

Tax Compliance in the New Millennium: The variables of understanding

Sue Yong^{*a}, Karen Lo^{**}, Brett Freudenberg^{***a}, and Adrian Sawyer^{****}

* AUT University (email: sue.yong@aut.ac.nz); ^a : contact author: sue.yong@aut.ac.nz; tel: +64 99219999 extension 5721; ** AUT University (email: Karen.lo@aut.ac.nz); *** Griffith University (email: b.freudenberg@griffith.edu.au); **** University of Canterbury (email: adrian.sawyer@canterbury.ac.nz)

Abstract

Taxpayer compliance is a critical part of any tax system, as without it the integrity of the tax system can be undermined. Previously, Jackson and Milliron published a comprehensive study reviewing and synthesising the findings of tax compliance studies conducted prior to 1986; and highlighted the variables, the methodologies and theories that had arisen. Richardson and Sawyer built upon this research by synthesising the work from 1986 to 1997. This current research seeks to build upon these two important studies and extend our knowledge by analysing tax compliance studies from January 1998 to 31 December 2017.

This analysis was facilitated by using text mining software, Leximancer, to visualise the frequency and co-occurrence of concepts and themes related to tax compliance factors. Data was drawn from 717 articles obtained from 15 research databases; and given the size of the data, this article is the first article of three, and focuses on the ‘variables’ that have been researched over the 20 year period.

Keywords: tax compliance, evasion, morale, compliance cost, Leximancer

Section One

INTRODUCTION

Taxpayer compliance is a critical part of any tax system, as without it the integrity of the tax system can be undermined, additional resources expended by revenue authorities and insufficient tax revenue raised. Governments around the world are confronting the problem of raising sufficient tax revenue to fund their public programs. Politically there can be resistance to just raising tax rates, and governments are becoming more acutely aware how there can be tax leakage given the perceived non-compliance by taxpayers. Taxpayers of concern may be multi-national corporations utilising transfer pricing strategies to small micro-businesses participating in the cash economy. Recently, it has been estimated that the cash economy can range between 3 to 15% of gross domestic product,¹ and the World Bank study estimating the ‘shadow’ economy in 2007 resulted in estimates of: United States of America (USA) 8.4%; Japan 10.3%; New Zealand 12.0%; China 12.1%; United Kingdom 12.2%, Australia 13.5% and Canada 15.3%.² If these estimates are accurate, the amount of tax revenue potentially being lost could be substantial. Additionally, more governments, especially developed nations, are relying on taxpayers to self-assess their tax liability, and for such systems to be able to operate they require a sufficient level of compliance by taxpayers. Accordingly, tax compliance is an important concept to consider, and is part of the reason that it has been researched through a range of disciplines from psychology to economics, and accounting.

There have been a number of theories postulated to try to explain why taxpayers may or may not comply, extending from the deterrence theory³ to motivational postures.⁴ It is likely that no one theory can completely explain the complex relationship between

¹ The Senate (Economics Reference Committee). (2004). ‘The Structure and Distributive Effects of the Australian Taxation System’. Commonwealth of Australia, Canberra, p 62.

² Friedrich Schneider Andreas Buehn Claudio E. Montenegro, (2010), “Shadow Economies All over the World New Estimates for 162 Countries from 1999 to 2007”, Policy Research Working Paper 5356, The World Bank Development Research Group Poverty and Inequality Team & Europe and Central Asia Region Human Development Economics Unit July 2010.

³ Nigel Walker, *Crime and Punishment in Britain*, (Edinburgh: Edinburgh University Press, 1965).

⁴ Braithwaite, V. (2017). ‘Tax and democracy: Bygone ideas or time for reinvention?’ Presentation at the 29th Australasian Tax Teachers Association Conference, Masterton, Victoria University of Wellington.

taxpayers and the tax system, as there are many types of taxpayers from salary and wage earners, to high wealth individuals and those conducting enterprises of various sizes, as well as taxpayers with different cultural and religious backgrounds, literacy, and age. Also, the tax culture of a jurisdiction may influence tax compliance, with the presence or absence of trust and/or corruption potentially influencing compliance behaviour.⁵ Furthermore, it could be that these factors alter with time, as there are changes in technologies, working habits and our understanding of tax compliance. In fact since the 1960s there have been many studies around the world exploring tax compliance in various jurisdictions. In isolation these prior studies have been important in illuminating various factors about why people do or do not comply with the tax system. However, what observations are available if we consider these prior studies as a whole? It is argued that such aggregated reviews of prior studies are important in illuminating important trends and commonalities.

To gain an understanding of our knowledge of tax compliance, Jackson and Milliron published a comprehensive study reviewing and synthesising the findings of mainly USA tax compliance studies conducted in the 1970's to the end of 1985.⁶ In this review, they identified compliance variables, the methodological issues and the theoretical areas considered. This initial study was then built upon by Richardson and Sawyer who analysed subsequent tax compliance research from 1986 to the end of 1997, to see to what extent the earlier identified compliance variables, methodological issues and theoretical areas did or did not continue to be found in studies from 1986 to 1997.⁷ Through this study, Richardson and Sawyer not only considered more countries in their analysis, they also identified new emerging areas of tax compliance research including the role of tax preparers and inducements.

Since 1997 there has not been another comprehensive analysis of tax compliance studies to consider what have been the emerging trends, and to what extent has our knowledge improved in terms of tax compliance. Given the growth of international

⁵ Faridy, N, Freudenberg, B, Sarker, T. (2016). "They are a tax fearing people": Deterrent Effect – Penalties, Audit and Corruption in a Developing Country. *Curtin Law and Taxation Review*, 3: 1-50.

⁶ Jackson, B. R., & Milliron, V. C. (1986). Tax compliance research: findings, problems and prospects. *Journal of Accounting Literature*, 5, 125-165.

⁷ Richardson, M., & Sawyer, A. J. (2001). A Taxonomy of the Tax Compliance Literature: Further Findings, Problems and Prospects. *Australian Tax Forum*, 16, 137-322.

trade, technological developments and sharing of information it is timely to for this thorough analysis to occur.

The aim of this article is to provide a comprehensive overview of the different tax compliance factors (concepts) that had been researched for the past twenty years (from January 1998 to December 2017) since the seminal papers by Richardson and Sawyer⁸ and Jackson and Milliron.⁹ This current study is a sequel to these two earlier studies, and given the 20 year time span and the size of the data, this article is the first article of three, and focuses on the ‘variables’, with the other forthcoming articles considering ‘methodology’ and ‘theories’ separately..

We reviewed tax compliance studies and present content analysis of the selected articles to map the different factors/concepts that have been researched from 1998 to 2017. In this way we can determine whether the tax compliance factors (variables) identified prior to 1998 remain important, and whether new tax compliance factors have emerged. Data was drawn from 717 articles obtained from 15 research databases pertaining to taxation, accounting, social sciences, law, economics, business, political sciences and psychology journals. Given the large corpus of data analysed in this study, 717 articles, a computer-aided program for content analysis, namely Leximancer was used. Leximancer was used to assist in visualising the frequency and co-occurrence of concepts and themes.

This study offers a starting point for academics to be directed as to what has been researched and what needs to be researched further to extend our understanding of tax compliance in the new millennium. Also it provides those in tax policy a comprehensive foundation as to our understanding of tax compliance.

In considering the tax compliance research over the last twenty years it appears that while the 19 variables previously identified continue to be considered in tax compliance studies, new variables/concepts have emerged. In particular, there are 19 new variables emerging such as tax evasion, tax enforcement, trust, power and sole traders. This analysis also demonstrates an apparent connectivity between some of the

⁸ Richardson, M., & Sawyer, A. J. (2001). A Taxonomy of the Tax Compliance Literature: Further Findings, Problems and Prospects. Australian Tax Forum, 16, 137-322.

⁹ Jackson, B. R., & Milliron, V. C. (1986). Tax compliance research: findings, problems and prospects. Journal of Accounting Literature, 5, 125-165.

concepts, such as that tax evasion connectively with the concepts of perceived opportunity, tax avoidance and tax perception.

Additionally, this analysis provides some insights into the gaps in tax compliance literature, as well as providing a pool of resources for new and experienced tax compliance researchers. With such a thorough analysis, it adds richness to our comprehension of tax compliance. Such findings build a strong foundation for future research as we continue to explore tax compliance in the new millennium.

Section Two of this article will provide a broad summary of the prior studies conducted by Jackson & Milliron and then Richardson & Sawyer. The third section will then provide an outline of the research methodology undertaken and explain how Leximancer can be used for data analysis. Section Four will then provide the results. Through the analysis of the results recommendations will be proposed, with future research being outlined in the fifth section of the article before concluding.

Section Two: Literature Review

When a study is itself a literature review, then the ‘literature review’ could be viewed as being redundant. In our view it remains important to provide an overview of prior literature reviews that have offered insights into what we know (and do not know) about taxpayer compliance.

Over time there has been a greater sophistication of our understanding of compliance, including cross-disciplinary research. For example, while income levels may be related to tax compliance levels, it may be that it has more to do with ‘types’ of income, as salary and wage earners have less opportunity to evade due to the withholding of taxes compared to other types of income (such as capital income). To consider how our understanding of income tax compliance has developed the two seminal papers by Jackson & Milliron (1986), and then Richardson & Sawyer (2001) are discussed.

Jackson & Milliron: 1970’s to 1985

In an attempt to aggregate our understanding of tax compliance, Jackson & Milliron (1986) reviewed 43 income tax compliance studies which mainly focused on the USA. Through this they synthesised 14 variables, and four methodological approaches from the studies examined.

This aggregation of studies demonstrated why there could be differences in research findings. For example, Jackson & Milliron (1986, p 138) when discussing ‘fairness’ note that a ‘more careful definition of the variable may be useful’; as the cause of the differences in the research studies could be the different definitions and then how it is then operationalised in studies.

Jackson & Milliron (1986, p 144) observed that studies demonstrate that those who use professionals to help prepare tax returns can have greater audit adjustments, as they are more aggressive with their tax planning and use of ‘grey’ areas. Jackson & Milliron (1986, p 156) concluded their study with the observation that ‘the state of tax compliance understanding can only be described as primitive’, and that ‘existing work should be examined critically for flaws’.

Richardson & Sawyer: 1986 to 1997

Richardson & Sawyer (2001) then set to examine the ten (10) years after the Jackson & Milliron study and due to technological developments and increased activity in the area they analysed more articles (approximately 170 studies), and had a greater global mix of studies beyond just the USA. They suggested that the increased interest in tax compliance could in part relate to the politically difficulty of raising tax rates (or to decrease public spending); and thereby if jurisdictions can improve tax compliance they can raise more tax revenue without awkward political decisions (Richardson & Sawyer, 2001, p 138).

To be able to aid comparison Richardson & Sawyer highlighted the importance of actually defining what is meant by ‘compliance’, as well as the difference between ‘intentional’ and ‘non-intentional’ non-compliance. Richardson & Sawyer (2001, p 142) favoured the definition of tax compliance by Roth, Scholz and Witte, (1989, p 21):

“Compliance with reporting requirements means that the taxpayer files all required tax returns at the proper time and that the returns accurately report tax liability in accordance with the Internal Revenue Code, regulations, and court decisions applicable at the time the return is filed.”

However, they observed that many of the studies analysed did not use this definition and some failed to have any definition at all. In this regard, Wu (2011) undertook a review of published research that examined ‘tax compliance’ to ascertain the degree to

which there is an ‘agreed’ common understanding or definition of tax compliance. Wu observes that there is a continuum with respect to the degree of severity of noncompliance, ranging from the unintentional to various degrees of intentional noncompliance. She concludes (pp 147-8, emphasis added):

“The number of studies conducted around the area of behavioural tax compliance and non-compliance has led to a number of different versions of definitions for the relevant concepts; therefore, it is an extremely difficult task to attribute ‘universal’ definitions for each of the concepts at issue. Nevertheless, a reasonable conclusion can be drawn that ‘tax compliance’ *involves taxpayers satisfying the applicable rules and regulations by paying the correct amount of taxes due at the right time*. In addition, there is enough evidence to support the argument on the *importance to distinguish between taxpayers who observe the ‘letter’ of the law, the ‘spirit’ of the law, or both, to determine the ‘degree’ or ‘extent’ of their compliance behaviour*. Although it is again a subjective issue in ascertaining what is intended to mean by complying with the ‘spirit’ of the law, it may be reasonably justified that taxpayers who are ‘committed’ to meeting their tax obligations without enforcement activities by the tax authorities could fit into this particular category.”

The matter of whether there should be a consistent definition for tax compliance remains unresolved. However, as a starting proposition, we agree that all studies need to clearly set out the definition of tax compliance that they have adopted.

Richardson & Sawyer (2001, p 203) in terms of their analysis of the 10 years since Jackson and Milliron’s study observed that:

“... some variables have received significantly greater attention since 1985 (ethics, fairness, income level and the probability of detection), while others have not (occupation and revenue authority contact).”

In term of our understanding how these variables may influence taxpayer compliance they concluded:

“Although the certainty associated with some variables has increased (revenue authority contact and sanctions) or been maintained (withheld income source, ethics and the probability of detection), the certainty associated with other variables has decreased (peer influence and tax rates).” (Richardson & Sawyer, 2001, p 203)

With the studies they analysed, Richardson & Sawyer (2001) compared to Jackson & Milliron (1986) found an additional five (5) variables (Compliance costs, Tax preparers, Framing, Positive inducements, Tax amnesties); one (1) new research method (process tracing), and one (1) new research theory (contingency theory). Table 1 contains a summary of both the Jackson & Milliron and the Richardson and Sawyer studies.

Richardson & Sawyer's (2001) analysis highlights the multi-dimensional relationship between variables, such as gender and education effects on compliance. They also report about non-compliance by peers brings into question about the direction of causality.

One of the conclusions by Richardson & Sawyer (2001, p 224) was that attention needs to be:

“... directed towards studies which develop and test their own hypotheses. ... as it is only through the replication of research findings (across populations, research methods, and time) that an increased level of certainty can be obtained about tax compliance behaviour.”

They also recommended that future non-compliance studies ‘examine separately the two main forms of taxpayer noncompliance — noncompliance through the overstatement of deductions, and noncompliance through the understatement of income’ (Richardson & Sawyer, 2001, p 224).

Considering the other side of the coin, Richardson & Sawyer (2001, p 168) observed that withholding and third party information reporting appears to improve compliance, although this needs to be balance with possible adverse increases tax compliance cost.

They also suggested that future studies need to delineate more for ethics (Richardson & Sawyer, 2001, p 178), given that it appears the ethics of the taxpayer will affect the factors that motivate them, as a less ethical taxpayer could be more influenced by probability of detection (compared to taxpayers with high ethics, who could comply regardless). However, the difficulty they highlighted with this was exactly how to identify and ‘measure’ ethics (Richardson & Sawyer, 2001, p 179).

Since Richardson & Sawyer's published study other researchers have provided ‘partial’ reviews of the literature. This includes Borrego et al (2013); Devos (2014); and Khelif and Achek (2015). Also a number of other studies have included relevant sections of the tax compliance literature as part of their empirical analysis.

Table 1: Aggregated Tax Compliance Studies

Study	Period covered	Number and type of papers	Jurisdictions and study focus	Databases/Source	Variables identified		Methodologies identified	Theories identified
Jackson & Milliron (1986)	1970's to 1985	43 Including working papers and unpublished studies	USA (+ few studies outside USA) Compliance studies of income tax	Not specified	14 Variables described: 1. Age 2. Gender 3. Education 4. Income level 5. Income source 6. Occupation	7. Peer Influence 8. Ethics 9. Fairness 10. Complexity 11. Revenue Authority Contact 12. Sanctions, 13. Probability of Detection 14. Tax Rates	4 Methodological Approaches 1. Surveys 2. Experiments 3. Analytical models 4. Regression analysis	4 Theories identified: 1. Prospect Theory 2. Deterrence Theory 3. Cognitive Structures Theory 4. Agency Theory
Richardson & Sawyer (2001)	1986 to 1997 (10 years)	Approx 170 Only published studies	Multiple-jurisdictional Compliance studies of income tax (inc CGT if CGT include in 'assessable income') Also personal collection. Checked bibliography of gathered articles Physical checks of leading international journals	4 databases identified: 1. ABI/Inform 2. EconLit 3. Social Sciences 4. Westlaw	19 Variables described J & M's 14 Variables: 1. Age 2. Gender 3. Education 4. Income level 5. Income source 6. Occupation 7. Peer Influence 8. Ethics 9. Fairness 10. Complexity	11. Revenue Authority Contact 12. Sanctions, 13. Probability of Detection 14. Tax Rates And then 5 additional compliance variables: 15. Compliance costs 16. Tax preparers 17. Framing 18. Positive inducements 19. Tax amnesties	5 Methodological Approaches J & M's 4 Approaches: 1. Surveys 2. Experiments 3. Analytical models 4. Regression analysis And then 1 additional methodology: 5. Process Tracing	5 Theories identified J & M's 4 Theories: 1. Prospect Theory 2. Deterrence Theory 3. Cognitive Structures Theory 4. Agency Theory And then 1 additional theory: 5. Contingency Theory

Overall, the studies by both Jackson & Milliron and Richardson and Sawyer provide an excellent summation and analysis of income tax compliance studies from the 1970s through to the end of 1997. However, since that time there has been continued interest and need for research. Additionally, with the advent of technology, global trade and changing work and investment habits, it is important for us to understand what studies from 1998 illuminate in terms of taxpayer compliance. The purpose of this study is to add to our understanding of tax compliance from published studies from 1998 through to 2017, with a particular focus on the variables identified. Given the magnitude of the data set, the other aspects of methodologies and theories in terms of tax compliance studies are the subject of other articles.

Section Three: Methodology

This section provides the justification and the process involved in the Leximancer data analysis, and discusses the settings chosen for the purpose of the research. The prior studies of Jackson and Milliron (1986), and Richardson and Sawyer (2001) use a more manual approach to their analysis, given the size of the data set a text mining software, Leximancer, was thought appropriate to assist in the initial analysis.

Leximancer

Leximancer is a text mining software that can be used to analyse the content of a collection of documents, as well as be able to visually display the extracted analysis through conceptual maps. These conceptual maps present the data through main concepts found and how these concepts are then related to each other.

To do this analysis, Leximancer uses empirically validated mathematical algorithms (Metropolis et al, 1953) to make two determinations. Firstly, it determines what are the most frequently used concepts within a body of text (thematic analysis); and more importantly, it determines if there are relationships between these identified concepts (semantic analysis) (Verreynne et al., 2011). By using a mathematical approach it leads to providing unbiased results by performing associational analysis of textual data in an independent manner (Liesch, et al. (2011, p 24).

Smith and Humphreys (2006), have demonstrated the advantage of Leximancer is that it can have face validity, stability, reproducibility, correlative validity and functional validity. It achieves this by employing:

“proximity value to analyse text and learns to automatically identify and map theme and concepts in textual data. The software uses word frequency and co-occurrence data to identify families of terms that tend to be used together in the text. Thus, Leximancer identifies what concepts exist in a set texts, allowing concepts to be automatically coded in grounded fashion.” (Cretchley et al., 2010, p 319).

In this process:

“Leximancer bootstraps an expanded list of associated terms that signify a concept from the textual data. This boosts the recall of relevant contents, and more closely resembles the work conducted by a human coder” (Liesch, et al. (2011, p 24-25)..

Important terms to understand in the use of Leximancer are: ‘concepts’, ‘themes’, ‘conceptual maps’, and within in these ‘circles’ and ‘lines’; each of which is discussed below.

Firstly, ‘concepts’ is used to describe words that occur very frequently. But in considering these ‘concepts’, Leximancer goes beyond merely counting words, and instead uses coding blocks of sentences (generally either one or two sentence blocks). In this way, Leximancer allows us to go beyond a simple keyword search (Smith & Humphreys, 2006, p. 262), and can provide a co-occurrence factors which indicate which concepts are occurring with other concepts at the same time. The stronger the co-occurrence indicates that these factors (concepts) have a strong relationship, and they are usually investigated together. It also allows for compound words to be considered such as “tax evasion” and “tax morale”. For consistency with the prior studies term ‘variable’ will be used a synonym for the ‘concepts’ identified by Leximancer. [Is this okay to do this?]

Secondly, the term ‘theme’ refers to a group of related concepts which is ascertained by Leximancer through its coding blocks of sentences. A ‘theme’ will contain concepts that co-occur often within the same coding block (Cretchley et al., 2010, p 319), with a theme represented by a circle on the concept map.

Through its analysis, Leximancer can then present ‘conceptual diagrams’ that present ‘concepts’, ‘themes’ and how they relate to each through the use of dots, circles and

lines. The conceptual diagram allows for the interpretation of data in a pictorial form which would otherwise be too cluttered in a pure numerical form. The concept maps have ‘themes’ which are represented by a circle around several concepts. Each theme is named after the ‘most prominent concept in that group, which is also indicated by the largest dot in the theme cluster; the researcher can rename themes when appropriate’ (Cretchley et al., 2010, p 319). The closeness between concepts indicates their co-occurrence in the data, and those concepts that are more central in the circle indicates its extent of connectivity to other concepts (Verreynne et al., 2011).

The size of the dot for each concept represents its prominence. The distance of the concepts within each circle also represents the proximity of concepts. The concepts are clustered according to weight and relationship. This means that concepts that appear together frequently in the data set or in similar situations are close together on the map. The grey lines show the most likely connections between concepts.

The advantage of Leximancer is that the Leximancer software can either run via default or can be directed to target key topics, as well as compare groups of different texts. As identified by Cretchley et al., 2010:

“Using the concept map, rank-ordered concept lists and text query options, the researcher can read instances of specific concepts to understand their relationships. Thus, the map visually represents the strength of association between concepts and provides a conceptual overview of the semantic structure of the data.” (Cretchley et al., 2010, p 319)

Who has used it?

Leximancer has been applied for various research purposes: including literature review, coding open-ended surveys, coding interview transcripts, analysing media articles, analysing website content (e.g. university site profile), and other text documents such as submissions from stakeholders. Table 2 summarises the common applications of Leximancer in accounting and management studies.

Leximancer has been used for data analysis in various academic social and business related research, including psychology (Cretchley, Rooney, & Gallois, 2010), accounting (Axelsen, Green, & Ridley, 2017; Crofts & Bisman, 2010a, 2010b; Palmer, 2013), legal duties (Voogt & Verreynne, 2018); management disciplines (Ho, Nguyen, Lo, McLean, & Teo, 2015; McKenna, Verreynne, & Waddell, 2016) (Cukier et al.,

2017; Maluwetig & Verreyne, 2011; Verreyne, Parker, & Wilson, 2011); and taxation (McCredie and Sadiq, 2018).

Table 2: Applications of Leximancer in Academic Research Settings

Application	Research Purpose	Type of Text	Output Options	Authors
Literature Review	To investigate the contextual concepts of accountability in accounting journals.	Analysed 114 journal articles related to accountability published in highly ranked accounting journals from 2000 to 2017.	Visual via concept maps and report via statistical output.	(Crofts & Bisman, 2010b) – accounting discipline
	To track the development of themes and concepts within the Journal of Cross-Cultural Psychology over 40 years.	Analysed 1416 articles mapped by decade.	Concept maps and representative quotes from text extracts.	(Cretchley et al., 2010) – psychology discipline
	To chart the movement of HR competency research and the signals given by organisations in the recruitment of HR professionals in New Zealand	Analysed 51 HR competency research articles from the Business Source Premier database from 1990-2012.	Concept maps, Insight Dashboard and statistical outputs.	(Ho et al., 2015) – management discipline
Coding Open-ended Surveys	To explore the perspectives of employees and managers from 50 small Australian firms to build a more complete and nuanced view of the interaction of human resources and capabilities with firm performance.	Analysed the qualitative open text data from six questions in the CEO survey and 30 questions in the employee survey.	Concept maps and representative quotes from text.	(Verreyne et al., 2011) – management discipline
Coding Interview Transcripts	To explore how work and life are discursively constructed by men and women in their daily practice of living	Analysed data from focus group and semi-structured interviews of 18 female and 19 male Australian managers in the financial and government sector	Concept maps and statistical output.	(McKenna et al., 2016) – management discipline
	To identify relevant “common statements” representing the prominent and shared perceptions of the IS auditor role amongst these auditors.	Analysed 32 interview transcripts with IS auditors on their roles in the public sector financial audit	Statistical output and representative quotes from text.	(Axelsen et al., 2017) – accounting discipline
Coding Media Articles	To explore the discursive constructions of equity and diversity in Canada at societal/macro level and provides original findings on how these discourses have been produced in the public sphere	125 articles published from 1986 to 2014 in The Globe and Mail (Canadian newspaper) and 26 report from the Canadian Employment Equity Act Annual Reports from 1988 to 2013.	Concept maps, statistical output and representative quotes from text.	(Cukier et al., 2017) – management discipline
Coding Website Content	To investigate the prevalence of social dimensions in business education in high-performing universities and to examine the relationship between their ranking and	Text data extracted from the university’s website of the top-ranking universities to reveal preferences, priorities, and	Concept maps	(Maluwetig & Verreyne, 2011) – management

Application	Research Purpose	Type of Text	Output Options	Authors
	social dimensions	recurring themes among high performers.		
Stakeholder Relationship Management	To explore attitudes to financial reporting in the not-for-profit (NFP) sector, by utilising submissions to the 2008 Australian Senate Economics Standing Committee.	Examined submissions from NFP entities and interested parties to the Senate	Concept Map and representative quotes from text.	(Palmer, 2013) – accounting
	To illustrate the complementary use of two computational semantic analysis techniques – latent semantic analysis and data mining - to examine a large textual data set.	Analysed 8544 abstracts from selected journals in Information Systems, Management, and Accounting over a period of 25 years.		(Indulska, Hovorka, & Recker, 2012) - IS, Management and Accounting
Coding Roles and Responsibilities	To determine the role and responsibilities of company directors form the codes and/or legislation from five jurisdictions: UK, Singapore, Australia, G20/OECD and South Africa.	Codes and legislative instruments governing company directors to determine key concepts and themes in their role and responsibility.	Concept Map and representative quotes from text.	(Voogt & Verreyne, 2018) – Business law.

Of course with any tool there are advantages and disadvantages. A clear advantage of Leximancer is its ability to effectively analyse a large quantity of text data (and for the purposes of this study this involved over 700 articles. See the later discussion). However, for data (articles) to be analysed they need to be ‘editable’ text format; this means that either they are in a Word document or editable PDF document. Note it is possible to convert a read only PDF into Word, though this may lead to some formatting errors. For example, the documents created may contain sentences truncated by paragraph marks, words containing unreadable symbols or incorrect spacing between words, which may result in inaccurate coding of concepts. Although it would be ideal to correct all these text processing errors before inputting the files into Leximancer, this process requires manual editing which is extremely time-consuming and prone to human error.¹⁰

While Leximancer allows for the rapid identification of concepts (Smith & Humphreys, 2006) (Sotiriadou et al., 2014); some of the data (words) can be irrelevant such as headers and footers, headings, references, etc that can appear in the formatting of an article. Thus requires manual removal of irrelevant text content (e.g., headings and references) prior to data analysis, while time consuming it provides for richer data for content analysis.

A clear advantage of Leximancer is objective data analysis as it avoids problems of researcher bias coder subjectivity (Smith & Humphreys, 2006) (Sotiriadou et al., 2014). However, this analysis only provides a bird-eye conceptual overview of the text data only. Articles that focus on specific concept(s) will have higher frequency counts for those concepts due to the narrow focus of the article (particularly in the results and discussion sections of the same article where the key concept(s) are mentioned

¹⁰ In some converted files, sentences that span over several lines may be truncated by paragraph marks, resulting in inaccurate coding of concepts and their co-occurrence with other concepts within the same sentence. Example one from Burton (2006, p.171):

“What much of this literature misses, however, is that the same scepticism which undermines absolute norms of substantive fairness also undermines any claim that a particular political process, such as deliberative democracy, is the ‘right’ process (Bohman, 1998; Mouffe, 1995).”

This example shows that the Leximancer has treated this quote as two sentences as there is a paragraph mark before the word ‘scepticism’, which led to a commencement of a new sentence.

Example two from Deak (2004, p.3):

“In reviewing tax avoidance schemes, it is essential to identify the inherent logic and the authentic content of a legal system. The validity of law is a category equivalent to the micro category of the coherence of law.”

This example shows that although there is a missing space between the first and second sentences Leximancer has still recognised them as two sentences because of the full-stop after the word ‘system’. However, the phrase ‘oflaw’ is treated as one single word due to a missing space in between.

repeatedly). Conversely, a literature review article will have more widespread frequency counts for a wider array of concepts.

Part of Leximancer's reliability is that little researcher intervention is required in content analysis process (Sotiriadou et al., 2014), and text segments are always coded in the same way given the same parameter settings (Smith & Humphreys, 2006). However, there can be some issues with jargon and similar terms for Leximancer. For the purposes of our study, we disabled the self-learning thesaurus and prescribed multiple meanings for certain concepts. For instance, the concept 'tax agents' is also referred to as 'tax advisors', 'tax consultants', 'tax lawyers' and 'tax practitioners' in the literature. However, this can lead to concepts in a compound can be implicitly stated anywhere in the text segment – for instance, the compound concept 'tax rate' was also coded where the term 'rate' was used to denote 'evasion rate', 'audit rate', 'penalty rate', and 'compliance rate'. We did not isolate coding of other terms for this compound concept, as they were also important to the study.

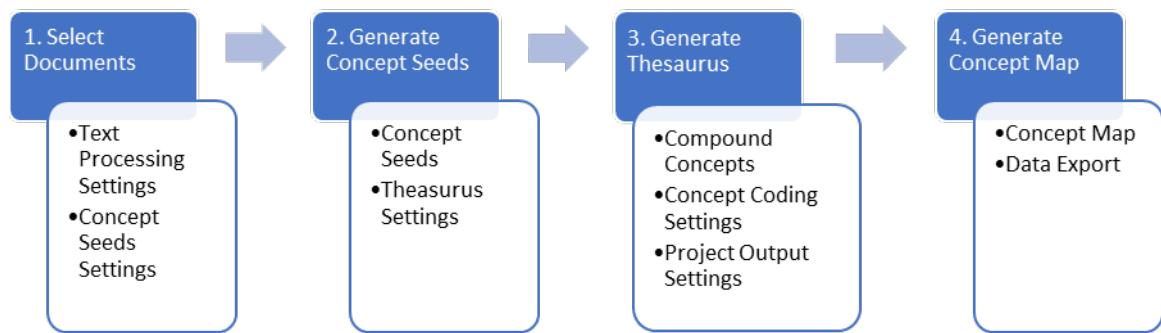
Also, the concept diagrams themselves are stochastic, meaning that the concepts on the map may settle in different positions each time a new map is generated from the software. This is because the concepts were initially distributed randomly in multidimensional space (Leximancer Manual version 4.5, 2016, p.89). However, any instability in the map does not affect the frequency statistics or the centrality rankings of concepts (Smith & Humphreys, 2006).

Overall, it was determined that it was appropriate to use Leximancer as a semantic mapping methodology to organise the large amount of literature on tax compliance, and to explore the conceptual complexity of tax compliance factors (variables) and the interrelationship between the factors. How this process occurred is described next.

Data processing and data analysis

Leximancer (version 4.5) was used for this study, and Figure 1 illustrates the four key processing stages in the Main Leximancer User Interface of: 1. Select Documents, 2. Generate Concept Seeds, 3. Generate Thesaurus and 4. Generate Concept Map.

Figure 1: Leximancer Project Processing Stages



Source: adapted from the Main Leximancer User Interface (<http://localhost:8085/leximancer/>)

In brief, the Select Documents stage enables the user to select the source of documents for analysis and specify configuring parameters for text processing and concept seeds.

In terms of the Select Documents studies on tax compliance factors were researched and developed from 1 January 1998 to 31 December 2017. Due to the multidisciplinary nature of tax compliance, articles were gathered from taxation, accounting, social sciences, law, economics, business, political sciences and psychology journals. Fifteen databases were used to access these journals, and they were:

- ABI-Inform Complete (Proquest)
- Accounting and Tax (Proquest)
- Business Source Complete
- ScienceDirect
- AGIS Plus Text
- Australian Taxation Law Library
- Google Scholar
- Heinonline
- JSTOR
- Kluwer Law
- Lexis Advance
- LexiNexisNZ
- Thomson Reuters WestLaw

- WestLaw NZ and Australia
- BestCase (via WestlawNext Canada).

The breadth of the current study appears to be more comprehensive than the prior two studies given the number of databases used to draw research from (refer to Table 1).

Within these databases, keywords used for the search were “tax* compliance” (truncation is used for ‘tax’ to include taxation and taxpayer) on peer reviewed, full-text, academic journal articles from 1998 to 2017 inclusively. The current study extended beyond ‘income tax’ compliance cost studies that was just the focus of Jackson & Milliron (1986) and Richardson & Sawyer (2001), and extended to VAT and GST compliance cost studies, although studies about fuel tax, inheritance tax and land tax were excluded. Also, limited search query (e.g. abstract only) and additional keywords (e.g. AND factor OR determinant OR variable; AND gender OR culture OR religion OR religiosity) were used for Klewer, Thomason Reuters WestLaw and WestLaw NZ databases to further refine the search results.

From this initial process a total of 917 articles were identified through the 15 databases. 132 articles were excluded as they were not accessible and were not relevant.¹¹ A further 128 articles were further excluded as they related to tax policies and tax reform articles which were not relevant for the purpose of this research.¹² This resulted in a total of 717 papers being eligible for further text analysis.

The 717 eligible papers were then divided into two main groups: conceptual and empirical. Conceptual papers are discussion papers without empirical data; whereas empirical papers were supported by sample data. Conceptual papers were tagged under the Literature Review folder and the empirical papers were tagged under the Quantitative, Qualitative and Mixed Method folders. The names of the folders reflect the different methodological approach. Of the 717 articles, 175 were conceptual papers, and 542 were empirical papers. From the 717 articles, 262 papers were from

¹¹ These articles came up in the initial search results but in the second stage of screening and sorting they did not fit into the research type criteria or they only mentioned tax compliance several times (less than 10 times and mainly within the reference/note section) in the paper, meaning that the focus of the paper is too broad and not suitable for the purpose of our research. A small number of papers were accessible but were only available in scanned and uneditable PDF format and could not be used for Leximancer analysis.

¹² Tax policies and tax reform articles were excluded because they did not focus on tax compliance variables but were more focused on legislative history, requirements and suggestions for reforms.

1998 to 2007 and 455 from 2008 to 2017. The significant increase in published tax compliance articles in the second decade demonstrates a greater level of activity in the topic, which could be due to the implications of tax compliance on tax revenues, policies and practice. Table 3 shows the breakdown of tax compliance articles for this study between the conceptual and empirical groups and the research methods used.

Table 3: Type of Research

Type of Research	1998-2017 (n=717)
Conceptual papers (tagged under Literature Review folder)	175
Empirical papers (tagged under Quantitative, Qualitative and Mixed Methods folders)	489
Quantitative Folder comprised of:	
- Survey	202
- Experiment	116
- Analytical Modelling	96
- Regression Analysis	75
Total Quantitative	489
Qualitative folder comprised of:	
- Interviews	19
- Case Study	8
- Focus Groups	1
- Observation	2
- Content Analysis	3
- Diary or Log Book Study	0
Total Qualitative	33
Mixed Methods	20

If nothing else there is a clear growth in research being conducted in this area, as 717 articles were obtained from 15 research databases compared to the modest 43 studies considered by Jackson & Milliron (1986), and approximately 170 studies in Richardson and Sawyer (2001). For this study this represents an average of 35 papers each year, compared to 14 papers each year for the ten years considered by Richardson and Sawyer.

In the text processing and concept seed setting we have adopted all default settings with the exception of setting the context blocks for coding as single sentences instead of using the default context blocks of two sentences. Single sentence context blocking was appropriate to ensure the context block was a cohesive statement and enables compound concepts (see stage 3 Generate Thesaurus) to be coded within a single sentence block. We tagged the document folders in the text processing setting to

generate concept maps that identifies and compares concept frequency and relationships between each research type (i.e., literature review, quantitative research, qualitative research and mixed methods research).

Although Leximancer allows the user to generate an automatic map by completing the Select Document stage followed immediately by the Generate Concept stage, we have also completed the Generate Concept Seeds and Generate Thesaurus stages with specified settings. In the Generate Concept Seeds stage, we removed all automated concept seeds and created a list of user-defined concept seeds instead. This is because our research focuses on the tax compliance literature which utilises a number of specific terminologies (which followed the terms used in Jackson and Milliron (1986) and Richardson and Sawyer (2001) see Table 1). In addition, we disabled the self-learning thesaurus function and manually merged concepts with similar meanings (e.g., religion and religiosity).

For the Generate Thesaurus Stage we created compound concepts which were concepts comprising of more than one term (e.g., ‘tax’ AND ‘evasion’, ‘trust’ AND NOT ‘financial’ AND NOT ‘foreign’) appearing within a context block. We then selected the concepts to be included in the output maps. The settings chosen for each of the above mentioned processing stages is listed in Table 4 .

Table 4: Leximancer processing and analysis settings

Leximancer Settings
Stage One: Select Document
<ol style="list-style-type: none"> 1. Open the Select Document Settings, select folders containing source documents 2. Open Text Processing Options, go to the General column - sentences per block, click on the drop down menu and select 1 to reduce corpus to 1-sentence 3. Go to the Tab column, tick the box next to Folder, click OK, close the Text Processing Options screen
Stage Two: Generate Concept Seeds
<ol style="list-style-type: none"> 1. Run the Generate Concept Seeds Process 2. Open Concept Seed Editor, click on the Auto Concepts Tab, select all concepts and click Remove, click OK 3. Click on the Auto Tags Tab, merge and/or edit folder names, click OR 4. Click on the User Defined Concepts Tab, add concepts and merge concepts with similar meanings, click OK, close the Concept Seed Editor screen 5. Open the Thesaurus Setting (Concept Learning), go to the General column – Learn Concept Thesaurus using Source Documents, untick the box to disable self-learning thesaurus, click OK, close the Thesaurus Setting screen
Stage Three: General Thesaurus
<ol style="list-style-type: none"> 1. Run the General Thesaurus 2. Open Compound Concepts, select name-like and word-like concepts lists in the first two columns, add to the right-hand column and create compound concepts using AND, OR and/or NOT functions, edit the names of compound concepts, click OK 3. Open Concept Coding Settings, go to the Mapping Concepts Tab, select all concepts in the Selected Mapping Concepts column and click on the right-hand arrow button to remove them from the list, then select folder tags from the Available Names column and mapping concepts from the Available Concepts column and click on the left-hand arrow button to add them to the Selected Mapping Column, click OK
Stage Four: Generate Concept Map
<ol style="list-style-type: none"> 1. Run Generate Concept Map 2. Open Concept Map 3. Adjust concept visibility, theme size and rotation 4. Edit font size, theme and concept labels 5. Save concept map 6. Export concept map, theme summary, ranked concept lists and query results

The initial analysis yielded 107 terms, with tax occurring most frequently and rights the least. We enhanced the semantic usefulness of Leximancer findings by further refining the analysis process with the following steps:

1. We deleted concepts of little relevance to the paper (e.g. data, study) and name-like concepts (e.g. FACTA and IRD);

2. We merged concepts which are synonymous with others (e.g. cost and costs, businesses and companies);
3. We added concepts that were of interest to our study but not identified as frequent concepts by Leximancer (e.g. culture, religion);¹³ and
4. We created compound concepts, which are comprised of two or more concepts that appear in the same sentence text (e. g. sole traders, tax morale).

Table 5 provides the thesaurus for the 38 tax compliance concepts. The thesaurus clarifies the tax compliance concepts currently reported in this study. It does so by reducing duplication of similar words such as ethics or ethical. The thesaurus also provides the derivation of compound words used such as income level and perceived opportunity. This is to provide transparency in the data processing for the 717 papers.

Table 5: Thesaurus for the 38 Tax Compliance Concepts

Tax Compliance Concepts	Thesaurus
age	taxpayer AND age
authorities contact	authorities AND contact
business taxpayer	business AND taxpayer
culture	cultural OR culture OR ethnic OR ethnicity OR race
education	taxpayer AND education
ethics	ethics OR ethical
framing	frames OR framing
gender	taxpayer AND gender
income level	income AND level
income source	income AND source
individual taxpayer	individual AND taxpayers
occupation	taxpayer AND (occupation OR status)
peer influence	peer AND influence
perceived opportunity	perceived AND opportunity
power	power
political	political
probability of detection	(probability AND detection) OR (tax AND audit)
religion	religion OR religiosity
reciprocity	reciprocity
sanctions	sanction OR sanctions
social norms	social AND norms
sole traders	self-employed OR SME(s) OR (sole AND (enterprises OR traders OR proprietors)) OR ((micro OR small) AND businesses)
tax agents	(tax OR taxation) AND (agent(s) OR accountant(s) OR advisor(s) OR consultant(s) OR lawyer(s) OR practitioner(s) OR preparer(s))

¹³ This means there appearance is because Leximancer was instructed to look for these concepts.

tax amnesty	(tax OR taxation) AND (amnesty OR amnesties)
tax avoidance	(tax OR taxation) AND avoidance
tax complexity	(tax OR taxation) AND complexity
tax compliance cost	(tax OR taxation) AND compliance AND cost
tax evasion	(tax OR taxation) AND evasion
tax enforcement	(tax OR taxation) AND enforcement
tax fairness	(tax OR taxation) AND fairness
tax inducement	((tax OR taxation OR positive) AND inducements) OR ((tax OR taxation) AND (penalty OR penalties))
tax morale	(tax OR taxation) AND morale
tax perception	(tax OR taxation) AND perception(s)
tax rate	(tax OR taxation) AND rate(s)
tax system	(tax OR taxation) AND system
trust	trust AND (NOT financial) AND (NOT foreign)
wealthy taxpayer	(wealthy OR (high AND wealth)) AND taxpayer(s)
Withholding tax	(withholding tax*)

To facilitate an analysis of different time series in terms of how tax compliance research may have altered between two decades (1st decade: 1998 to 2007; 2nd decade: 2008 to 2017) we created subsets of data in 10-year periods.

The next section provides a summary of the results of the Leximancer analysis.

Section Four: Results

This paper contributes to research on tax compliance in three ways. First, it collects and summarises multiple tax compliance literature over the last two decades from 1998 to 2017 inclusive. Second, it provides a snapshot overview on the different concepts and themes that were pertinent and researched since the Richardson and Sawyer's (2001) paper. Third, it compares the different themes and concepts that were researched between those in the prior studies of Jackson and Milliron; and Richardson and Sawyer

There are three segments to in this section which are:

1. Research concepts and themes identified from 1998 to 2017
2. The top five themes researched over the last 20 years; and
3. Concepts researched between 1998 to 2017 compared to Jackson & Milliron and Richardson & Sawyer

Research concepts and themes identified from 1998 to 2017

Through the analysis of the 717 articles, 38 variables (concepts) were identified, which means that an additional 19 variables have emerged since the prior studies of Jackson and Milliron (1986) and Richardson and Sawyer (2001). Table 6 provides an overview of the results of this study, including the variables/concepts identified.

Table 6: Variables Identified 1998 to 2017

Study	Period covered	Number and type of papers	Jurisdictions and study focus	Databases/Source	Variables identified
Yong et al.	Jan 1998 to Dec 2017 (20 Years)	717 Only published refereed articles (262 for 1998 to 2007) (455 for 2008 to 2017)	Multiple-jurisdictional Compliance studies which looked at income tax, CGT, VAT (GST). (Excluded compliance cost studies about fuel tax, inheritance tax and land tax)	15 databases utilised: 1. ABI-Inform Complete (Proquest) 2. Accounting and Tax (Proquest) 3. Business Source Complete 4. ScienceDirect 5. AGIS Plus Text 6. Australian Taxation Law Library 7. Google Scholar 8. Heinonline 9. JSTOR 10. Kluwer Law 11. Lexis Advance 12. LexiNexisNZ 13. Thomson Reuters WestLaw 14. WestLaw NZ and Australia 15. BestCase (via WestlawNext Canada).	38 variables described: J & 's 14 Variables: 1. Age 2. Gender 3. Education 4. Income level 5. Income source 6. Occupation, 7. Peer Influence, 8. Ethics 9. Fairness 10. Complexity 11. Revenue Authority Contact 12. Sanctions, 13. Probability of Detection 14. Tax Rates R & S's 5 additional compliance variables: 15. Compliance costs, 16. Tax preparers 17. Framing 18. Positive inducements 19. Tax amnesties Additional 19 variables (concepts) identified: 20. Tax evasion (more a theme than a factor) 21. Tax enforcement 22. Trust 23. Power 24. Sole traders 25. Culture 26. Tax System 27. Political 28. Individual taxpayer 29. Business Taxpayer 30. Tax morale 31. Social Norms 32. Religion 33. Withholding taxes 34. Reciprocity 35. Tax perception 36. Perceived opportunity 37. Tax avoidance 38. Wealthy taxpayer

These 38 concepts are associated with the five themes of Tax Evasion, Trust, Tax System, Tax Morale and Tax Enforcement. The way that the concepts are clustered around themes is detailed in Table 7 (with new additional concepts denoted by ‘*’).

Table 7: Tax Compliance Concepts contained in each Theme

Themes	Tax Compliance Concepts contained in each Cluster	Number of Concepts
Tax evasion	1. Tax evasion* 2. Tax rate 3. Probability of detection 4. Tax inducement 5. Tax fairness 6. Tax perception* 7. Tax enforcement* 8. Income level 9. Occupation 10. Peer influence	10
Trust	11. Trust* 12. Tax agents 13. Power* 14. Sole traders* 15. Culture* 16. Age 17. Gender	7
Tax system	18. Tax system* 19. Tax compliance cost 20. Political* 21. Tax complexity 22. Individual taxpayer* 23. Business taxpayer* 24. Education	7
Tax morale	25. Tax morale* 26. Ethics 27. Social norms* 28. Tax amnesty 29. Religion* 30. Framing 31. Perceived opportunity*	7
Tax avoidance	32. Tax avoidance* 33. Sanction 34. Income source 35. Withholding taxes* 36. Reciprocity* 37. Authorities contact 38. Wealthy taxpayer*	7

Note: * denotes tax compliance variables identified in our study in addition to those examined by Jackson and Milliron (1986)¹⁴ and Richardson and Sawyer (2001).¹⁵ See Table 1 and Table 6 for more detail.

¹⁴ Jackson and Milliron (1986) examined fourteen compliance variables: age, gender, education, income level, income source, occupation, peer influence, ethics, fairness, complexity, revenue authority contact, sanctions, the probability of detection and tax rates.

Hence additional new tax compliance concepts (variables) researched since Richardson and Sawyer (2001) paper are: Tax evasion (more a theme than a factor); Tax enforcement; Trust; Power; Sole traders; Culture; Tax System; Political; Individual taxpayer; Business Taxpayer; Tax morale; Social Norms; Religion; Withholding taxes; Reciprocity; Tax perception; Perceived opportunity; Tax avoidance and Wealthy taxpayer. These issues were given more attention over the last twenty years. It could be that these terms are inter-related to the prior variables identified but now are more nuanced, or they could be new variables explored. For example the new concept of ‘tax enforcement’ is likely to be related to the prior variable of ‘sanctions’.

Additional information regarding the concepts is provided in Table 8, which ranks the concepts by count and relevance scores. This shows the number of times each concept has been mentioned within the texts of the 717 articles. As shown in Table 8, *tax evasion* has the highest count followed by *tax rate*, *tax system*, and *probability of detection*. On the other hand, the lowest count were from *wealthy taxpayer* and *perceived opportunity* concepts with only 43 and 30 mentions within the texts. While this does not indicate importance or which factors (variables) has the largest influence on tax compliance it does indicate the number of times concepts are mentioned. What the prominence of ‘tax evasion’ could be indicating that to understand ‘compliance’ we need to better understand the reason why people evade their tax. The relevance count would tend to suggest that the following concepts are more relevant to tax compliance in the new millennium as they are over 30%: tax evasion; tax rate; tax system; probability of detection; tax morale; trust; tax agents and tax compliance cost. Whereas the following concepts appear weaker with 5% or less reliance: income source; framing; reciprocity; withholding taxes; education; occupation; Age; gender; authorities contact; peer influence; wealthy taxpayer and perceived opportunity. However, it could be that these factors are less researched than the others (which may be due to methodological issues, or difficulties in measuring).

¹⁵ Richardson and Sawyer (2001) added five other variables: compliance costs, tax agents, tax amnesties, framing of tax decisions and positive inducements.

Table 8: Ranked List of Tax Compliance Concepts by Count and Relevance Scores

Concepts	Count	Percentage of Count Relative to the Top Concept
tax evasion	8422	100%
tax rate	5523	66%
tax system	4385	52%
probability of detection	4361	52%
tax morale	3560	42%
trust	3497	42%
tax agents	3011	36%
tax compliance cost	2810	33%
sole traders	2439	29%
tax inducement	2347	28%
power	2118	25%
ethics	2074	25%
culture	1913	23%
tax avoidance	1837	22%
tax enforcement	1625	19%
political	1565	19%
tax fairness	1543	18%
sanction	1466	17%
tax perception	1414	17%
social norms	1272	15%
tax complexity	1054	13%
individual taxpayer	1037	12%
income level	1019	12%
tax amnesty	769	09%
business taxpayer	764	09%
religion	518	06%
income source	444	05%
framing	432	05%
reciprocity	328	04%
withholding taxes	301	04%
education	217	03%
occupation	138	02%
Age	138	02%
gender	72	01%
authorities contact	65	01%
peer influence	56	01%
wealthy taxpayer	43	01%
perceived opportunity	30	00%

Note:

Count = the number of times the concept appears in the entire data set (per one-sentence block)

Percentage of Count Relative to the Top Concept (or relevance percentage) = the count value of each concept divided by the single highest count value. The relevance percentage for tax rate is calculated by taking the count value for tax rate (5523) then divide it by tax evasion's count value (8422) then times 100% = 66%. Thus, the most frequent concept will always have 100% relevance, regardless of whether it occurs in all context blocks.

Top Five Concepts

Table 9 ranks the list of tax compliance themes by count and connectivity scores. This information is important in determining the strength of the relationships (if any) between the concepts. It shows the connection of each concept to one another rather than the count relative to the most important concept which is tax evasion.

Table 9: Ranked List of Tax Compliance Themes by Count and Connectivity Scores

Theme	Concept	Count	Total Count	Connectivity	Total Connectivity
1. Tax evasion	1. Tax evasion*	8422	26448	4100	17688
	2. Tax rate	5523		3296	
	3. Probability of detection	4361		2833	
	4. Tax inducement	2347		2028	
	5. Tax fairness	1543		1745	
	6. Tax perception	1414		1677	
	7. Tax enforcement*	1625		1005	
	8. Income level	1019		712	
	9. Occupation	138		165	
	10. Peer influence	56		127	
2. Trust	11. Trust*	3497	13188	1656	5763
	12. Tax agents	3011		1252	
	13. Power*	2118		1107	
	14. Sole traders*	2439		706	
	15. Culture*	1913		718	
	16. Age	138		164	
	17. Gender	72		160	
3. Tax system	18. Tax system*	4385	11832	2507	6047
	19. Tax compliance cost	2810		1031	
	20. Political*	1565		537	
	21. Tax complexity	1054		938	
	22. Individual taxpayer*	1037		425	
	23. Business taxpayer*	764		375	
	24. Education	217		234	
4. Tax morale	25. Tax morale*	3560	8655	1340	3495
	26. Ethics	2074		913	
	27. Social norms*	1272		598	
	28. Tax amnesty	769		252	
	29. Religion*	518		274	
	30. Framing	432		84	

	31. Perceived Opportunity*	30		34	
5. Tax avoidance	32. Tax avoidance	1837	4484	941	2345
	33. Sanction	1466		624	
	34. Income source	444		360	
	35. Withholding taxes*	301		157	
	36. Reciprocity*	328		108	
	37. Authorities contact	65		134	
	38. Wealthy taxpayer	43		21	
Total			64607		35338

Each concept has a count value and connectivity score.

Count value measures the number of times a concept appears within the data set. For instance, Table 9 shows that the concept *tax evasion* has a count score of 8422, meaning that this concept has been coded 8422 times in the text. Therefore, the total concept counts for the theme *Tax evasion* = tax evasion 8422 + tax rate 523 + probability of detection 4361 + tax inducement 2347 + tax fairness 1543 + tax perception 1414 + tax enforcement 1625 + income level 1019 + occupation 138 + peer influence 56 = 26448.

Connectivity score measures the number of times a concept appears concurrently with another concept within the same sentence in the data set. Note the '*' highlights the new variables/concepts compared to the prior study of Richardson and Sawyer (2001).

Figure 2 illustrates that when the concept *tax evasion* is selected in the concept map, a number of red radiant lines appears connecting this concept with various other concepts, and the connectivity scores for this concept is displayed on the count column on the right. According to the connectivity scores, *tax evasion* is coded 13 times with *perceived opportunity*, 516 times with *tax avoidance*, 258 times with *tax perception*, etc, which add up to a total connectivity score of 4100 for the concept *tax evasion*. Hence, the sum of the connectivity scores for the *Tax Evasion* theme = tax evasion 4100 + tax rate 3296 + probability of detection 2833 + tax inducement 2028 + tax fairness 1745 + tax perception 1677 + income level 1005 + income level 712 + occupation 165 + peer influence 127 = 17688. Count and connectivity provides different information. For example, a concept can appear very often (high count) but only in a very particular context, so its hit count would be high but its connectivity would be low. The connectivity is important as it shows what variables are being researched simultaneously.

Figure 2: Leximancer Interface showing the Connectivity for the Concept Tax Evasion

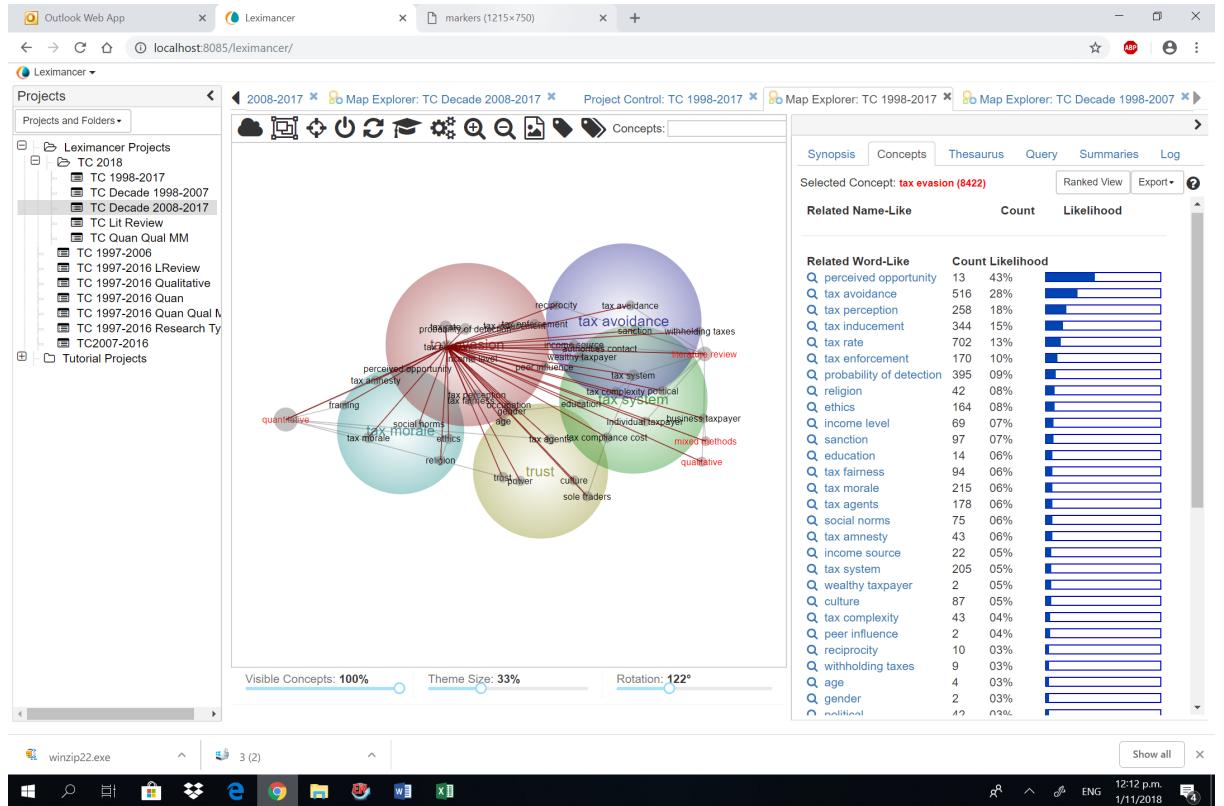


Table 8 demonstrates that *tax evasion*, *tax rate*, *tax system*, *probability of detection* and *tax morale* are the top five concepts across the entire text. Further exploration of the indirect relationships between the concepts and their connectivity is detailed in Table 10.

Table 10: Top 5 Concepts and Their Connected Concepts

Top 5 Concepts	Total Count	Connected Concepts	Co-occurrence with Connected Concepts	Total Count of Connected Concepts	Connectivity Percentage
Tax evasion	8422	Perceived opportunity	13	30	43%
		Tax avoidance	516	1837	28%
		Tax perception	258	1414	18%
		Tax inducement	344	2347	15%
Tax system	4385	Tax fairness	361	1543	23%
		Tax complexity	241	1054	23%
		Tax perception	235	1414	17%
Probability of detection	4361	Peer influence	19	56	34%
		Authorities contact	18	65	28%
		Tax inducement	545	2347	23%
Tax rate	4256	Authorites contact	15	65	23%
		Peer influence	10	56	18%
		Tax inducement	289	2347	12%
Tax Morale	3560	Religion	91	518	18%
		Culture	170	1913	09%
		Trust	170	3497	05%

Tax evasion

Tax evasion (count = 8422) is in strong association with other identified concepts such as *perceived opportunity* (count = 30). These two concepts were coded 13 times together within the same sentence block across the data set, with a connectivity score of 43%, which is calculated based on their co-occurrence score (i.e. 13) divided by the total number of times the connected concept (i.e. *perceived opportunity*) is coded across the entire data set (i.e. 30). However, it is important to note that a query search for the concepts ‘tax evasion’ and ‘perceived opportunity’ shows that of the 13 co-occurrences for these two concepts, 11 co-occurrences were from the same source - a quantitative survey research article by Keraro et al. (2017). This could suggest that while Keraro et al. (2017) considered these two concepts closely, other work largely did not. Keraro et al.’s study considered whether what effect, if any, perceived opportunity had for tax evasion on tax compliance for small and medium enterprises (SME) in Kenya. What the study concluded was that as the perceived opportunity for tax evasion increased, tax compliance reduces among SMEs.

Tax evasion is also in strong connection with *tax avoidance* (count = 1837), with a co-occurrence score of 516 and connectivity score of 28%, and with *tax perception* (count = 1414) with a co-occurrence score of 258 and connectivity score of 18%. Here are some quotes that illustrate how the concept tax evasion is used with these two terms:

“The main difference between tax evasion and tax avoidance is not only the interpretation of the laws but the exploitation of legal ambiguities by powerful taxpayers.” (Bergman, 1998, p.57) – Literature Review

“The dividing line between tax avoidance (legitimate tax planning) and tax evasion (abuse) it is not a natural or unanimous one, but rather depends on policy decisions of “each state.” (Barreto, 2015, p.829) – Literature Review

“If the perception becomes widespread that the government is not willing to detect and penalize evaders, then such a perception legitimizes tax evasion.” (Alm, 1998, p.265) – Literature Review

“In contrast, equity theory predicts that a given level of tax payments, as the increase in public goods improves the taxpayer’s perception of exchange equity, the taxpayer will be less likely to participate in tax evasion.” (Kim, 2002, p.772) - Quantitative research

An exploration of the concept *tax evasion* using a query showed the quotes extracted from the text on what contributes to tax evasion:

“Based on *Jackson and Hatfield* (2005) and *Highthouse and Yuce* (1996), the risky alternative of tax evasion (or under-reporting) will be perceived as an “opportunity” by taxpayers in the loss domain and as a threat to taxpayers in the gain domain.” (Sanders et al. 2008, p.8) – Quantitative research

“Tax officers believed that taxpayers’ intention, role of tax agents, taxpayers’ ignorance of tax law and financial strain were the main reasons for non-compliance.” (Abdul-Jabbar & Pope 2009, p.5)- Quantitative Research

“This research seeks to incorporate into the compliance analysis psychological factors such as shame, embarrassment, and guilt, as well as taxpayer perceptions of fairness, satisfaction with the Government, and moral commitments to honest reporting” (Mazz, 2003, p. 1068) – Mixed Methods research.

The following quote demonstrates the tension between the two terms:

“The dividing line between tax avoidance (legitimate tax planning) and tax evasion (abuse) it is not a natural or unanimous one, but rather depends on policy decisions of each state.” (Barreto, 2015, p. xx) – Literature review

“Tax non-compliance can be divided into tax avoidance and tax evasion.” (Suffian, 2017, p.159) – Quantitative Research

“Tax compliance may be seen in terms of tax avoidance and tax evasion.” (James, et al, 2005 p. 163) – Quantitative research (survey)

“Tax evasion and tax avoidance are deviations from what is to be found legal, fair and just.” (Deak, 2004, p. 41) – and “Tax evasion and tax avoidance concern more than simply the problem of losing public revenues.” (Deak, 2004, p.64) – Literature Review

Tax system

The second most important concept *tax system* (count = 4385) is also closely connected to *tax fairness* (count = 1543, co-occurrence = 361, connectivity = 23%), *tax complexity* (count = 1054, co-occurrence = 241, connectivity = 23%) and *tax perception* (count = 1414, co-occurrence = 235, connectivity = 17%).

“Tax scholars, at least those that follow in the footsteps of Kenneth Carter, often manifest a preoccupation with tax equity and fairness as a preliminary step on the way to advocating the importance of a robust, comprehensive income tax system.” (Brooks, 2015, p.383) – Literature Review

“The taxpayer is thus left in the predicament where in order to be compliant in a system of increasing complexity, he or she must pay tax specialists to interpret and manage tax affairs.” (Langham et al., 2012, p.393) – Quantitative research

“Participants clearly mentioned that their negative perceptions of the income tax system, particularly on the tax rate structures and government spending, had motivated them to avoid and evade paying tax.” (Saad, 2013, p.104) – Qualitative Review

Probability of detection (also referred to as tax audits)

The third most important concept *probability of detection* (count = 4361) is closely connected to *peer influence* (count = 56, co-occurrence = 19, connectivity = 34%), *authorities contact* (count = 65, co-occurrence = 18, connectivity = 28%) and *tax inducement* (also defined as tax penalties) (count = 2347, co-occurrence = 545, connectivity = 23%).

“However, beliefs about the probability of apprehension may also be affected by personal experience, peer influence and mass media exposure.” (Devos, 2007, p. 205) – Literature review

“The last dimension is the tax system or structure which deals with complexity of the tax system, contact with tax authorities, sanctions, detection probability and tax rates.” (Ho & Wong, 2008, p. 372) – Qualitative research

“Allingham and Sandmo’s (1972) classical model of tax evasion assumes that behaviour is influenced by factors such as the tax rate (which determines the

benefits of evasion), the penalties for fraud, as well as the probability of detection which determine the cost)." (Oberholzer & Stack, 2013, p. 227) – Qualitative research

Tax rate

According to the Leximancer ranked concept list, the fourth most important concept was *tax rate* (count = 4256).¹⁶

The concept *tax rate* (count = 4256) is closely connected to *authorities contact* (count = 65, co-occurrence = 15, connectivity = 23%), *peer influence* (count = 56, co-occurrence = 10, connectivity = 18%) and *tax inducement* (count = 2347, co-occurrence = 289, connectivity = 12%).

“Specifically, an increase in the probability of tax detection is associated with a decrease in the tax rate.” (Hung, 2015, p.206) – Quantitative research

“This is due to an awareness that the analysis research using only deterrence variables (such as tax audits, tax rates, and tax penalties) are not sufficiently able to explain the level of tax compliance.” (Darmayasa & Aneswari, 2015, p. 143) – Qualitative research

Tax Morale

The fifth important concept *tax morale* (count = 3560) is closely connected to *religion* (also defined as religiosity, count = 518, co-occurrence = 91, connectivity = 18%), *culture* (count = 1913, co-occurrence = 170, connectivity = 9%), and *trust* (count = 3497, co-occurrence = 170, connectivity = 5%).

“Most previous studies have used available data including from the *World Values Survey* to study the role of religiosity in tax compliance or, more specifically, tax morale.” (Pope & Mohdali, 2010, p.17) – Literature review

“However, more recent research shows that former *East Germans* demonstrate higher tax morale than former *West Germans* and that “the cultural background seems to have an effect on tax morale”. (Yong 2016, p.297) – Qualitative research

¹⁶ Initially the analysis demonstrated that for the concept ‘tax rate’ in the coded text blocks shows that the term ‘rate’ was used also to denote ‘audit rate’ (362 times), ‘penalty rate’ (205 times), ‘compliance rate’ (214 times), ‘evasion rate’ (51 times), and ‘unemployment rate’ (23 times). We did not isolate coding of other terms for this compound concept as they were also important to the study. The query results show that the term ‘tax rate’ appeared only 4256 times in the coded text blocks, which decreases the ranking for this concept to five instead of two

“Torgler (2003) found evidence that trust in government, pride, and religious conviction have systematic positive influences on tax morale even after controlling for a range of individual characteristics. .” (Hashimzade et al., 2013, p.xx) – Quantitative research

Comparison with Jackson & Milliron (1986) and Richardson & Sawyer (2001)

To further understand the changes of tax compliance concepts between the two decades, a comparison is also made with the Richardson and Sawyer (2001) and Jackson and Milliron (1986) studies in Table 13 .

Table 11: Comparison of Tax Compliance Variables Coded in the 1998 – 2007 Decade and the 2008 – 2017 Decade

Study	Concepts	1 st Decade		2 nd Decade		Total Period	
		1998-2007		2008-2017		1998-2017	
		n	%	n	%	n	%
Jackson and Milliron (1986)	1. age	67	2%	71	1%	138	2%
	2. gender	40	1%	32	1%	72	1%
	3. education	101	3%	116	2%	217	3%
	4. income level	426	14%	593	11%	1019	12%
	5. income source	195	7%	249	5%	444	5%
	6. occupation (or status)	45	2%	93	2%	138	2%
	7. peer influence	27	1%	29	1%	56	1%
	8. ethics	1053	36%	1021	19%	2074	25%
	9. tax fairness	649	22%	894	16%	1543	18%
	10. tax complexity	336	11%	718	13%	1054	13%
	11. authorities contact	49	2%	16	0%	65	1%
	12. sanction	748	25%	718	13%	1466	17%
	13. probability of detection (or tax audit)	1697	57%	2664	49%	4361	52%
	14. tax rate	2282	77%	3241	59%	5523	66%
Richardson and Sawyer (2001)	1. tax compliance cost	1077	36%	1733	32%	2810	33%
	2. tax agents	1333	45%	1678	31%	3011	36%
	3. framing (of tax decisions)	210	7%	222	4%	432	5%
	4. tax inducement (or tax penalties)	1079	36%	1268	23%	2347	28%
	5. tax amnesty	273	9%	496	9%	769	9%
Additional Tax Compliance Variables identified in this study	1. tax evasion	2958	100%	5464	100%	8422	100%
	2. tax system	1628	55%	2757	50%	4385	52%
	3. tax morale	1353	46%	2207	40%	3560	42%
	4. tax avoidance	673	23%	1164	21%	1837	22%
	5. tax enforcement ↑	493	17%	1132	21%	1625	19%
	6. power ↑↑	351	12%	1767	32%	2118	25%
	7. trust ↑↑	599	20%	2898	53%	3497	42%
	8. political	663	22%	902	17%	1565	19%
	9. social norms	526	18%	746	14%	1272	15%
	10. religion (or religiosity)↑	118	4%	400	7%	518	6%
	11. culture ↑	586	20%	1327	24%	1913	23%
	12. tax perception	500	17%	914	17%	1414	17%
	13. perceived opportunity	3	0%	27	0%	30	0%
	14. sole traders ↑↑	505	17%	1934	35%	2439	29%
	15. individual taxpayer↑	291	10%	746	14%	1037	12%
	16. business taxpayer	318	11%	446	8%	764	9%
	17. wealthy taxpayer	27	1%	16	0%	43	1%
	18. reciprocity ↑	90	3%	238	4%	328	4%
	19. withholding taxes↑	101	3%	200	4%	301	4%
Total		23470		41137		64607	

Note:

- n = the number of times a concept is coded within the time period
- % = n divided by the number of times the top concept is coded for that decade

- ↑↑ = Major increase in prominence, ↑ = Some increase in prominence

Table 12 reveals an overall orientation toward *tax evasion*, *tax rate*, *tax system*, *tax morale*, and *probability of detection* as determinants of tax compliance. There was an early emphasis on *tax agents*, *tax compliance cost*, *ethics*, *tax inducement*, *sanction*, *tax fairness* and *political* in 1998 - 2007. In 2008 - 2017, there has been a strong emphasis on *trust*, *power*, *sole traders and culture*. The concepts *tax enforcement*, *religion*, *individual taxpayer*, *reciprocity* and *withholding taxes* also have slightly increased in prominence.

Jackson and Milliron (1986) identified fourteen main compliance variables: two had continued to receive significant attention in the period covered by this study (*probability of detection* and *tax rate*), four had remained prominent (*tax fairness*, *tax complexity*, *ethics* and *sanctions*), while others have not (*age*, *gender*, *education*, *income level*, *income source*, *occupation*, *peer influence* and *authority contact*). Richardson and Sawyer (2001) added five other variables: three had remain prominent since 1998 (*tax compliance cost*, *tax agents* and *tax inducements*), but not *tax amnesty* and *framing (of tax decisions)*. This finding concurs with the suggestion by Jackson and Milliron (1986) that the first four variables of ‘Tax compliance cost, tax agents and tax inducements and tax amnesty’ were potential growth areas in the tax compliance literature.

Our study has identified another 19 variables (1. *tax evasion*, 2. *tax system*, 3. *tax morale*, 4. *tax avoidance*, 5. *tax enforcement*, 6. *power*, 7. *trust*, 8. *political*, 9. *social norms*, 10. *religion / religiosity*, 11. *culture*, 12. *tax perception*, 13. *perceived opportunity*, 14. *sole traders*, 15. *individual taxpayer*, 16. *business taxpayer*, 17. *wealthy taxpayer*, 18. *Reciprocity* and 19. *withholding taxes*). Of which, the tax compliance variables that have remained prominent or received greater attention over the 1998 – 2017 period are:

- *Tax evasion (more a theme rather than a concept)*, *tax system*, *tax morale*, *power* and *trust* had received major attention in the tax compliance literature since 1998
- *Sole traders* and *culture* have increased in prominence in 2008 – 2017 with an increase in likelihood score from 17% to 35%, and 20% to 24%, respectively

- *Individual taxpayers* and *business taxpayers* have remain prominent in 2008 – 2017 (with likelihood scores of 14% and 8%, respectively)
- Although *religion, reciprocity and withholding taxes* concepts have low occurrence scores when compared to other concepts, these concepts have increased in prominence over the last 20 years (with an increase in relevance scores from 4% to 7%, 3% to 4%, and 3% to 4%, respectively)

Overall, although some concepts have lost prominence over the years (*e.g., age, gender, occupation, education, income level, income source, peer influence, perceived opportunity*) further examination revealed that these concepts exist through indirect relationships with other more prominent concepts. For instance:

Age, gender, occupation and religion/religiosity are connected to tax morale:

“More recently tax morale has also been related to the characteristics of the environment in which the taxpayer takes his decision and a substantial effort has been exerted in order to identify external variables that play a significant role in shaping tax morale besides that played by individual characteristics (age, religiosity, gender, marital and occupational status) that have been regularly found to.” (Fillipin, 2013, p xx)

Culture and religion are also connected to *tax morale*:

“The tax morale part of the diagram at Figure 2 is now extended (upwards) to include six main attitudinal factors, including norms (encompassing what some researchers prefer to call “culture”), religiosity, education, attitudes to government/tax authority and to society, and perception of 598 fairness.” (Pope, 2011, p xx).

Peer influence is grouped together with *tax fairness*:

“Based on the fourteen tax compliance variables identified by Jackson and Miliron (1986), Fisher, Wartick and Mark (1992) have grouped ethics or morality, tax fairness and peer influence together representing factors influencing an individual’s tax compliance behaviour.” (Pope & Mohdali, 2010, p xx).

Section Five: Recommendations, Limitations and Conclusions

Recommendations

Undertakings of this magnitude are not expected to be regular events. The growth in studies reviewed from 43 by Jackson and Milliron (1986), to approximately 170 by Richardson and Sawyer (2001), are relatively small in comparison to this study (717) as at the end of 2017, although part of this is that this study considers 20 years compared to ten years by Richardson and Sawyer. To make future reviews of the tax compliance literature less time consuming and timelier, updates should be taken more often, perhaps once each decade (especially given the average number of papers has increased per decade).

With the substantial growth in the number of published studies, it is vital to make use of technology that assists in analysing the research. In this regard use of tools such as Leximancer are essential to not only reduce the sheer number of hours of analysis, but also to ensure the major themes are captured. We would also encourage the next research team to use Leximancer to assist with undertaking comparative analysis with our study.

Limitations of this research and opportunities for future research

With all studies there are limitations in the analysis and scope of the research, as well as areas for future research. This applies even in instances where the study comprises a comprehensive literature review.

Limitations

Leximancer is analysing the whole of the article (including the literature review that many academic articles will have); this means part of the analysis presented is not just the results of the ‘new’ research, but also the summary of the prior research. It would be interesting to see if analysis of the method, results and conclusion parts of the articles wold produce different results.

You still need to go to source document to ascertain the direction of effect (if any) of the concepts discussed and the tax compliance. Additionally, you need to be mindful of jurisdictional differences that could influence whether the results would be applicable in other jurisdictions.

Additionally, articles (especially as they are sourced for a variety of disciplines), may be using inconsistent terms or attaching different meanings to concepts.

In a similar manner to Richardson and Sawyer (2001), no books, discussion papers or conference papers form part of the data analysed. Also, only articles published in English were analysed, along with only peer-reviewed articles with full text from respectable databases. Through these restrictions imposed on the analysis it is expected that some important contributions have been excluded. In some instances these could form part of a future review if the works are subsequently published in peer reviewed journals. In other cases, the absence of these contributions will be a limitation on the analysis.

The substantial increase in published research on tax compliance has also meant the level of analysis reproduced in this paper is significantly less than that of Richardson and Sawyer (2001). That said, the insights that Leximancer provides we believe offers other valuable insights.

Future research

Future research could consider the development of the methodologies utilised over the study period, and compare these to those identified by Richardson and Sawyer (2001). Additionally, future research could consider which theories have underpinned the tax compliance studies from 1988 to 2017. It is such analysis the will form the forthcoming articles from the authors.

Future research could analyse all of the concepts/factors and associated variables in detail, and within the two 10 year blocks. Additionally, studies reported in languages other than English could be analysed. Other databases, beyond the present 15 used could be explored for publications not captured in this study. Additionally, tax compliance studies published in books, as well as potentially conference papers and other respected outlets, could also be analysed.

Other taxes

An analysis of compliance studies about fuel tax, inheritance tax and land tax could occur. The last types of taxes could be particular importance given that we are about to witness the largest inter-generational wealth transfer, as well as arguments for an increase emphasis on land tax given concerns about the mobility of capital.

Conclusions

Part of the importance of research is to add to our understanding of a phenomenon. In this regard it is important to take stock and analyse how knowledge and understanding has evolved. Over 30 years ago Jackson & Milliron set about to review in the USA the current understanding of tax compliance, and in their analysis of 43 paper they identified 14 variables. Then Richardson & Sawyer set about reviewing the decade after Jackson & Milliron from 1986 to 1997; and by doing so reviewed nearly four times the amount of papers from a wider jurisdictional basis and five additional variables were identified.

This current study is the sequel to these prior two studies, and analysis how our understanding of tax compliance has matured over the two decades since Richardson and Sawyer's review. Given the size of the data set of 717 published articles a text mining software, Leximancer, was utilised. This software allowed not only the identification of concepts, but also the potential relationship of concepts to each other, thus providing the ability to determine if there could be any connectivity between concepts and tax compliance. Such software also meant the analysis was done on a more objective basis. In our study not only has there been an increase in the number of compliance variables examined, but the relative emphasis on variables examined by Jackson and Milliron, and Richardson and Sawyer, has altered.

Through this analysis it was identified that in addition to the 19 variables already identified in the prior two studies, that another new 19 variables had been used in research to try to aid our understanding of tax compliance. In particularly, of these new variables it appeared that

While tax compliance is an ever-moving fest, which could continue to change, with such a comprehensive understanding we are better placed about how future research cold consider taxpayer compliance. As tax compliance is an multi-faceted phenomena that requires consideration of many aspects from the political to social norms.

Reference List

Refer to separate document.