- ○ Strongly agree
- ○ agree
- ○ Neutral
- ○ disagree
- ○ Strongly disagree

I needed to learn a lot of things before I could get going with this dashboard.

- ○ Strongly agree
- ○ agree
- ○ Neutral
- ○ disagree
- ○ Strongly disagree

Appendix AP

THEMATIC ANALYSIS OF THE GENERAL COMMENTS

Theme	Interpretations	Example of Quotations	Partici-pant
Student Feedback	Participants argued that data fusion or linking student attendance, performance, engagement data with SET data could support the teachers to make informed-decisions in new ways	The online evaluation system now closely mirrors the conditions of anonymous online comment. These conditions seldom produce thoughtful feedback and they far too often invite vengeful, abusive and discriminatory comments.	p15
		I think it's very important to be able to match responses in different questions. e.g. if I ask how many lectures the student attended and how effective I've been as a teacher, I want to be able to pair the responses. I suspect that students who attend very few lectures tend to rate my effectiveness lower than the students who come to class this is important to know when assigning credit for good teaching.	p34
		The quality of the data is so bad that I do not consider it worth interpreting. The only useful part is the comments section, which sometimes provides insights but usually demonstrates that the respondents have not been attending classes.	p39
		Most data is biased in ways that make it unhelpful. Would be good to separate bias from constructive feedback.	p47
Promotion	Participants also claimed that their potential manipulation of the evaluations process by academics and it is important to have multiple point of feedbac	it has to be triangulated to be valid. It is a human factor and skills to get good scores. I know exactly wat to do to manipulate and get high scores- be extremely -(pretend)nice to students, pretend you care and you get good scores!	p7
		I think the XXXX know the surveys are flawed but when used for promotion, its the best option. If it wasn't being used for promotion, just to improve teaching, then the academic literature points to many other ways to evaluate teaching. It's not an unknown!	p28
Improve teaching quality	Participants believe that qualitative comments contribute to improving teaching performance compared to the quantitative Likert scale items.	Being able to interpret teaching evaluation data is a difficult but important part of improving teaching quality, and closing the feedback loop.	p56
		It is complex and nuanced. I like qualitative data over quantitative data.	p32
		I think there should be far less reliance on student teaching evaluations, and greater reliance on other forms of feedback and reflection.	p21
Low Response	data can be misrepresented another wll like teaching dashboard to address issues with regards to data misrepresentations	The student evaluation form summaries often present data as percentages. This is very misleading when class numbers are so small or when the proportion of students who complete evaluations is so low.	p36
		There is no official guide of what counts as 'good' in the evaluations or any guidance on how people should use the information provided. I've figured it out on my own, but I talk to senior colleagues who are baffled by it. It is also hard to get a good response from students since moving online. I have gone from 80-90% responses to 10-25% responses, and that is considered good amongst my peers. It raises the question as to whether the data is valid, or even worth interpreting as you tend to get responses from polar opposites, those who love you or hate you.	p38
		For promotion/hiring/evaluation purposes there needs to be more attention to sample size representativeness. The response rates are so low and I've found them a bit bimodal, mostly content students but a few unhappy students. This can give a really biased view.	p45
Problem with current evaluation System	Some issues have been raise about the current online student evaluation system	Online surveying is USELESS I got far better response rates under the old system [p29]. -	p29
		Would like to see a better offering of tools by Otago.	p33
		How to get students to actually fill them out?	p43

Continued...

Theme	**Interpretations**	**Example of Quotations**	**Partici-pant**
		I have a big problem which is procedural. Online forms of student evaluation are much much less successful at capturing a significant, representative proportion of the class. Interpretation has no value for data that does not capture a significant, representative proportion of the class. Since XXXX now only support online systems I run my evaluations on paper outwith the XXXX system. I will continue to do this until XXXX returns to supporting paper evaluations or demonstrates an effective way to increase participation in online evaluation. This is not me being a dinosaur- it's simply how students operate.	p40
		I think the current set of core teacher and core paper questions could be improved.	p41
		The form in which we collect data limits the value of these in improving our teaching.	p44
		The problem with teaching evaluation data is not its interpretation but the questions and how the data is gathered.	p53
		Will your dashboard help teachers who get only low % response rates to student feedback forms to understand how to interpret partial data with statistically informed confidence? If not, what is the point? Will your dashboard help those who ask a colleagues to peer review some aspect of their teaching to present their findings with some sense of humility, integrity and clarity? If not, what can we do about it?	p57

End of Table

Key: XXXX represents name of a department, faculty or institution.