

**Toward Assessing Scientific Thinking:  
A qualitative analysis of student reasoning among psychology undergraduates**

Shaughn Bezuidenhout

**Thesis presented in fulfilment of the requirements for the degree of Master of Arts  
(Psychology) at the University of Stellenbosch**



Supervisor: Prof.S A Kagee

Co-supervisor: Associate Professor Saalih Allie, Physics Department and Academic Development

Programme, University of Cape Town.

March 2011

## DECLARATION

By submitting this thesis electronically, I, declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

.....

Signature

.....

Date

Copyright © 2011 Stellenbosch University

All rights reserved

## **ABSTRACT**

The effectiveness of a research methods course on the scientific thinking skills of a group of second year psychology students was recently reported on by Kagee, Allie and Lesch (2010). As part of this study they developed an instrument, The Scientific in Psychology Scale, comprising 11 questions each of which required (a) the endorsement of one of two binary choices and (b) a written explanation detailing the reasons for the choice. However, their findings were based only a statistical analysis of (a) which allowed for a comparison between a control group and an experimental group as a whole. The present study aims to characterize the patterns of thinking at a more detailed level, by analysing the qualitative data for one of the questions. To this end, an alphanumeric scheme was devised to code the data for the two groups mentioned; namely, first year psychology students who comprised the control group, and second year psychology students who comprised the experimental group. The coding was performed at a fine-grained level from which broader categories were constructed.

## OPSOMMING

n Verslag op die effektiwiteit van 'n navorsingsmetodiek kursus op die wetenskaplike denkvaardighede van 'n groep tweedejaar sielkunde studente, was onlangs waargeneem deur Kagee, Allie en Lesch (2010). As deel van hierdie studie het hul 'n instrument ontwikkel, Die Wetenskaplik Denking in Sielkunde Skaal, wat bestaan uit 11 vrae wat elk 'n (a) borg van een of twee binêre keuses en (b) 'n geskrewe verduideliking wat die redes vir die besluit, omskryf. Hul bevindings was egter net gegrond op 'n statistiese analise van (a) wat toegelaat het vir 'n vergelyking tussen 'n kontrole groep en eksperimentele groep as geheel. Hierdie studie beoog om die patrone van denke op 'n meer gedetailleerde vlak te karakteriseer, deur analise van kwalitatiewe data van een van die vrae. Ten eiende dit te bereik, is 'n alfanumeriese skema geskep om die data van die twee reeds genoemde groepe te kodeer; naamlik, eerste jaar studente wat deel gevorm het van die kontrole groep, en tweede jaar studente wat deel gevorm het van die eksperimentele groep. Die kodering was uitgevoer op 'n hoogs gedetailleerde vlak waaruit wyer kategorieë gekonstrueer is.

## ACKNOWLEDGEMENTS

I would like to convey a sincere expression of gratitude to Prof. Ashraf Kagee from the University of Stellenbosch, as well as Prof. Saalih Allie from the UCT Department of Physics. It goes without saying that their lucid insight has made an invaluable contribution to the content of the present thesis. The detailed and thought provoking feedback they provided presented much food for thought and has consequently, I hope, allowed for the completion of a cogent discussion that adds to a large body of literature concerning a very important pedagogical aim.

I am further grateful for the opportunity to partake in discussions with post-graduate students from both the Stellenbosch Department of Psychology, as well as the UCT Department of Physics. These informal exchanges of ideas proved essential as a means to gain a more nuanced and comprehensive understanding of the data, and of the difficulties in teaching students to think critically.

And finally, a much deserved and special expression of gratitude to my family. Without their support, financially and emotionally, this would not have been possible. I am, and will forever be grateful.

## CONTENTS

DECLARATION.....	i
ABSTRACT.....	ii
OPSOMMING.....	iii
ACKNOWLEDGEMENTS.....	iv
CONTENTS.....	v
LIST OF TABLES.....	ix
LIST OF FIGURES.....	x
CHAPTER 1: Introduction.....	1
1.1 Introduction	1
1.2 The Need for Scientific Thinking	1
1.3 Scientific Thinking and Psychology	5
1.4 Purpose of the Present Study	11
CHAPTER 2: Scientific Thinking.....	14
2.1 Literature Review: History of Scientific Thinking	14
2.2 Literature Review: Defining Scientific Thinking	15
2.3 Studies Examining Scientific Thinking Acquisition	21
2.4 Scientific Thinking Assessments	24
CHAPTER 3: Research Methodology.....	28
3.1 Introduction	28
3.2 Research Design	29
3.3 Participants	29

3.4 Instrument	30
3.5 Data Analysis – Alphanumeric Code Scheme	35
3.6 Data Analysis – Quantitative Analysis	57
CHAPTER 4: Results of Analysis.....	59
4.1 Scientific Thinking Framework	59
4.1.1 Music Related Considerations	60
4.1.2 Formalists	62
4.1.3 Evasive	64
4.1.4 More Scientifically Based	66
4.2 Individual Change	68
4.3 More Scientific Reasons for Endorsement at Both Testing Opportunities	72
4.4 Quantitative Analysis	74
CHAPTER 5: Conclusion and Discussion.....	76
5.1 Summary	76
5.2 Conclusion	78
5.2.1 Proxy Analysis	78
5.3 Limitations	81
REFERENCES	83
APPENDIX A: Cognitive Skills Associated with Critical Thinking –The Delphi Report.....	95
APPENDIX B: Geertsen’s List of Revised Micro-thinking Skills.....	97
APPENDIX C: Vignettes Comprising the Scientific Thinking in Psychology Scale.....	96

APPENDIX D: Chi-square Table Representing the Relationship between the Frequency of MSR for Endorsement between the Control Group at Post-test and the Experimental Group at Pre-test.....	103
APPENDIX E: Summary of Code Descriptors, and Frequency of Occurrence, for the Music as the Catalyst Reasons for Endorsement at Pre- and Post-test – Experimental Group.....	104
APPENDIX F: Summary of Code Descriptors, and Frequency of Occurrence, for the Type of Activities Reasons for Endorsement at Pre- and Post-test – Experimental Group.....	105
APPENDIX G: Summary of Code Descriptors, and Frequency of Occurrence, for the More Research Reasons for Endorsement at Pre- and Post-test – Experimental Group.....	104
APPENDIX H –Summary of Code Descriptors, and Frequency of Occurrence, for the Endorsement Based on Personal Experience Reason for Endorsement at Pre- and Post-test – Experimental Group.....	106
APPENDIX I: Summary of Code Descriptors, and Frequency of Occurrence, for the Individual Considerations Reason for Endorsement at Pre- and Post-test – Experimental Group.....	107
APPENDIX J: Summary of Code Descriptors, and Frequency of Occurrence, for the Reliance on the Research Process Reasons for Endorsement at Pre- and Post-test – Experimental Group.....	108
APPENDIX K: Summary of Code Descriptors, and Frequency of Occurrence, for the Confounding Variables Reasons for Endorsement at Pre- and Post-test – Experimental Group.....	109
APPENDIX L; Summary of Code Descriptors, and Frequency of Occurrence, for the Identification of Potential Bias Reasons for Endorsement at Pre- and Post-test – Experimental Group...	111
APPENDIX M: Summary of Code Descriptors, and Frequency of Occurrence, for the Methodological Concerns Reasons for Endorsement at Pre- and Post-test – Experimental Group.....	111



APPENDIX N: Summary of Code Descriptors, and Frequency of Occurrence, for the Reference to  
Technical Terms Reasons for Endorsement at Pre- and Post-test – Experimental Group..112

## LIST OF TABLES

Table 1– <i>Frequency of Music Related Considerations as Reason for Endorsement – Experimental Group</i>	61
Table 2 – <i>Frequency of Formalist Line of Reasoning as Reason for Endorsement – Experimental Group</i>	63
Table 3 – <i>Frequency of Evasive Line of Reasoning as Reason for Endorsement – Experimental Group</i>	65
Table 4 – <i>Frequency of MSR for Endorsement – Experimental Group</i>	66
Table 4.1 – <i>Frequency of MSR for Endorsement – Control Group</i>	67
Table 5 – <i>Number of Respondents Comprising the Individual Change Categories – Experimental Group (n=70)</i>	69
Table 5.1 – <i>Number of Respondents Comprising the Individual Change Categories – Control Group (n=36)</i>	72
Table 6 – <i>Contingency Table Representing the Frequency of MSR for Endorsement between the Experimental and Control Groups at Post-test</i>	74
Table 7 – <i>Chi-square Analysis</i>	75
Table 8 – <i>Extent to which the Selection of the MSO Acts as a Valid Proxy for a MSR for Endorsement, in the Experimental Group, at Pre- and Post-test (n=76)</i>	80
Table 9 – <i>Extent to which the Selection of the MSO Acts as a Valid Proxy for a MSR for Endorsement, in the Control Group, at Pre- and Post-test (n=37)</i>	80

**LIST OF FIGURES**

<i>Figure 1. An Organogram Illustrating the Developed Scientific Thinking Framework</i>	60
<i>Figure 2. A Proxy Table Outlining the Goal of Instruction</i>	78

## **CHAPTER ONE**

### **Introduction**

#### **1.1 Introduction**

The work reported on in the present study forms part of a larger ongoing research programme within the Department of Psychology at Stellenbosch University. The ongoing research programme seeks to determine the extent to which a course in Research Methodology alters the critical reasoning ability of psychology undergraduate students. There has long been a call for education to foster within students the skills and dispositions necessary for effective critical thought (El Hassan & Madhum, 2007). Similarly, the need within psychology education is to foster, within graduates, an understanding of psychological phenomena that sets that understanding apart from lay perceptions. As such, the ability to critically evaluate the veracity of claims relating to psychological phenomena and human behaviour needs to be conveyed as a means to ensure the integrity of psychology as a scientific pursuit. The current chapter will outline and justify the research aims of the present study, with special reference to psychology.

#### **1.2 The need for Scientific Thinking**

Van Gelder (2000) points out that each year, in hundreds of higher learning institutions across the world, students enroll in one semester courses known as critical thinking, informal logic, and introduction to reasoning. The assumption which often leads to enrolment is that these courses will instill within the student an ability to think critically, or to engage more effectively in informal reasoning processes. Formal education has been identified as the primary vehicle through which to facilitate scientific thinking amongst learners, with schooling that promotes scientific thinking skills long being sought after. One of the goals of education is therefore to develop critical thinkers who

can adapt, solve problems and make use of logical/analytical knowledge in their interaction with the world (El Hassan & Madhum, 2007).

Paul (2005) however suggests that the extent to which critical thinking has been included in educational practice highlights three disturbing issues. The author contends that most college faculties do not realise that they lack a substantive concept of critical thinking. Instead, Paul suggests most college faculties believe they understand critical thinking sufficiently and that they are already successfully teaching it within their discipline. The author goes on to conclude that despite “reform” efforts, lectures, rote memorization, and short-term study strategies are still common practice in contemporary college instruction and learning.

The above findings are echoed in Willingham (2008). The author notes that a primary but insufficiently met goal of schooling is the enablement of scientific thinking amongst students. Similarly, Nosich (2005) points out that many text books which incorporate critical thinking exercises do not require any more critical thought to be completed successfully than they do rote regurgitation. The author goes on to specifically identify psychology text as being guilty of this practice, with the majority of textbook exercises advocating critical thinking as only a minor aspect of learning a discipline.

Within the current curriculum, many methods of teaching allow students to progress and progress well, despite a limited ability to engage critically with the course content. This limited extent to which students are able to engage within academic disciplines is inherently problematic if one considers that well-focused instruction to promote critical thinking would be far more memorable and meaningful than skills specific to a particular discipline of knowledge (Yeh, 2001). In a related vein, other authors suggest that the level of critical thinking of many high school graduates

are insufficient to meet the demands of modern society, and that over half the population cannot consistently exhibit even the most basic skills of general reasoning and argument (Hummel & Huitt, 1994; Kuhn, 1991). The reality is therefore that despite a call for an increased emphasis on critical thinking skills, the present situation is one which fails to adequately meet those demands.

The considerations listed above gain special pertinence when one considers the state of affairs which may manifest themselves in the absence of scientific thinking. That is, if one were to contrast scientific thinking with unreflective thinking which as Fisher (2001) suggests, is best defined as the type of thinking undertaken when one jumps to a conclusion or accepts evidence/claims at face value without due consideration. Indeed Ennis suggests that the alternate to scientific thinking is in some respect to, “believe everything that you read and hear” (cited in Davidson, 1998, p. 121).

Advances in knowledge and increasing public exposure to unsubstantiated information, have elevated the importance of critical reading and thinking (Douglas, 2000). As citizens and consumers, students are bombarded with conflicting information, some of which are misleading and/or erroneous (Commeyras, 1993). Similarly, Halpern (2003) claims that the information explosion, environmental considerations, and the role of citizen as chooser, necessitate the acquisition of critical thinking skills amongst students, educators, and the general public alike. The notion of an information explosion, refers to the fact that with readily available information, comes the need to evaluate, interpret, digest and apply that information in a critical manner. Additionally, concerns regarding the impact humans have on the environment have elevated the need for individuals to critically evaluate the manner and means by which they interact with their environment. Lastly, the notion of a citizen as a chooser, refers to the fact that citizens are required to make countless important decisions on a daily basis. For example, the responsibility associated

with democratic participation necessitates such decisions be made with the requisite critical reasoning abilities.

The considerations listed above are clearly delimited from conceptual knowledge of an academic discipline. As such, these considerations serve to illustrate the notion that the debate around critical thinking should not be held within the halls of academia, which often conceptualize critical thinking in terms of educational outcomes, but rather that critical thinking has a role to play in the day to day activity of contemporary society. Indeed, Gieves (1998) suggests critical thinking be considered more as a social practice and not as a decontextualised cognitive skill. West, Stanovich and Toplak, (2008) provide further discussion on the extent to which aspects of critical thinking are related to important real-world considerations, such as personal finance, employment, health, and public policy.

One only need take a cursory glance at contemporary titles to get an idea of the extent to which discussions around critical thinking have pervaded aspects of everyday life. Titles such as, *How to win every argument; The use and abuse of logic* (Pirie, 2006), *Guidelines for critical thinking; Reasons explanations and decisions* (McKay, 2000), *Critical thinking for students* (van den Brink-Budgen, 2000), *Research methods and statistics; A critical thinking approach* (Jackson, 2003), *Critical thinking; How to prepare students in a rapidly changing world* (Paul, 1995), as well as *Critical thinking about sex, love and romance in the mass media* (Galician & Merskin, 2007), abound. Titles such as these, and many others, set out to facilitate within the reader the use of critical thinking when negotiating everyday considerations.

In summary, the present section seeks to highlight the extent to which scientific thinking has become an educational and societal demand. Discussions around scientific thinking are played out

in everyday discourse with the promotion of critical thinking skills not assumed as the sole responsibility of pedagogical processes. Situated within the context as outlined above, the subsequent sections will justify the aim of the present study, with special reference to the need for scientific thinking as a function of psychology education.

### **1.3 Scientific Thinking and Psychology**

Stanovich (1998) suggests that psychology, more than any other science, requires scientific thinking skills that enable students to, “separate the wheat from the chaff that accumulates around all sciences” (p. xvii). There is a large body of literature examining scientific thinking in psychology. Many authors such as Dunn, Halonen and Smith (2008), Stanovich (2004), Lawson (1999), provide insightful commentary on the topic. The following section seeks to outline some of the more relevant considerations that define the debate around critical thinking in psychology, and how these considerations impact on the aims of the present study.

The incorporation of critical thinking into psychology texts has received much attention. Griggs, Jackson, Marek, and Christopher (1998) for example, highlight the fact that 65% of the 37 full length introductory psychology text books reviewed mention scientific thinking as an educational objective. However, research into how psychology students define critical thinking elicit responses very different from those advocated in mainstream literature. The responses which emerged from the psychology students were for example, scientific thinking being the split second one takes to brake before slamming into a car or as thinking which requires effort, or as thinking which is done ‘before I fall asleep, or when I wake up and don’t have to go to class’ (Kuebli, Harvey, & Korn, 2008, p. 137). The authors continue with the suggestion that students of psychology do not understand scientific thinking as a complex construct, containing cognitive processes such as logical reasoning, the evaluation of evidence and the examination of data from multiple perspectives.



The need to override biases in everyday thinking has obvious implications from an educational perspective. Within psychology however, this need is increased exponentially when one considers that decision making based on more than intuition, is vital in the clinical setting (Aegisdottir, White, Spengler, Maugherman, Anderson, Cook, et al., 2006). Similarly, belief biases such as irrational belief persistence have implications that extend beyond the classroom and informal clinical practice (Baron, 2008).

The implications are that educational practice that seeks to facilitate scientific thinking skills needs to be more comprehensive than the simple conveyance of technical knowledge. Jakoubek (1995) for example, highlights the notion that psychology students are frequently unable to apply the understanding they gained about research in psychology to evaluating what they read. Similarly, it is not enough to simply provide students with heuristics that seek to foster critical thinking skills (Bailin, Case, Coombs, & Daniels, 1999), and informing students of cognitive bias does little to ameliorate the impact of those biases (Dunbar, 2000). The promise of improved scientific thinking ability among students has implications that extend beyond the field of psychology. Bessick (2008), for example, points out that universities would undoubtedly benefit from increased scientific thinking skills among students. The author suggests that students who exhibit an ability to engage critically with the course content display the promise of increased retention ability. These students also seem to show increased persistence and graduation rates. They, are likely to be better prepared for entry into the workforce, and are more likely to provide assurance to the public who expect greater accountability with regards to student progress. Finally, these students are likely to foster greater confidence from the wider community regarding the ability of universities to produce productive and responsible members of society (Bessick, 2008).

It is therefore important for the integrity of psychology as a scientific pursuit, that those principles associated with the scientific process be emphasized in curricula seeking to promote the scientific study of psychological phenomena. By placing such importance on an empirically orientated framework, educators not only equip students with the necessary skills associated with critical thinking, they also foster within students the disposition necessary for effective critical thought. A course in RM therefore needs to equip students with both the technical understanding of procedures related to social science research, as well as convey a deeper understanding of the means by which to apply that newly acquired knowledge in novel contexts. Essentially, teaching students to think like psychologists means teaching them to ask questions and examine methodologies as a means to answer those questions (Edman, 2008).

Within psychology, the tendency, for example, to base a potentially life-changing intervention on little more than intuition or clinical experience (see Garb & Boyle, 2003), is a cause for concern. The tendency within psychology to make a decision in the absence of critical reflection or empirical support may lead to the perpetuation of misconceptions and inadvertently inform pseudoscientific practices. The implications within psychology are obvious in that the perpetuation of therapeutic interventions void of any empirical base may lead to the practice of unsubstantiated, fraudulent and potentially harmful therapy (Wade, 2008). The author goes on to suggest that psychology was developed as a means to replace explanations of behaviour on more than whim or wishful thinking, and that from its conception, psychology has had to contend with 'psychobabble' that continues to pervade the discipline.

Pseudo-scientific therapeutic interventions such as, crisis therapy, primal fear therapy, alpha wave therapy, expectation therapy and 'do it now' therapy (Tavris, 2003), are often seized upon by lay people as well as professionals and used without concern for their lack of empirical backing. The

intuitive appeal of pseudoscience contributes to the tendency of pseudoscientific practices to be readily adopted. The proliferation of pseudoscientific practices might be explained in terms of two related notions.

Firstly, many pseudoscientific therapies are often punctuated with reference to the scientific process. Makgoba (2002) suggests that the distinction between pseudoscience and science becomes hazy when one is unable to separate science from an ideology, or when scientific truth becomes merged with an ideology. Others, such as Popper (1963) suggest that the distinction lies in the fact that science is bound by the inductive approach and that the theories it produces are falsifiable. Lillienfield (1998) however points out that science and pseudoscience do not necessarily differ in kind but rather in degree. As such, a ready distinction is not always apparent.

Secondly, as Tavris (2003) points out, pseudoscience fulfills the desire for certainty so prevalent in many societies. The author goes on to suggest that the emphasis placed on immediate relief, gratification and results which pervades contemporary society, has resulted in the mass marketing of empirically untested interventions. Within the context of the South African health field, a proliferation of controversial and unsubstantiated statements have gone on to inform public discourse about HIV/Aids. Authors such as Makgoba (2000) for example, suggest the debate around HIV/Aids has been characterized by absurd, pseudoscientific claims. Similarly, baby Einstein/baby Mozart products have been widely advertised as capable of fostering early childhood development. Sales from the developmental products generate in excess of 200 million Euros per annum (Meredith, 2009). Research conducted has however not found any support for the advertised developmental gains (McKelvie & Low, 2002; Thompson, Schellenberg, & Hussein, 2001). Such empirically unfounded notions have the propensity to filter down through the perpetuation of misconceptions and inadvertently inform pseudoscientific practice. Indeed, the American

Association for the Advancement of Science's (AAAS) Project 2061 warns that without the ability to think critically, 'citizens are easy prey to dogmatists, flimflam artists and purveyors of simple solutions to complex problems' (cited in Zeidler, 1997, p. 483). For a further discussion highlighting the extent to which pseudoscientific principles have come to inform thinking around nutrition, education, and HIV/Aids – with special reference to South Africa – see Goldacre (2009).

In essence, pseudoscience is popular and widely accepted because it has a tendency to confirm that which is widely believed. Science on the other hand, is unpopular because of the tendency it has towards disconfirmation. Science essentially questions that which is believed through the demands that it places on empirically falsifiable observations.

For a further discussion of the role politics and the media play in perpetuating pseudoscientific practice in South Africa, see Makgoba (2002). For a further discussion of examples in psychology relating to the uncritical acceptance of claims concerning behaviour, see Wade (2008), Kagee, Harper and Giorcos (2008) and Kagee (2009), additionally, Lilienfield, Lynn and Lohr (2004) provide an overview of the potentially harmful consequences resulting from the implementation of pseudoscientific techniques in clinical practice.

Jackson (2003) suggests that students often enrol for a course in psychology with the notion that it is not a science. The author goes on to suggest that this misconception may stem from the widely held belief that subject matter alone defines what constitutes a science. Instead, that which constitutes a science should relate to the manner in which something is studied, as well as the content matter. It follows therefore that psychologists should seek to apply principles associated with the scientific method to the study of psychological phenomena as a means to protect the integrity of psychology as a scientific pursuit. It is important that psychology courses prepare

students to think critically about what they read, and develop within students the ability to aptly apply their knowledge in appropriate ways. Such an educational aim should be undertaken with the view to facilitating, within students, specialized knowledge about psychological functioning, that should be different from lay understanding of psychological phenomena (Kagee, 2009).

When one acknowledges that a person's decision-making behaviour is a manifestation of that individual's knowledge structures and his/her epistemological beliefs (Yang & Anderson, 2003), one acknowledges that the educational process cannot simply foster more effective thought without taking into account pre-existing knowledge structures. Nosich (2005) for example, points out that the prior beliefs held by students are more than simply information. The author goes on to suggest that there is a logic that may be correct, erroneous, clear, unclear and as such the educational process is not a question of writing on a blank slate. The problem of facilitating more effective thinking is deeper and more entrenched than that.

It should therefore be noted that critical thinking is not a natural ability for every student (O'Donnel, Francis, & Mahurin, 2008) with the acquisition of critical thinking skills unlikely to develop naturally among psychology students (Kagee, 2007). It should further not be assumed that students are aware of the need for the separation of fact from opinion in their own thinking without being told of the need to do so, given directions on how to do so, and being held accountable for doing so (Osborne, Kriese, Tobey, & Johnson, 2009). The implications are that, without being provided with the appropriate technical skills, or being told of the need to separate fact from opinion, students are unlikely to simply acquire the skills necessary for effective critical thought through 'passive' tertiary attendance (Facione, 1990b). As such, educators need to make a concerted effort to promote the skills and dispositions necessary for critical thinking.

Within psychology departments, a special emphasis is placed on the teaching of critical thinking skills within Research Methodology (RM) classes (Kagee, 2007). Proficiency in RM is an important requirement in the ability to think critically and evaluate empirical claims within psychology. This critical engagement is important as a proliferation of pseudoscientific perceptions and commonly made scientific mistakes are a threat to the integrity of psychology as a scientific pursuit. It is therefore necessary that courses in RM play a role in promoting scientific ways of approaching and engaging with psychology as a discipline. In addition to teaching the technical skills involved in research design, courses in research methods aim to train students to think scientifically and to avoid common errors in thinking. A course in psychology which incorporates training in RM therefore provides an ideal platform from which to determine and improve the extent to which students are able to engage in critical thought processes (Beins, 2008).

#### **1.4 Purpose of the Present Study**

The assumption of this research project is that the concept of critical thinking should be brought to the fore of psychological enquiry, and be encouraged as a means to further psychology as a scientific endeavour. The present study therefore draws its broad purpose from the notion that critical thinking is one of education's most valued goals and central outcomes (van Gelder, 2001). However, assessing the degree to which a course has impacted on the scientific thinking skills is not easy and various instruments have been developed for this purpose. Chapter two will highlight some of the more widely used instruments.

The present study forms part of a larger research project, the purpose of which sought to determine the extent to which a course in RM altered the critical reasoning process of psychology undergraduates. The data were collected via respondents completing a new instrument, *The Scientific Thinking in Psychology Scale (STPS)* (Kagee, Allie, & Lesch, 2010). *The STPS, presented*

in Appendix C comprises 11 questions each consisting of a short vignette followed by two parts to which the students have to respond. The first part requests the respondents to endorse one of two assertions; the second requests required detailed written explanations for the choice. Each binary pair was set so that an endorsement of one option would be regarded as being compatible with scientific thinking while the endorsement of the other would not. The findings reported within Kagee et al. (2010) were, however, based on the binary responses. Thus, based on the statistical analysis that was performed conclusions about the groups as a whole were drawn. The authors report on an analysis of variance showing non-significant differences between the groups at pre-test and a significant difference ( $p < 0.05$ ) at post-test. The results are interpreted to mean that a course in RM was responsible for increasing students' level of scientific thinking. However, the selection of the more scientific binary option, did not in itself represent the presence of a more scientific reason or selection. As such, no clear insights into student reasoning was gained and thus, no detailed pedagogical strategies can be developed in order to improve the effectiveness of the RM course.

The present study aims to complement the previous analysis by investigating a portion of the associated qualitative data. However, no previous such work was carried out in order to act as a guiding framework. Thus, the present work reports on both the development of an analysis approach to such qualitative data (see Chapter three) as well as the detailed findings of the first question that flow from this approach (see Chapter four). It should be pointed out that while the present study is not aimed at validating the STPS instrument the degree to which an endorsement serves as a proxy for scientific reasoning for an individual respondent is a key issue that is discussed. The application of the method that is developed in the present work can therefore serve as a basis to analyze the qualitative aspects of the remaining questions from which the overall

“goodness” of the instrument as a coarse group measure of scientific thinking can then be constructed.

In summary, the present chapter outlined the rationale and purpose of the present study. The subsequent chapter will outline those considerations relevant to the research question as outlined above. More specifically, chapter two will outline the history of scientific thinking, as well as those constructs that have come to define the debate around scientific thinking. Additionally, difficulties associated with the teaching and assessing of scientific thinking will be presented. Thus, Chapter two will also present an overview of some of the presently used measures of scientific thinking, as well as highlight studies examining the extent to which scientific thinking skills have been successfully facilitated. Chapter two will conclude with an overview of the theoretical framework that defines the present research.



## CHAPTER TWO

### Scientific Thinking

#### 2.1 Literature Review: History of Scientific Thinking

As both Scientific and Critical Thinking have many areas of overlapping constructs, and both come to represent subsets of what may be coarsely described as ‘good’ thinking, the present research seeks to make no distinction, nor elaborate on, the largely semantic differences between the two. For the purposes of the present discussion, the terms Scientific Thinking and Critical Thinking will be used interchangeably (O’Donnel et al., 2008).

The history and development of critical thinking has received much attention within existing literature (see Bessick, 2008; Fisher, 2001; Paul, Elder, & Bartell, 1997). The present overview starts with Descartes’s book, *Rules for the Direction of the Mind*. In it, Descartes developed a method of thought based on the principle of systematic doubt. In essence, Descartes argued for a thinking process which is defined by questioning, doubt and continuous testing (Paul et al., 1997). Descartes’ concept of critical thinking is reiterated within contemporary literature by Paul et al., (1997). The authors suggest that critical thinkers analyze thinking, they assess thinking, and as a result they improve thinking. The authors therefore argue for a conceptualization of critical thinking which entails continuous improvement as a result of revision and replacement. The result is that new and better thinking is the consequence of healthy critical thought (Paul, 2005). The process of improving one’s critical thinking abilities by essentially ‘thinking about your thinking’ is often referred to within the literature as ‘metacognition’ (Fisher, 2001).

The notion that effective thought can be regarded as a process of inquiry, and that it incorporates a self-corrective component finds support elsewhere (for example Lipman, 2003). Constant revision

and the improvement of weak thought processes as a means to improve scientific thinking is perhaps best surmised via Popper's notion of falsification (Popper, 1963), Falsification denies the possibility of certain knowledge, instead falsification asserts the notion that nothing is secure, that our knowledge remains speculative, falsifiable, and consequently open to improvement.

With the publication of *Folkways*, William Graham Sumner brought scientific thinking into the realm of modern academia. He highlighted the tendency of schools to serve the (uncritical) function of social indoctrination. Sumner was adamant in his criticism of social institutions, suggesting that schools produce individuals who are all of one mould. A proliferation of uncritical opinion has the propensity to undermine scientific thinking through the perpetuation of popular perceptions which Sumner defined as containing "broad fallacies, half truths, and glib generalizations" (Sumner, 1906, p. 631).

## **2.2 Literature Review: Defining Scientific Thinking**

Since the 1980's, there has been greater emphasis placed on the development of scientific thinking instruction within academia (Ennis, 1993). Arguably the most ambitious attempt to formulate a concrete definition of critical thinking was undertaken at the start of February 1988. The Delphi Report (Facione, 1990a) comprised a panel of 46 experts in critical thinking. The experts were each asked to submit their thoughts about critical thinking to a central mediator. The panelists answered questions which required thoughtful and detailed responses. A set definition was derived through continuous revision, sharing of opinions, and a willingness to consider alternative ideas and explanations.

The Delphi panel identified six cognitive skills as central to the concept of scientific thinking. These skills are, interpretation, analysis, evaluation, explanation, inference, and self-regulation (Facione,

1990a). A detailed description of the skills listed above is contained within Appendix A. The synopsis of the 2-year effort concludes with the suggestion that the ideal critical thinker is;

habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeing results which are as precise as the subject and the circumstances of inquiry permit. (Facione, 1990, p.1)

The conclusions reached by the Delphi report have become widely used and provides the template for two of the most frequently used assessments of critical thinking, namely the California Critical Thinking Skills Test, as well as the California Critical Thinking Dispositions Inventory (Alvarez-Ortiz, 2007).

Research conducted in 1997 suggested that some 89% of college professors surveyed within a California Commission of enquiry into critical thinking claimed that critical thinking was one of their teaching objectives. However, only 19 percent could adequately explain what critical thinking was, and less than 10 percent were clearly teaching it on a regular basis (Paul, Elder, & Bartell, 1997). Similar findings have been reported elsewhere. Ruminski and Hanks (1995) for example, found in a survey of 172 colleges, journalists and mass communication educators that the majority had no clear concept of what critical thinking is, although 89 percent of the educators believed themselves to be teaching it. Similarly Phillips and Bond (2004) highlights the fact that less than half of a sample of final year undergraduates at an Australian University experienced learning in a way that can be regarded as representative of critical thinking.

The significance of these findings suggest that more than two decades since the publication of The Delphi Report, the extent to which critical thinking has been incorporated into pedagogical practice is marred by inconsistencies. Consequently, literature which discusses the extent to which the teaching of critical thinking skills has been successful is punctuated with conflicting views and results. Studies like those mentioned above, leads Paul to bemoan the fact that, “research demonstrates that most college faculties lack a substantive concept of critical thinking, though they mistakenly think otherwise” (2005, p. 27).

Resnick (1987) maintained that while thinking skills resist a precise form of definition; higher order thinking skills can be recognized when they occur. While many would agree that the concept of critical thinking is often understood at a vague, intuitive level, this is of little use to those who seek to conceptualize and measure critical thinking.

Alvarez-Ortiz (2007) points out that two approaches have come to define the contemporary framework for conceptualizing critical thinking. The traditional approach centres on the logical analysis of information, while the ‘second wave’ approach to critical thinking emphasizes the use of imagination, intuition and the active participation of emotions and values in addition to the ability to apply methods of logical analysis (e.g Walters, 1994). The proponents of the ‘second-wave’ have however failed to provide empirical evidence that it constitutes an improved conceptualization of critical thinking (Alvarez-Ortiz, 2007).

The critical thinking process is generally regarded within the literature as containing both a dispositional as well as a skills component (Facione, 1990a; McPeck, 1981; Miri, David, & Uri, 2007). For guidelines as to what constitutes the skills components, the National Science Education Standards suggests that critical thinking involves the

identification of assumptions, use of critical and logical thinking, and consideration of alternative explanations; the analysis of firsthand events and phenomena as well as the critical analysis of secondary sources; testing reliability of knowledge they have generated; and the critical abilities of analyzing an argument by reviewing current scientific understanding, weighing the evidence, and examining the logic so as to decide which explanation and models are best. (cited in Bailin, 2002, p. 361)

The dispositional component is perhaps best summarised through Siegel's (1988) reference to the 'critical spirit'. The critical spirit is taken as one that respects authorities of truth within reason, but does not trust authority blindly, this critical spirit is challenging of authority but does not regard one's own ideas as the sole authority (Wang, Tsai, Chiang, Lai, & Lin, 2008). The critical spirit is further perceived as a disposition or attitude that works in conjunction with critical thinking to prompt the asking of the question, 'why?' (Tyler, 2006). The notion of a critical spirit has therefore come to represent an inquisitive and skeptical attitude which is best coupled with a willingness to consider alternate opinions and review one's own opinion in light of new evidence. For a further discussion of the dispositional component as predictive of critical thinking performance, see Ku and Ho (2009).

John Dewey is widely regarded as the father of modern critical thinking (Fisher, 2001). Dewey referred to the critical thought process as 'reflective thinking' and defined it as, 'Active, persistent, and careful consideration of a belief or supposed form of knowledge in the light of the grounds which support it and further conclusions to which it tends'(cited in Fisher, p. 2).

According to Commeyras (1993), Robert Ennis has developed the most substantive and detailed conceptualization of critical thinking. The definition offered by Ennis (1989) conceptualizes critical

thinking as a reasonable, reflective process that is focused on deciding what to believe or do.

Although many would agree with Ennis' conceptualization, still more would argue that the 'detailed and substantive' conceptualization offered by Ennis, fails to take into account numerous important considerations.

Paul for example, advocates a more rigid definition of critical thinking through the suggestion that effective thought requires discipline and restraint, with the intention of raising our thinking to a level of perfection or quality that is not natural or likely in undisciplined, spontaneous thought (cited in Geertsen, 2003).

Gallagher (1998) defines critical thinking as the ability to assess different paradigms and choose the one which is best suited to the situation at hand. These thoughts are echoed within Halpern (2003), who defines critical thinking in terms of mental processes which increase the likelihood of obtaining a desirable outcome. Griggs, Jackson, Marek and Christopher (1998), reflect the problem solving conceptualization of critical thinking through the suggestion that critical thinking is a process of evaluating evidence for certain claims, and determining whether presented conclusions logically flow from the evidence while considering alternative explanations.

Yeh (2001) conceptualizes critical thinking in terms of argumentation and suggests that the empirical evidence which examines the tendency for children to interact in terms of assertion, defense and negotiation, may be a useful base on which to build critical thinking skills. These thoughts are echoed by Mayer and Goodchild who define critical thinking as 'an active and systematic attempt to understand and validate arguments' (cited in Jakoubek, 1995, p. 57).

Geertsen (2003) has a slightly different conceptualization of critical thinking. Geertsen considers critical thinking and reflective thinking as complimentary forms of higher order thinking.

He therefore bases his conceptualization of critical thinking on the differences between lower and higher level thinking. Bloom, Bertram and Krathwohl (1956) first used this approach in their taxonomy of educational objectives, in which they identified a linear continuum of six skills related to knowledge acquisition. A full table outlining Geertsen's revised taxonomy can be found in Appendix B.

Kuhn (1991) sees critical thinking as involving the abilities to: a) differentiate opinions from evidence, b) support opinions with non-spurious evidence, c) propose opinions alternative to one's own and to know what evidence would support these, d) provide evidence that simultaneously supports one's own opinions while rebutting the alternatives, and e) take an epistemological stance that involves weighing the positives and negatives of what is known.

Skills generally regarded as representative of scientific thinking include: identifying assumptions, identifying and dealing with equivocation, making value judgements, analyzing arguments, asking and answering questions of clarification and/or challenge, judging the credibility of a source (Anderson, Howe, Soden, Halliday, & Low, 2001). Cognitive domains such as causal reasoning, probabilistic reasoning, hypothetical thought, theory justification, assessment of the co-variation of events, and the tendency to think of alternative explanations are also legitimately classified as part of critical thinking (West et al., 2008), as is an emphasis on the inference of appropriate conclusion regarding empirical evidence (Yanchar et al., 2008). Such considerations should be accordingly adopted in the field of psychology if research relating to human behaviour is to be studied in a scientific manner. When inquiry, in any discipline, is not based on sound critical thinking, it may result in error, misunderstanding, myth, or illusion (Paul, 2005).

For the purposes of the discussion presented in this thesis, the definition offered by Griggs et al. (1998) will be adopted. The definition put forth by the authors derives from an analysis of those notions most commonly associated with psychology specific critical thinking. The authors contend that critical thinking is a process of evaluating evidence for certain claims, determining whether presented conclusions logically follow from the evidence, and considering alternative explanations. Critical thinkers exhibit open-mindedness; tolerance of ambiguity, and a skeptical, questioning attitude.

### **2.3 Studies Examining Scientific Thinking Acquisition**

There is a large body of literature examining the acquisition of critical thinking skills. Van Gelder (2001) points out that two strategies have come to conceptualize the debate around the acquisition of critical thinking skills. The indirect approach assumes students pick up critical thinking skills as they progress through their studies. The measured gains for this indirect approach show modest improvements with authors such as Pascarella, suggesting the best estimate is that the first three years of tertiary education (without a specific emphasis on the teaching of critical thinking skills), provide disappointingly slow gains in critical thinking skills of about .55 of a standard deviation or 20 percentile points. (cited in van Gelder, 2001). King too, suggests that many students increase critical thinking skills as a direct result of attending college, most, however, only reach low levels of these skills by graduation (cited in Bessick, 2008).

Others, such as Facione (1990b), suggest that the notion of critical thinking improving as a natural byproduct of college instruction is doubtful. As such, the second approach aimed at fostering critical thinking skills centers around the intensive and purposeful teaching thereof (for example Bessick, 2008; Van Gelder, 2001, 2007). Very often, these specially developed critical thinking courses are comparatively short in nature, usually not longer than a semester. The empirical results



of studies examining improvement over the length of a semester show varied degrees of success. Van Gelder (2000) provides a summary as to the effectiveness of a single semester courses in promoting critical thinking skills. The author ends with the suggestion that it is difficult to make a compelling case as to the effectiveness of teaching heuristics aimed at fostering critical thinking, and that there is a need for more and better research on the topic.

Discussions regarding the extent to which different academic disciplines improve critical reasoning are similarly met with varying interpretations. The section below will seek to highlight some contemporary discussions regarding the extent to which critical thinking skills have been successfully taught.

A comprehensive longitudinal case study undertaken by Miri, David, & Uri (2007) employed a pre, post and post-post research design. The California Critical Thinking Disposition Inventory (CCTDI), as well as the California Critical Thinking Skills Test (CCTST) was used. High school students were divided into three groups. The experimental group (n=57) consisted of science students who were exposed to teaching strategies designed for enhancing higher order thinking skills. Two other groups: science (n=41) and non-science majors (n=79), were taught traditionally, and acted as a control group. Results indicate that the experimental group showed a statistically significant improvement on critical thinking skills components. The authors go on to suggest that if one 'knowingly, persistently and purposely' teaches for the promotion of higher order thinking, there are good chances of success amongst students (Miri et al., 2007).

Renaud and Murray (2008) report on a control group, pre-post test research design that suggests larger gains in critical thinking ability were achieved in a subject specific test (introductory

psychology), than in a test that sought to measure general topics. Studies like these lend credence to the conceptual subject specificity approach to critical thinking.

Alvarez-Ortiz (2007), while examining the extent to which a semester course in philosophy was able to improve critical thinking among students, suggests that there is not sufficient evidence to be confident that philosophy improves critical thinking skills any more than studying other academic disciplines. The thoughts are echoed by others. Lampert (2006), for example, reports on mean scores for tertiary students using the CCTDI, and suggests that there is no significant difference between tertiary students from different academic disciplines. Similarly, Miri et al. (2007) report on non-significant differences between science and non-science majors as assessed by the CCTDI and the CCTST. Others such as Lehman and Nisbitt (1990) however, suggest that training in different academic disciplines have an important influence on the type of reasoning improved, and the extent to which reasoning is improved.

Solon (2006) suggests that a moderate infusion (10 hours of class time activity and an additional 20 hours of homework exercises) of critical thinking material emphasizing active learning principles and guided practice, can produce significant growth in the critical thinking ability of students.

Frijters, Ten Dam and Rijlaarsdam (2008), report that a dialogic approach (group based), resulted in a greater positive effect on the critical thinking competences of students. The dialogic approach represents a learning environment in which emphasis is placed on interaction, mutual understanding and collaborative decision making. The premise assumed within the study is that a dialogic approach is more beneficial on the grounds that it stimulates students beyond the simple conveyance of knowledge, and that questions asked within an interactive setting stimulates further thinking of the issue under discussion. The notion that this collaborative process, characterized by interaction and the sharing of ideas, leads to improved critical thinking ability finds support

elsewhere (for example Anderson et al., 2001; Angeli & Valanides, 2009; Wild & Pfannkuch, 1999).

Van Gelder (2001) reports on findings made possible via the use of an interactive software programme. The *Reason!Able* software programme (later modified into the Rationale software programme) allows individuals who have had no previous instruction in the principles of reasoning and argument, to acquire those concepts and skills relevant to critical thinking. The programme makes use of argument maps or argument trees. An argument map can be best defined as a graphical representation of reasoning in which the relationships among claims are made visually explicit (Van Gelder, 2001). The notion underlying this gradual process of scaffolding and feedback is the quality practice hypothesis (QPH). The QPH is based on the premise that critical thinking skills only improve through extensive, quality practice. The authors report significant gains, over a 12 week period, in critical reasoning ability. The authors go on to suggest that the QPH fosters improvements in critical reasoning that outpace those one might expect through an ‘indirect’ – just being at university – approach. The implications of the results bear promise when one considers the extent to which computer software programs can ease the logistical concerns of group interaction as necessary for critical thinking skills acquisition (for example Frijters et al., 2008), or the extensive demands imposed by advocating an intensive tutorial approach as a means for critical thinking skills acquisition (for example Bessick, 2008). The use of computer-aided methods for teaching critical thinking receives further support elsewhere (for example Osborne et al., 2009; Hopson, Simms, & Knezek, 2001).

#### **2.4 Scientific Thinking Assessments**

The *Watson-Glaser Critical Thinking Appraisal* (WGCTA; Watson & Glaser, 1980) is the oldest and most widely used measure of critical thinking (Bernard et al., 2008). The WGCTA is comprised

from five skills taken as representative of critical thinking, namely inference, recognition of assumptions, deduction, interpretation and the evaluation of arguments. Each of these theoretical constructs are assessed within the assessment, and as such, the *WGCTA* comprises five subscales. A principle component analysis of the psychometric properties of the *WGCTA*, however, reports that a one factor solution accounts for approximately 80% of the total variance across both forms of the assessment (Bernard et al., 2008). This finding leads the authors to conclude that there is no clear subscale structure, and that the *WGCTA* is context specific, performing well with some learners, and poorly with others. Additionally, reliability coefficients vary from 0.23 to 0.73, with guessing associated with the multiple choice response format potentially accounting for the variability in reliability values (Wagner & Harvey, 2003).

A confirmatory factor analysis showed that a two factor model, that is, a model that contains aspects related to both the dispositional and skills components of critical thinking, provided a better fit for the data from a group of university students than did a single factor model (Taube, 1997). The *WGCTA* loaded solely on the skills component, while open ended assessments such as the *Ennis-Weir Critical Thinking Essay Test* (Ennis & Weir, 1985) loaded significantly on both factors. These findings suggest that the dispositional and cognitive components of critical thinking are differentiable and that open ended test formats would capture more of the dispositional aspect of critical thinking than multiple response formats would (Ku, 2008). For a further discussion on the extent to which dispositional components associated with critical thinking contribute to unique variance in critical thinking scores, see Toplak and Stonovich (2003); West et al. (2008).

*The California Critical Thinking Skills Test* (CCTST) comes in two forms, an A and a B. The CCTST seeks to address the constructs of critical thinking outlined in the Delphi Report (Facione, 1990a). These have been presented above. The CCTST was constructed using a bank of 200

multiple choice questions. These were eventually refined to 34 items. These 34 items were designed to reflect three sub-scores, namely; analysis, evaluation and inference (Jacobs, 1995).

Form B was developed by rewriting 28 of the 34 items appearing on Form A. Different terms were substituted, names, concepts, and contexts were changed. The type of topic or problem involved was maintained, as well as the specific critical thinking behavior assessed by the original item (Jacobs, 1995). The CCTST is composed of 19 four-option multiple-choice items and 15 five-option multiple-choice items. Reliability coefficients range from 0.68 to 0.70 (Facione cited in Jacobs, 1995). Further studies have however found that the CCTST to have poor construct validity, unstable reliability, and low comparability between its two forms (Bondy, Koenigseder, Ishee, & Williams, 2001).

The *California Critical Thinking Disposition Inventory* (CCTDI) was developed to assess one's critical thinking disposition. The CCTDI was designed to assess dispositional components such as systematicity, inquisitiveness, self-confidence and maturity (Yeh, 2002). The CCTDI comprises a six point, forced choice Likert scale of agree-disagree alternative responses. The higher the score a respondent achieves, the stronger the critical thinking disposition. Overall sub-scale alpha is reported at 0.90, with alpha of the seven sub-scales ranging from 0.71 to 0.80 (Facione cited in Yeh, 2002)

The *Cornell Critical Thinking Test* (CCTT) consists of two levels, X and Z. The CCTT seeks to assess seven aspects of critical thinking, namely; induction, deduction, credibility, assumptions, semantics, definition, and prediction (Ku, 2008). The test is story-based comprising only multiple choice questions. Level X contains 71 items and is designed for grade 4 to college students. Level Y contains 62 items and is designed for gifted high school and college students. Concerns regarding the construct validity of the CCTT have however been expressed (Michael, Devaney, & Michael,

1980). Additionally, researchers examining the comparison between the WGCTA and the CCTT, unanimously prefer the WGCTA, primarily due to the availability of equivalent forms, data regarding technical characteristics, and the extensive normative data for the WGCTA (Jacobs, 1995)

It may be worth noting that the developed assessment bares a similarity to other psychology specific measures of critical thinking. Most notably the *College of Mount St. Joseph's Psychological Critical Thinking Exam* (Lawson, 1999). The instrument presented in Kagee et al. (2010) made use of binary options, in the direction of either scientific or non-scientific responses, each were coded 1 and 2 respectively. The maximum score obtainable was therefore 22, and the minimum 11. The construct comprising the scale is presented in greater detail in chapter three.

In summary, chapter two has sought to outline considerations relevant to the teaching of scientific thinking. The literature surveyed suggests that scientific thinking skills can be promoted, but that such a promotion it is not without difficulty. As such, varying empirical results have come to define the research around the acquisition of scientific thinking skills. Chapter three will introduce the research methodology and instrument of the present study; as well as provide a detailed description of the derived alpha-numeric code scheme.

## CHAPTER THREE

### Research Methodology

#### 3.1 Introduction

The original method of data analysis was quantitative in nature, with more emphasis being placed on the endorsed binary answer as opposed to the underlying reason for endorsement (Kagee et al., 2010). Such an analysis assumes that an endorsement of Walker would indicate the presence of a more scientifically sound reason for endorsement, and that an endorsement of Rider was done so for less scientifically sound reasons.

While this approach allowed inferences to be drawn for the groups as a whole, it failed to accurately capture the complexity of student reasoning at respondent level. A qualitative approach was therefore deemed as more appropriate as it allowed for greater insight into the presented reasons for endorsement; thereby providing a rich narrative as to the underlying reasoning process. A Grounded theory approach was adopted with the presented reasons for selection forming the themes that comprise the assigned codes. In addition to the adopted Grounded theory approach, the present research made use of a Chi-square analysis as a means to determine the significance of any identified change in critical reasoning ability.

The present chapter will discuss the instrument and the developed alphanumeric code scheme adopted for the present study. The present chapter will further highlight the participants, design and administrative process of the larger study.

### **3.2 Research Design**

The present research employed a pre-test, post-test control group research design. The specially developed vignettes comprising the, *The Scientific Thinking in Psychology Scale* (Appendix C), were administered at the start, and then again at the end of the fourth academic term (September to November). Respondents were asked to complete the assessment during class time. The assessment took approximately one hour to complete.

With respect to the larger study, the independent variable was the course in research methodology (RM). The RM course sought to instill within the students an understanding of the research process. The course content covers, inter alia, aspects of the research process such as correlation, causality, control and experimental groups, equivalence as well as processes related to research design. The dependent variable was the extent to which a critical thought process emerged post a course in RM. The extent to which a critical thought process emerged was determined via a qualitative analysis of the reasons presented for the binary answer endorsed.

While the second year students were exposed to the independent variable, the first year students were exposed to an introductory psychology course. The nature of the first year course provides a brief introduction to aspects of psychology such as psychotherapy, developmental psychology, psychopathology, group identity and poverty.

### **3.3 Participants**

First and second year psychology students were asked to take part in the study. The first year students (n=37) comprised the control group while the second year (n=76) students acted as the experimental group. The experimental and control groups were informed that participation was entirely voluntary and that non-participation would have no ill-effect on their academic



performance. Participants who completed the assessment at pre, as well as a post-course received a R30 voucher redeemable at the nearby student centre. Both groups of respondents were presented with an outline of the nature and purpose of the research. Informed consent was solicited from respondents who acknowledged to understanding the nature of the research and wished to partake. A qualitative analysis of the reasons for selection among both groups of respondents indicated a great deal of similarity at pre-test, both in terms of the nature of themes present, as well as the relative frequency with which categories emerged. Further, Appendix D highlights no significant difference between pre-test experimental group respondents and post-test control group respondents in the frequency with which more scientific reasons for endorsement emerge. As such, any identified change in scientific reasoning ability is therefore more likely the result of a course in RM – or additional courses comprising the second year syllabus - than it is the result of a course in introductory psychology, or from the cognitive maturity one might expect to result from at least one year's exposure to tertiary education.

### **3.4 Instrument**

The limitations to the construction of a valid assessment of scientific thinking have been discussed in chapter two. It was with these limitations in mind that an 11 item instrument was developed. The *Scientific Thinking in Psychology Scale* (Appendix C) required participants to read vignettes and respond to specific questions that reflect the aspects of critical thinking of interest. The instrument sought to assess the following 9 constructs, as outlined within Kagee et al. (2010) and serve as the most comprehensive conceptualization of scientific thinking with regards the needs of this research. The 9 constructs are as follows; hypothetical counterfactual; the availability heuristic; reversed burden of proof; reliance on intuition; post hoc ergo propter hoc; overreliance on testimony; correlation equals causation; confirmation bias; and hindsight bias. These constructs are discussed in greater detail below.

### **Hypothetical counterfactual**

Typically, when presented with an apparent cause and effect relationship, scientific thinking requires that one imagine whether the effect would also be observed in the absence of the apparent cause. This envisaged hypothetical counterfactual set of circumstances permits the conclusion that the stated or apparent cause is indeed the cause of the observed effect. In experimental terms such a set of circumstances is given expression in the form of a control or comparison group. In the absence of a control or comparison condition it is often considered scientifically incorrect to make a causal attribution, even though change in the apparent cause is observed in tandem with change in the apparent effect.

### **The Availability Heuristic**

West and Stanovich (2007), posit that the override of heuristics and the avoidance of biases are related to critical thinking. The availability heuristic is a rule of thumb or cognitive shortcut where one bases a prediction of an outcome on the vividness and emotional impact of an event rather than on actual probability (Rusco, 2000). People generally make a judgment based on what they remember, rather than complete data. The availability heuristic is particularly used for judging the frequency or likelihood of events. Thus people often remember information about a few cases and assume that this is representative of a population.

### **Reversed Burden of Proof**

The burden of proof in science rests on the person who makes the scientific claim, not on the skeptic or critic (Shermer, 1997). It is therefore inappropriate to expect that the skeptic should demonstrate that a claim is false (for example effectiveness of a new technique). Instead, the proponent of the claim must show that the claim is likely to be true. Thus, if the evidence in favour of the effectiveness of a certain psychological procedure is not forthcoming, a reasonable response

is one of skepticism rather than a retort that there is no evidence against the procedure and therefore the procedure is valid.

### **Reliance on Intuition**

There is an assumption among many psychology students that clinical training will develop in them an intuitive sense about the clients with whom they work. Psychologists are typically called upon to make assessments, diagnoses, and predictions about events pertaining to clients such as future violence, recidivism, hospitalization, diagnosis, prognosis, and suicide attempts. However, in most studies comparing actuarial and clinical methods of prediction, actuarial methods significantly outperformed clinical methods (Aegisdottir et al., 2006). Invariably, objective data such as test results yields outcomes that are superior to what is known as “clinical intuition” (Garb & Boyle, 2003).

### **Post Hoc Ergo Propter Hoc**

*Post hoc ergo propter hoc* is a Latin term that translates as “After this therefore because of this”. This proposition is based on the mistaken notion that simply because one event happens after another, the first event was a cause of the second event (Pinto, 1995). While there are many sequences of events that may be both temporally and causally related, temporality is only one condition out of several, for two events to be causally connected. In and of itself temporality does not equal causation.

### **Overreliance on Testimony**

Without dispute anecdotes are good educational devices. However, they are not generally useful as a basis for generalization or as evidence, as they are typically not representative of the population of cases from which the anecdotes are drawn (Casscells, Schoenberger, & Graboys, 1978). For any

empirically demonstrated relationship there may be outliers that run afoul of the apparent relationship. This does not mean that the relationship is invalid but merely that exceptions to the rule do occur. For example, in general, men are taller than women. However, in some instances women may be taller than men. Similarly, there is an undisputed relationship between smoking and lung cancer. Yet, everyone knows an elderly person who has smoked heavily for decades, but whose health is excellent. An individual case does not invalidate the empirically demonstrated relationship between smoking and cancer. When discussing empirically demonstrated relationships between variables, students may sometimes cite individual cases that go contrary to the data, mistakenly believing that individual cases invalidate such relationships.

### **Correlation Equals Causation**

If two variables are shown to correlate with each other, a naïve explanation would be to say that one causes the other. Thus if the correlation coefficient between shoe size and reading ability among children is found to be high, it would be erroneous to conclude that having large feet causes children to read better. An alternative explanation is that reading well causes feet to grow, but a more likely explanation is that age, associated with cognitive development, results in physical growth as well as an increase in reading ability. Age is therefore a third variable in the equation that is the causal agent. The conclusion therefore is that correlation by itself does not equal causation and is only one of several conditions to be satisfied for a causal relationship to be determined (Kerlinger & Lee, 2000).

### **Confirmation Bias**

Confirmation bias is selective thinking whereby one tends to notice and look for events that confirm pre-existing beliefs, and to ignore or undervalue the relevance of those that contradict those beliefs (Stanovich, 2004). The motivation to find support for preferred beliefs 'often leads a person to

overlook even glaring faults in the data, because it is difficult to find what is not sought (Dawson, Gilovich & Regan, cited in Lam, 2007). The ability to decouple prior beliefs and opinions from the evaluation of evidence and arguments is deemed to be a skill of paramount importance (West et al., 2008). Within the literature this skill is often conceptualized as the ability to avoid myside bias and belief bias. Macpherson and Stanovich (2007) make the distinction in that while myside bias is defined as the tendency to evaluate evidence and test hypotheses in a manner biased towards one's own opinions, belief bias is defined as the tendency of accepting or rejecting an argument based on its believability rather than its logical validity. Belief bias therefore arises from people's factual knowledge about the world, while myside bias arises from reasoning biased towards personal opinions or stances.

### **Hindsight Bias**

Hindsight bias is the tendency to state after an event occurred that the event was predictable (Hawkins & Hastie, 1990). It is an inclination to see past events as being predictable and reasonable to expect after the fact, rather than before they have occurred.

The present study seeks to report only on the first vignette. Respondents were presented with the scenario outlined on the following page and asked to indicate which of the mutually exclusive binary options they most strongly endorse. The respondents were then asked to elaborate on their reasons for endorsement. While an endorsement of Walker may be regarded as the more scientifically sound binary choice, an endorsement of Walker does not in itself necessitate the presence of a more scientific reason for endorsement. For this reason, the data analysis process focused more strongly on the underlying reason for endorsement. These written reflections were examined for emerging themes. Allowing for written elaborations as a means to assess critical thinking ability is in line with literature that suggests a multi-response format, encompassing

multiple choice as well as allowing for written reflections, is more appropriate than a single response format, such as multiple choice alone (Ku, 2008).

### Vignette 1

---

Professor Rider announces to a group of workers that he wants to study the effect of playing music on their productivity. After a week of playing to them he finds that productivity has indeed increased. He then turns up the volume and after another week finds that the productivity has increased further. He thus concludes that music causes an increase in productivity. His colleague Professor Walker however, says this is not a correct conclusion.

With whom do you most strongly agree, Professor Rider or Professor Walker?

Professor Rider

Professor Walker

---

### **3.5 Data Analysis – Alphanumeric Code Scheme**

The Grounded theory method was used as a means to analyze the data. Originally developed by Glaser and Strauss, Grounded theory is a data analysis tool that aims to produce a theory that is grounded in the data that have been collected (Glaser & Strauss, 1967). The coding process outlined within the present research will be grounded in the data. A comprehensive discussion of the Grounded theory approach is beyond the scope of this discussion. For a comprehensive discussion of the history, methodology and advances of Grounded theory, see Glaser and Strauss (1967);

Glaser (1978); Strauss and Corbin (1998). A brief outline with a view to highlighting the most relevant aspect related to the present research will however be discussed below.

The constant comparison of data is the key component to the Grounded theory approach (Hawker & Kerr, 2007). The constant comparative method differs from other qualitative data analysis tools in subtle ways. Content analysis for example, is similar to the constant comparative method, in that data is coded. Instead of grouping the codes however, content analysis counts the frequency of the codes to determine which codes are most cited in the data. The constant comparative method on the other hand, concerns itself more with the development of a theory or set of themes (Leech & Onwugbuezie, 2008). Further, Grounded theory is a suitable method for exploratory and explanatory research but it distinguishes itself from other qualitative analysis methods in that the Grounded theory approach seeks to be more than just descriptive (Payne, 2007). Grounded theory does not concern itself with the discovery of the theory, but rather of a theory that aids understanding and interpretation (Heath & Cowley, 2004). Instead of the prevailing *a priori* approaches to qualitative data analysis, the analytic approach of comparative analysis, allows for the explicit coding of data, as well as the development of theory from the data (Glaser, 1978). The constant comparative method, as suggested by Glaser and Strauss (1967), is outlined in four stages. Namely; comparing incidents applicable to each category, that is, the analyst starts by coding aspects of the data into as many units of analysis as possible. Secondly, integrating these coded categories and their properties, thirdly, delimiting and refining the theory, and finally, writing the theory.

As a means to illustrate the application of the constant comparative method, a discussion of the coding process adopted for the purposes of the present study is presented below. The proposed data analysis was based upon a conceptualized framework for understanding measurement amongst

physics students (Allie et al., 1998; Volkwyn, 2005). The reasons for endorsement provided by the respondents within the present analysis was similarly examined and coded.

Stage one involved coding the data into units of analysis for each category. An example of the proposed coding procedure is as follows. Vignette 1, presented previously, provides a forced choice selection, respondents are able to endorse either Professor Rider, or Professor Walker, and then elaborate on their reason making process for the chosen answer. Respondents siding with Walker were assigned a code of W, as well an arbitrary numerical tag representative of the prominent theme which emerges upon analyses of that respondent's provided explanation. For example, W10, the 10 in this case denotes an agreement with Walker based on the idea that *music is the catalyst* for the reported change in productivity. Similarly, a code of W20 was assigned when a respondent suggested that the *type of work* being done by the research subjects is an important consideration for the reported change in productivity. Slight variations of these themes (*music as the catalyst* and *nature of work* respectively) were assigned codes of W11, W12, W13 and W20, W21 etc. A code of W11 for example, denotes the notion that music is an important consideration but specifically that the, like or dislike of music is an important consideration. Similarly, a code of W12 denotes agreement with Walker on the grounds that music is an important consideration for the reported change in productivity but that more specifically, the type of music is important. Respondents siding with Professor Rider on the basis of music as the catalyst, or type of work, were similarly coded. A code of R10 or R20 respectively denotes agreement on the grounds that music is an important consideration, while R20 denotes the notion that the type of work is important.

The proposed code scheme is presented with a view to identifying as many unique ideas and themes as possible. In many instances respondent relied on more than one reasoning process to resolve the presented problem. This finding is in line with other studies that seek to assess scientific reasoning



amongst students (Sadler & Zeidler, 2005; Yang & Anderson, 2003). In order to capture the nuanced nature of these dual code responses, a secondary code scheme was adopted. As a means to illustrate the construed code scheme, verbatim responses, and the assigned codes, are presented below.

### **Music as the Catalyst (R/W10)**

The following examples serve to illustrate the R/W10 category. A code of R10 or W10 was assigned to those respondents who suggested that the music acted as a catalyst for the reported change in productivity. An endorsement of each of the possible binary options (Rider or Walker) is mutually exclusive. As such, the examples to follow are those of separate respondents.

I agree with Rider because;

*I agree with Prof. Rider, music is very stimulating especially when the music contributes to creating a good atmosphere. (R10)*

I agree with Walker because;

*I feel that the music stimulated the individuals and therefore they could work faster and thus the increase in productivity - The music might have another effect on the individuals if further tests should take place. Therefore professor Rider's statement is not 100% correct. (W10)*

The examples above suggest some skepticism regarding the research findings. We see this through the suggestion that Rider's findings are not 100% correct, and that further tests might be necessary. Such a suggestion would normally incur the secondary code of *mr* (W30 – more research). However, the suggestion that “music might have another effect” clearly illustrates that this

respondent regards the influence of music as the major determinant for the reported change in productivity.

### **Music as the Catalyst – Like of Music (R/W11)**

Respondents were coded as R11/W11 if they endorsed the notion that music is the catalyst for the change in productivity, but more specifically, that productivity is dependent on the like or dislike of music.

I agree with Rider because;

*Music makes work nicer and makes the workers feel like working. If it is music that they like, they will feel happier and work better. Nice music will/can also make the time go quicker. (R11)*

I agree with Walker because;

*This case applies only to a small group of people. To make it a concrete finding, it must be applied to a larger test group. The productivity depends on whether or not the people have a preference for the music. If they do not, productivity will decrease. (W11mc)*

A secondary code of *mc* was assigned to capture the notion that this respondent highlighted methodological concerns around the sample size (W92). Considerations associated with the use of the W90 category will be discussed in greater detail later in the present chapter. It is noted that the instructions within the vignette expressly informed the respondents to base their decisions only on the presented information. For the purposes of the code scheme however, it was felt necessary to incorporate themes which are representative of the knowledge necessary for the successful completion of social science research.

**Music as the Catalyst – Type of Music (R/W12)**

A code of R/W12 was assigned to those respondents who suggested that the type of music is an important consideration for the reported change in productivity. The examples below are representative.

I agree with Rider because;

*The increase in productivity with Prof.Rider's work definitely indicates that music leads to an increase in productivity, but there is not enough research done regarding the type of music. (R12)*

I agree with Walker because;

*The type of music makes a difference. If relaxing music is played and a person is tired, it will decrease productivity. (W12)*

**Music as the Catalyst – Volume of Music (R/W13)**

A code of R13 was assigned to those respondents who suggested that music volume was an important consideration when assessing the presented vignette.

I agree with Rider because;

*There is a clear improvement in productivity after a week and more improvement with an increase in the volume. It may be that the workers work harder because they know they're being tested but this is probably not all too significant. (R13id)*

The example above was assigned the secondary code, *id*, to indicate that the respondent highlighted the notion that potential bias may influence the study.

I agree with Walker because;

*Higher volume of music will not necessarily lead to better productivity. (W13)*

### **Music as the Catalyst – Nature of Question (W14)**

The W14 category was specific to Walker. The example below is representative.

I agree with Walker because;

*Prof. Rider did at the start get the right results for the research question that he asked. If he plays music, productivity increases. But when he changed the volume, he actually changed the question; how does the volume of music influence productivity. He should have instead removed the music to see if productivity again declines as a means to confirm his first results. (W14mc)*

A code of W14 was assigned to capture the notion that this respondent regarded music, but more specifically the volume of music, as an important consideration for the reported change in productivity. The code of W14, as opposed to, *music as the catalyst – volume* (W13) was assigned as it was felt the ‘nature of question’ consideration so prevalent within the above example could not be accurately encapsulated within the W13 category. The identifier, *mc* was assigned as this respondent exhibited methodological concerns regarding the research process. More specifically, the example highlights the idea that a counterfactual condition (W91) would be necessary to determine the extent to which the music is responsible for the reported increase in productivity.

### **Music as the Catalyst – Confounding Musical Considerations (R/W15)**

The above category was assigned the code of R/W15 as representative of the idea that music, but more especially confounding variables in the form of the type or volume of music are important

considerations when evaluating the presented vignette. The code of R/W15 differs from R/W12 and R/W13, in that respondents who were assigned the code of R/W15 were taken to consider the type or volume of music as potentially confounding variables. This is in contrast to R/W12 or R/W13, in which the type or volume of music were considered more as slight variations of the music itself. The examples below are representative.

I agree with Walker because;

*Because there may be other causes that increased the productivity i.e. the type of music that was being played. (W15)*

I agree with Walker because;

*There can be more things than the music that plays a role like, for example, the type of music. Prof. Rider's sample size is not mentioned. (W15mr)*

### **Music as the Catalyst –‘Factual’ Considerations (R16)**

The code of R16 was specific to Rider. The code was assigned when respondents referred to other sources as a means to substantiate their endorsed answer. The example below is representative.

I agree with Rider because;

*It is already proven that music, e.g. Baroque music, stimulates the brain and helps productivity with, especially children who struggle to learn. Prof. Rider's study has also shown the positive influence of music on productivity. (R16rp)*

A secondary code of *rp* was assigned when a respondent exhibited a *reliance on the research process* (R/W60) in addition to another reason for endorsement.

### **Music as the Catalyst – Confounding Variables (W17)**

The code of W17 was assigned when respondents suggested that music may have an important influence on the reported increase in productivity, but that other (non-musical) variables might also be responsible for the apparent increase in productivity. The example below is representative.

I agree with Walker because;

*It would have been a cause for the increased productivity at work but it is not the main reason for it. (W17)*

### **Type of Work as Important (R/W20)**

The code of R/W20 was assigned to those respondents who suggested that the type of work is an important consideration when assessing the presented claim. Very often the type of work category was used in conjunction with another reasoning process. As a result, the secondary code, *wo*, was assigned. The example below is representative.

I agree with Rider because;

*I agree more than I tend to disagree because it would depend on the type of work they are doing. If it is a job that requires a lot of focus and attention like a traffic controller or a magazine writer for example the music could to some extent distract the individual from their job. In general I feel that music does increase productivity as you tend to enjoy the work more because of the good effect of the music. (R10wo)*

The primary code of R10 was assigned as a result of the respondent suggesting music may make work more enjoyable. The secondary code of *wo* was assigned to capture the notion that improvement in productivity would also depend on the type of work being done.

I agree with Walker because;

*It depends on what activities the workers were busy with. Workers' motivations to react to the music can also differ. There can be people who will react differently and whose productivity will decrease. (W20ind)*

The secondary code of *ind* (W50) was assigned to capture the notion that this respondent suggested the tendency for people to react differently to the music, as well as the type of work being done, are important considerations when evaluating the reported change.

### **More Research Necessary (R/W30-R/W33)**

The *more research* code was assigned to those respondents who called for more research/more information in order to substantiate the answer endorsed. The more research category comprises *more research* (R/W30), *more research – hypothetical considerations* (R/W31), *more research – sample considerations* (R/W32), as well as the, *more research – more time necessary* (R/W33) category. The examples below are representative.

### **More Research (R/W30)**

A code of R/W30 was assigned to those respondents who endorsed either Rider or Walker on the notion that more research or further studies were necessary.

I agree with Rider because;

*There is more information needed and more different conclusions can be made – more research `needs to be done. (R30)*

I agree with Walker because;

*You need to do further studies to see if this hypothesis is correct.(W30hy)*

### **More Research – Hypothetical Considerations (R/W31)**

The category, R/W31 was assigned to respondents who called for the fulfilment of certain requirements. It is noted that while the scepticism displayed by these respondents lends their reasoning process to a more critical paradigm, the overarching nature of the arguments presented lends itself more to the evasive, *more research* category. The example below is representative.

I agree with Rider because;

*If the playing of music is the only factor that can contribute to the increased productivity, then it is the playing of music that increased productivity. If Prof. Walker can prove that it is not the playing of music that increased productivity, then he can say that the conclusions are incorrect. (R31)*

### **More Research – Sample Considerations (W32)**

Respondents were assigned a code of W32 when they called for more research to be conducted, but more specifically referred to groups or samples. The W32 category was only applicable to Walker endorsing respondents.

I agree with Walker because;

*Prof. Rider cannot just on that one experiment (1group) assume that music has an effect on productivity. (W32)*



**More Research – More Time Necessary (R/W33)**

A code of R/W33 was assigned to those respondents who suggested that more time was necessary to be sure of the findings.

I agree with Rider because;

*For the time length of the study I feel it is necessary to allow a longer time for the efforts to be recorded. (R33)*

I agree with Walker because;

*The period in which these workers were tested and exposed to the music was too short to make a scientific and trustworthy assumption. (W33)*

**Endorsement Based on Personal Experience (R/W40)**

An *endorsement based on personal* experience came to represent those respondents whose reasons for endorsement were derived from a largely subjective experience of music. The examples below are representative.

I agree with Rider because;

*I personally work better and faster if I listen to music. It puts me at ease if I am nervous about a task. (R40)*

I agree with Walker because;

*Music - or any noise – is for me highly distracting. I prefer to learn in dead silence where nothing and nobody can disturb me. (W40)*

### **Individual Considerations (R/W50)**

Respondents who highlighted issues around the notion that individual considerations were an important determinant for the change in productivity were assigned a code of R/W50. The examples below are representative.

I agree with Walker because;

*Not every person is able to cope/produce under the same situations. Thus, some people may be productive others not. (W50)*

The example below serves to illustrate the use of the secondary code, *ind*. The respondent below endorsed *music as the catalyst* (R10), in addition to the suggestion that an individual's response to the music will vary.

I agree with Rider because;

*If music increases the productivity levels then it is clearly a motivating factor for the learners. It must be kept in mind that that not everybody's response will be the same. Individuals differ from one another. (R10ind)*

### **Reliance on the Research Process (R60-R62 and W60)**

A code of R60 was assigned when respondents exhibited a *reliance on the research process*. Slight variations include a reliance on the *research process – burden of proof* (R61) and *reliance on the research process – repeatability* (R62). The R61 code highlights issues around the burden of proof in social research. Typically, the burden of proof falls on the claimant. With respect to the vignette the onus is therefore on Rider to provide evidence as a means to substantiate his findings.

Respondents were assigned a code of R61 when they called for the onus to be on Walker to provide

some sort of evidence to back up the criticism leveled against Rider's findings. The code of R62 (reliance on the research process – repeatability) was assigned when respondents exhibited a reliance on the research process, but more specifically, suggested that the subsequent increase in volume and associated increase in productivity, or that the notion that the findings were repeated, lends credibility to Rider's results. The examples below are representative.

### **Reliance on the Research Process (R/W60)**

The respondents below were assigned the code of R/W60 (*reliance on research process*) as a result of the suggestion that the presented information has been experimentally proven. The examples below are representative.

I agree with Rider because;

*The results of the experiment confirm Mr Rider's hypothesis. (R60)*

The secondary code for the *confounding variables* category (*cv*), was assigned to those respondents who, despite exhibiting a reliance on the research process, suggested that confounding variables should be taken into account.

I agree with Walker because;

*Other factors still need to be taken into account despite being experimentally proven.*

(W60cv)

### **Reliance on the Research Process – Burden of Proof (R61)**

The following examples serve to highlight the reliance on the research process – burden of proof category. Respondents assigned a code of R61 suggested that the onus should be on Walker to refute Rider's findings.

I agree with Rider because;

*Prof. Rider has made his conclusion based on experimental results, thus concluding evidence for his findings. No info is given about how or why Prof. Walker disagrees with Prof. Rider. (R61)*

I agree with Rider because;

*He was the person conducting the experiment & therefore has 'first hand' evidence for his conclusions. On what basis does Prof. Walker draw his conclusions...? (R61)*

### **Reliance on the Research Process – Repeatability (R62)**

The following two examples serve to illustrate the R62 category. The code of R62 was assigned when a respondent suggested that the consecutive increase in productivity, or the notion that the test was performed more than once, lends support to the research process.

I agree with Rider because;

*It shows very clearly that music had a positive outcome on the workers' productivity. If it was just tested once, I would not say that it is necessarily a correct conclusion, but it is tested on another occasion and once again showed the same outcome (actually better).(R62)*

I agree with Rider because;

*Prof. Rider has tested this hypothesis over a short period and with every testing the productivity of the workers has increased.*(R62)

### **Confounding Variables (W70-W74)**

The category of W70, *confounding variables*, differs from R/W15 (confounding musical considerations) and R/W17 (music as the catalyst – confounding variables). The code of R/W15 was assigned when respondents suggested that ‘other’ variables; specifically the type or volume of the music was responsible for the reported change in productivity. A code of R/W17 was assigned when respondents suggested that music was in some way responsible for the reported change, but that other (non-musical) variables should be taken into account. A code of W70 was assigned to those respondents who suggested confounding variables need to be taken into account.

A code of W71 was assigned to those respondents who suggested confounding variables may be responsible for the reported change and that a control/comparison group was necessary to be sure of the findings. A code of W72 was assigned when a respondent specifically identified the Hawthorne effect as a potentially confounding variable. A code of W73 was assigned to those respondents who suggested specifically that confounding variables in the form of group cohesion might be responsible for the reported change. A code of W74 was assigned to those respondents who suggested the notion that correlation does not equate to causation. The examples below are representative.

### **Confounding Variables (W70)**

A code of W70 was assigned to those respondents who suggested music may be in no way responsible for the reported change and that confounding factors should be taken into account. An example is presented below.

I agree with Walker because;

*There can be many other external factors that played a role that Prof. Rider has not taken into account. (W70)*

### **Confounding Variables – No Control Group (W71)**

A code of W71 was assigned when respondents identified potentially confounding variables and made explicit reference to a control group. The example below is representative.

I agree with Walker because;

*There is no control group so other factors that could influence the study are not excluded.(W71)*

### **Confounding Variables – Hawthorne effect (W72)**

A code of W72 was assigned when a respondent specifically identified the Hawthorne effect as a potentially confounding variable.

I agree with Walker because;

*It is possible, but are there other factors that have been taken into account that could influence productivity? e.g. Hawthorne effect.(W72)*

### **Confounding Variables – Group Cohesion (W73)**

A code of W73 was assigned when respondents suggested that confounding variables in the form of group cohesion was responsible for the reported increase in productivity.

I agree with Walker because;

*Both Professor Rider and Professor Walker are right to some context, however I agree with Professor Walker because as much as Professor Rider argues/concludes that music can cause increase in productivity it is not fair to make it a full conclusion, because there are probably reasons why productivity has increased eg it could be group cohesion within the members or other reasons. (W73)*

### **Confounding Variables – Correlation does not equal Causation (W74)**

A code of W74 was assigned to capture the notion of ‘correlation does not equal causation’.

I agree with Walker because;

*There might well be a positive correlation between music and productivity, but that does not mean it is a causal relationship. (W74)*

I agree with Walker because;

*There seems to be a correlation between the music played to the students and the level of productivity, however there is no control group so an accurate conclusion of the effect of the music as the only variable cannot be made. (W74co)*

### **Identification of Potential Bias (R/W80-W82)**

A code of R/W80 was assigned when respondents highlighted the fact that Rider informed his participants that they are going to be studied. The secondary code of *id* was assigned when the identification of potential sources of bias theme emerged in conjunction with another category. The examples below are representative. A code W81 was assigned when a respondent expressed concerns regarding potential bias and made explicit reference to the lack of/need for, a control

group. A code of W82 was assigned when a respondent expressed concerns regarding potential bias and made explicit reference to the Hawthorne effect. The examples below are representative.

### **Identification of Potential Bias (W80)**

A code of R/W80 was assigned when respondents highlighted the fact that Rider informed his participants that they are going to be studied. The secondary code of *id* was assigned when the identification of potential sources of bias theme emerged in conjunction with another category. The examples below are representative.

I agree with Walker because;

*The workers had been told that the researcher wants to find out if music in any way influences productivity. This could be the possible reason why productivity was high. (W80)*

I agree with Walker because;

*Productivity could have increased for other reasons too, especially the reason that the workers knew they were being watched. (W80)*

### **Identification of Potential Bias – Need for a Control Group (W81)**

A code W81 was assigned when a respondent expressed concerns regarding potential bias and made explicit reference to the lack of/need for, a control group.

I agree with Walker because;

*The fact that he said what the study was about could have meant that the students worked harder, and there is no mention of a control group. (W81)*



### **Identification of Potential Bias – Hawthorne Effect (W82)**

A code of W82 was assigned when a respondent expressed concerns regarding potential bias and made an explicit reference to the Hawthorne effect.

I agree with Walker because;

*The Hawthorne effect may be taking place (influence of researchers on participants). (W82)*

I agree with Walker because;

*Professors Rider's results are the results of the Hawthorne effect, where people's behaviour changes when they are monitored. (W82)*

### **Methodological Concerns (W90-W96)**

A code of W90 was assigned when a respondent expressed concerns regarding the extent to which the findings could be generalized. Calls for the assessment of productivity in the absence of music were assigned W91 (counterfactual condition). Concerns expressed by respondents regarding the sample used by Rider were assigned a code of W92. Calls for a control or comparison group were coded as W93. The suggestion that the improvement at Post-test may be due to the idea that the sample had time to practice was assigned a code of W94. Methodological concerns regarding constructs/operational definitions were assigned a code of W95. A code of W96 was assigned when a respondent made explicit reference to the lack of a Pre-test as a methodological flaw. The examples below are representative.

### **Methodological Concerns –Generalized (W90)**

A code of W90 was assigned when a respondent expressed concerns regarding the extent to which the findings could be generalized.

I agree with Walker because;

*Hy kan ook nie die gevolgtrekking na die totale populasie veralgemeen nie, omdat die eksterne geldigheid laag is. (W90)*

### **Methodological Concerns – Counterfactual Condition (W91)**

Calls for the assessment of productivity in the absence of music were assigned a code of W91 (counterfactual condition).

I agree with Walker because;

*He should maybe first have removed the music to see whether productivity increased in order to confirm his first results. (W91)*

### **Methodological Concerns – Sample Concerns (W92)**

Concerns expressed by respondents regarding the sample used by Rider were assigned a code of W92.

I agree with Walker because;

*He might have studied the group of workers who likes music, so they may have enjoyed working while music is playing. (W92)*

### **Methodological Concerns – Control/Comparison Group (W93)**

Calls for a control or comparison group were coded as W93.

I agree with Walker because;

*There are not two groups involved so you cannot make a comparison. (W93)*

**Methodological Concerns – Practice (W94)**

The suggestion that the improvement at post-test may be due to the idea that the sample had time to practice was assigned a code of W94.

I agree with Walker because;

*They may have acquired more information in the time span of this test & have had more time to practice that is why they have done better.*

**Methodological Concerns – Operational Definition (W95)**

Methodological concerns regarding constructs/operational definitions were assigned a code of W95.

I agree with Walker because;

*Workers should be operationally defined eg what if they are deaf workers. (W95)*

**Methodological Concerns – Pre-test (W96)**

A code of W96 was assigned when a respondent made explicit reference to the lack of a Pre-test as a methodological flaw.

I agree with Walker because;

*There was never a pre-test done to see what the productivity of the workers was before the test. (W96)*

**Methodological Concerns – Maturation (W97)**

I agree with Walker because;

*The increase in productivity can also be the result of maturation.(W97)*

### **Reference to Technical Terms (W100)**

Terms such as control group and Hawthorne effect were used as a means to substantiate or guide the presented reasoning process. Examples of these have been presented above. The category of W100 was assigned when respondents made reference to technical terms (control group or Hawthorne effect, Pre-test, independent/dependant variables), but failed to apply those terms correctly or to elaborate on the relevance of the term to the presented scenario. Respondents who were assigned a code of W100 therefore often used technical terms associated with research methodology in a haphazard/erroneous way

I agree with Walker because;

*It is called the Hawthorne effect. Workers like the music and know that if they work hard for a little bit, they will be allowed to permanently listen to music. (W100ha)*

I agree with Rider because;

*This study has a pre-post test. Therefore Prof. Rider can make these conclusions. (W100pr )*

### **3.6 Data Analysis – Quantitative Analysis**

In addition to the qualitative analysis, a Chi-square test was performed as a means to determine whether or not there was a significant association between the type of training, and the extent to which respondents presented more scientific reasons (MSR) for endorsement. The Chi-square test revealed significantly greater than chance differences. The results of the Chi-square analysis will be discussed in greater detail in the following chapter

In summary, the present chapter sought to discuss the instrument and methodology adopted for the present study. As a means to illustrate the generated code scheme, representative examples were

presented. Chapter four will present the results of the generated code scheme as applied to the data. Chapter four will further highlight the developed Scientific Thinking framework that emerges as a result of the application of the code scheme as outlined above; as well as present the results of the Chi-square analysis.

## CHAPTER FOUR

### Results of analysis

#### 4.1 Scientific Thinking Framework

Themes construed from the data were highlighted in chapter three. In summary, the categories that emerged are; *Music as the catalyst* (R/W10), *Type of work* (R/W20), *More research* (R/W30), *Endorsement based on personal experience* (R/W40), *Individual considerations* (R/W50), *Reliance on the research process* (R/W60), *Confounding variables* (W70), *Identification of potential sources of bias* (R/W80), *Methodological concerns* (W90), as well as the *Reference to technical terms* (R/W100) category. The present chapter will illustrate the relationships between those emergent themes when conceptualized in terms a framework that comes to represent the reiterative nature of scientific thinking.

For example, *reliance on the research process*, an *endorsement based on personal experience*, and the notion that *music is the catalyst*, came to represent music-related considerations and were considered as less scientific reasons (LSR) for endorsement. Similarly, those respondents who used technical terms associated with the research process incorrectly, the Formalists, as well as those respondents who were Evasive in the presented reason for endorsement, were all conceptualized as exhibiting LSR for endorsement. If one were to contrast these less scientific reasons for endorsement with the more critically orientated reasons for endorsement, themes such as the *identification of potential sources of bias*, *methodological concerns* and the *acknowledgment of potentially confounding variables* come to the fore. An endorsement based on the above-mentioned reasons would lead the more critically orientated respondents to disregard Rider's findings on the basis that the methodology reported in the vignette is flawed. Failing the ability to recognize these potential flaws (such as telling the participants about the nature of the research, or failing to control

for potentially confounding variables), it was hoped that the more critically orientated respondents would suspend judgment in accepting the results of Rider's findings (Edman, 2008). The extent to which these themes manifest themselves within the construed code-scheme helped place the emergent themes along a framework. Figure one illustrates that framework.

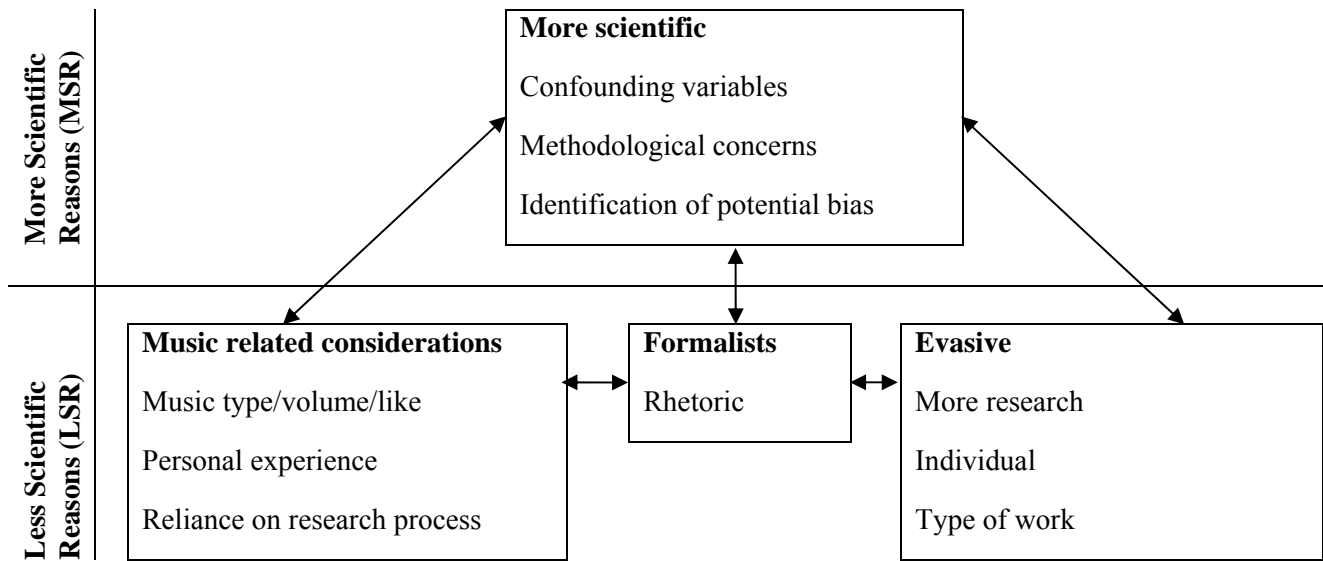


Figure 1. An Organogram Illustrating the Developed Scientific Thinking Framework

The identification of emergent themes, as well as the conceptualization of those themes along the scientific thinking framework outlined above, allowed for an examination of the extent to which respondents altered between their pre and post-test reasons for selection.

#### 4.1.1 Music Related Considerations

The, *music as the catalyst category*, as well as the *endorsement based on personal experience*, and the *reliance on the research process* categories, were collectively deemed as representative of music-related considerations. That is, reasons for endorsement comprising this broader category were regarded as derivatives of music related considerations. For example, *music as the catalyst*, came to the fore amongst those respondents who regarded music as an important consideration for

the reported change. Similarly, an *endorsement based on personal experience*, characterized a reason for endorsement based on the subjective effect of music on the respondent. Finally, an endorsement characterized by a *reliance on the research process*, came to represent those respondents who regarded Rider's study as re-affirming the positive effect of music on productivity.

Table 1, which follows, illustrates the frequency with which the themes comprising the Music related considerations emerge at pre and post-test. Most significantly, the table illustrates the fact that one respondent exhibited a personally based reason for endorsement at post-test, compared to 11 at pre-test. Such a finding is positive in that these respondents were, post a course in RM, less inclined to bring their personal considerations to bear on the presented vignette, but rather, to base their reason for endorsement on the evidential and contextual considerations of the vignette. Further, fewer respondents were inclined to suggest that music, or some derivative thereof, type, volume etc, were important considerations when assessing the reported change.

Table 1

*Frequency of Music Related Considerations as Reason for Endorsement – Experimental Group*

Code descriptor	Code	Frequency of occurrence	
		Pre-test	Post-test
Music as the catalyst	R/W10-R/W17	40	23
Personal experience	R/W40	11	1
Reliance on the research process	R/W60-62	9	11
Total		60 (60%)	38 (41%)

Therefore, at pre-test, amongst the respondents in the experimental group, music related considerations emerged as an important consideration for resolving the presented scenario 60% of



the time. This figure dropped to 41% at post-test. A noteworthy exception is the fact that there was an increase in the number of respondents who exhibited a reliance on the presented research process. Of the 9 respondents at pre-test who endorsed Rider on the notion that the presented research yields a valid result, four endorsed Walker at post-test for more scientifically sound reasons, while one endorsed Walker at post-test for less scientific reason. Of the remaining four respondents, one endorsed Rider at post-test for reasons other than a reliance on the research process, the remaining three respondents exhibited no change in the presented reason for endorsement and again exhibited a reliance on the research process at the second testing opportunity. Of the eight respondents who therefore comprise the new 'additions' to the post-test R/W60 category, three endorsed Rider at pre-test for reasons other than a reliance on the research process, one endorsed Walker in a less scientifically sound manner. The remaining three respondents all exhibited an endorsement of Walker in a more scientifically sound manner at pre-test and therefore exhibited a LSR for endorsement at post-test compared to pre-test. Possible reasons as to why a more critical reason for endorsement at pre-test did not in itself guarantee a similarly sound reason for endorsement at post-test, will be discussed in greater detail later in the present section.

#### **4.1.2 Formalists**

Table 2, to follow, refers to those respondents who endorsed Walker in a formalistic, but unsubstantiated manner. Respondents in this category applied terms associated with the research process in an incorrect/haphazard way.

Respondents who suggested that Rider had done sufficient research to prove his hypothesis (R60hy), or that more research was necessary to prove his hypothesis (W30hy), displayed a superficial understanding of the concepts relating to hypothesis-testing. In social science research

there is much emphasis on the fact that one deals with probabilities and not certainties. Hypothesis are therefore never proved as correct, but are instead accepted as the best standing explanation for an observed phenomena until the fallibility thereof is tested (Field, 2005). Four out of the eight respondents who made use of technical terms in an erroneous manner, did so via the incorrect use of the term, hypothesis. The examples below are representative.

I agree with Rider because;

*Because he had a hypothesis and was able to prove it, that makes him eligible to make the conclusion. He did not just assume. (R60hy)*

I agree with Walker because;

*You need to do further studies to see if this hypothesis is correct. (W30hy)*

Table 2

*Frequency of Formalist Line of Reasoning as Reason for Endorsement – Experimental Group*

Code descriptor	Code	Frequency of occurrence	
		Pre-test	Post-test
More research to prove hypothesis correct	W30hy	-	2
Rider has proved his hypothesis correct	R60-61hy	1	3
There is a pre-post test so Rider can draw these conclusions	W100pr	-	1
Called the Hawthorne effect – workers work hard because they like the music	W100ha	-	1
Total		1(1%)	7(8%)

### 4.1.3 Evasive

Table 3, to follow, refers to those respondents whose provided reasons for endorsement could be best described as evasive. Respondents in this category tended to provide rather obvious considerations when elaborating on their reason for endorsement. The examples below are representative.

I agree with Rider because;

*If the playing of music is the only factor that can contribute to the increase in productivity, then it is indeed the playing of music that increased productivity. (R31)*

I agree with Walker because;

*Not every person is able to cope/produce under same situations. Thus, some people may be productive others not. (W50)*

In addition to the *more research* and *individual considerations* category, those respondents who suggested that the type of work was an important consideration comprised the remainder of this theme. The example below is representative.

I agree with Walker because;

*During work that is repetitive it is definitely the case. It will have to be tested with more than one group of workers. It is therefore not an accurate reflection. (W20mr)*

Within the experimental group, at pre-test, an Evasive reason for endorsement emerged, as a percentage of total reasons for endorsement, 20% of the time. At post-test, the frequency with which these reasons for endorsement emerge declined to 9%. While one could argue that calls for more research, or more information, can be regarded as critical in that it represents a tendency of

suspending judgement before reaching conclusions, such a notion is unfounded. For example, an endorsement based on that fact that Rider’s sample size is not mentioned (W32), does not say anything as to whether or not that respondent is critically evaluating the presented vignette. Conversely, an endorsement based on the notion that, ‘workers should be operationally defined, e.g. what if they are deaf workers’ (W95), suggests a more nuanced understanding of sampling procedures. While both reasons for endorsement therefore relate to sampling considerations, concerns expressed toward the methodology, as opposed to calls for more research/information, were considered as more scientific (MSR) in nature. So, while those respondents comprising the *methodological concerns* category – to be discussed in the next section – exhibited specific concerns regarding the methodology adopted, respondents who called for more research did not. The notion that doubting everything, just as accepting everything, are two sides of the same coin that represents ineffective thinking, is highlighted elsewhere (see Korzybski, 2003). Table 3, below, is representative.

Table 3

*Frequency of Evasive Line of Reasoning as Reason for Endorsement – Experimental Group*

Code descriptor	Code	Frequency of occurrence	
		Pre-test	Post-test
Type of work	R/W20	7	0
More research necessary	R/W30-R/W32	9	7
Individual considerations	R/W50	4	2
Total		20 (20%)	9 (9%)

#### 4.1.4 More Scientifically Based

The previous sections discussed those reasons for endorsement that were deemed less scientific. (LSR). Those respondents who presented a more scientific reason for endorsement are discussed below. The more scientific category comprised those students who endorsed Walker on the grounds that potentially confounding variables have not been taken, into account, have expressed methodological concerns, or have identified issues around potential bias. Examples of each of these reasons for endorsement have been outlined in chapter 3.

Table 4, below, refers to those respondents whose elaborated reasons for endorsement were more scientific in nature. Within the experimental group, a more scientifically sound reason for endorsement emerged, as a percentage of total reasons for endorsement, 21% of the time. At post-test, the frequency with which these more critical reasons for endorsement emerged rose to 43%.

Table 4

#### *Frequency of MSR for Endorsement – Experimental Group*

Code descriptor	Code	Frequency of occurrence	
		Pre-test	Post-test
Confounding variables	W70-W74	9	19
Identification of potential bias	W80-W82	5	13
Methodological concerns	W90-W96	7	8
Total		21 (20%)	40 (43%)

For comparative purposes, the frequency with which the more scientifically sound reasons for endorsement emerge within the control group are presented in Table 4.1.

Table 4.1

*Frequency of MSR for Endorsement – Control Group*

Code descriptor	Code	Frequency of occurrence	
		Pre-test	Post-test
Confounding variables	W70-W74	2	1
Identification of potential bias	W80-W82	1	1
Methodological concerns	W90-W96	1	3
Total		4 (8%)	5 (12%)

In summary, the data presented above suggests that the vignette triggered an emotional response in the respondents. Within the experimental group, 80% of pre-test reasons for endorsement, and 57% of post-test reasons for endorsement, were characterized by reasons that can be deemed as less scientific. These responses came to represent a reason for selection based on subjective opinion and personal experience.

The data presented above supports earlier work conducted by Ellsworth and Ross (1983). The authors report on a study that suggest while it is possible for respondents to believe capital punishment is morally wrong, yet effective as a deterrent for serious crime , or morally acceptable yet ineffective, almost none of the sample held such a combination of beliefs. Those who found it ineffective also found it morally wrong and vice versa. The implications are that students are not necessarily aware of the need to separate a belief based on personal experience, from one based on objective data. The notion that students are inclined to present reasons for endorsement that are largely guided by emotive, subjective experience, is supported elsewhere (for example Sadler & Zeidler, 2005; Stanovich & West, 2007).

It must be noted that decision making guided by emotive, considerations does not necessarily equate to 'bad' thinking. Baron (2008) for example suggests that when one stops using the term emotion as a substitute for irrational, emotions often emerge as the goals of decisions and as such can, in moderation, help us to make more effective goal orientated decisions.

#### **4.2 Individual Change**

The frequency with which the themes comprising the framework outlined in Figure 1 emerge, have been discussed above. While an analysis focusing on cumulative frequencies is of use, it does not capture the extent to which individual respondents altered in their presented reasons for endorsement. The analysis presented below seeks to address this shortcoming.

As a means to determine the extent to which a course in research methodology altered the critical reasoning ability of second year psychology students, an analysis of specific change needs to be undertaken. Specific change in the sense that it is employed here, refers to those respondents who presented a more scientific reason (MSR) for endorsement at post-test, compared to the less scientific reason (LSR) presented at pre-test. Those reasons deemed as more scientifically sound have been previously highlighted.

A respondent was regarded as exhibiting a change in reasoning process if a shift emerged between the presented reasons for endorsement at pre compared to post-test. A shift is defined in terms of reasons for endorsement that may be regarded as more scientific versus less scientific. Such a shift is represented as a vertical movement along the conceptualized framework.

Category A represents a positive change. A positive change is regarded as a move toward a more scientifically based reason for endorsement at post-test than what was presented at pre-test.

Category B may be thought of as no change in the presented reasoning process. Category B comprised those respondents who presented a MSR, or LSR for endorsement at pre and post-test. Category C represents those respondents who exhibited a less critical reason for endorsement at post-test than that which was presented at pre-test. Those respondents who exhibited a more scientifically sound reason for endorsement across both testing opportunities were excluded from the analysis. The experimental group contained six such respondents. The control group contained one such respondent. Table 5, to follow, is representative.

Table 5

*Number of Respondents Comprising the Individual Change Categories – Experimental Group  
(n=70)*

	Category	No. of respondents
Positive change	A	28 (40%)
No change	B	37 (53%)
Negative change	C	5 (7%)

Table 5 highlights the fact that 28 of the 70 (40%) respondents in the experimental group exhibited an improvement, in that a more scientifically sound reason for endorsement was presented at post-test than that which was presented at pre-test. That is, they may have endorsed Rider at pre-test (n=13), or Walker in a manner that is less scientifically sound (n=15), but based their post-test reasons for endorsement on more critically orientated notions. The example below is representative. The example is that of a particular respondent at pre and post-test.



Pre-test - I agree with Rider because;

*It was measured via scientific principles and is, according to me, more correct than Prof.*

*Walker's general statement. (R61)*

Post-test - I agree with Walker because;

*Considering there was no pre-test done, the productivity levels cannot be compared to*

*indicate a so called improvement. (W96)*

Category B comprised those respondents who exhibited no change in the extent to which they critically engaged with the presented vignette. Of these 43 respondent, 6 presented MSR for endorsement at pre and post-test. The remaining 37 respondents presented LSR for endorsement at pre and post-test. Therefore, while a course in RM has seemingly proved effective at facilitating a more critically based reason for endorsement amongst 40% of the less scientifically sound respondents, the majority, 53%, of the respondents in the experimental group have not displayed a move toward a more scientifically sound reasoning process. That is, these 37 respondents might have shifted along the framework, but their pre and post-test reasons for endorsement were characterized by either, music related considerations, an endorsement in a formalistic, unsubstantiated manner, or an endorsement representative of an evasive reasoning process.

Of these 37 respondents who exhibited a less scientific reason for endorsement at pre and post-test, 13 endorsed Rider at both testing opportunities, 11 endorsed Rider at pre-test and Walker at post-test in a less scientifically sound manner. Of the remaining 13 respondents, 9 endorsed Walker across both testing opportunities with a less scientific reason for endorsement. The remaining 4 respondents endorsed Rider at post-test and Walker at pre-test. The example below is representative. The example is that of a particular respondent at pre and post-test.

Pre-test: I agree with Rider because;

*I find that music increases my productivity. (R40)*

Post-test: I agree with Rider because;

*Music can increase productivity especially if the music is enjoyed. (R11)*

The five respondents comprising the negative change category displayed a less scientifically sound reason for endorsement at post-test than that which was presented at pre-test. That is, while the reasons for endorsement at pre-test were more scientific in nature, the reasons for endorsement at post-test comprised either an endorsement of Rider (two), or an endorsement of Walker for LSR (three). The example below is representative. The example is that of a particular respondent at pre and post-test.

Pre-test: I agree with Walker because;

*In order to conclude that the above is true, there needed to be a control group. (W93)*

Post-test: I agree with Rider because;

*This is because the workers are motivated to work harder in a music environment. (R10)*

For comparative purposes, four out of 31, 11% of the respondents in the control group displayed a similar improvement in the presented reason for endorsement. This improvement is negated however by the fact that five, 14% of the respondents in the control group, displayed a less critical reason for endorsement at post-test, than that which was presented at pre-test. Further, the vast majority, 75% of respondents in the control group, exhibited a less scientific reason for endorsement across both testing opportunities. The results of the control group are presented below.

Table 5.1

*Number of Respondents Comprising the Individual Change Categories – Control Group (n=36)*

	Category	No of respondents
Positive change	A	4 (11%)
No change	B	27 (75%)
Negative change	C	5 (14%)

The present section sought to illustrate the extent to which individual respondents altered in their presented reasons for endorsement. As a means to illustrate this specific change, the conceptualized framework was introduced and discussed. The data suggests that 40% of respondents in the experimental group changed favorably in that their post-test reason for endorsement were more scientifically sound than those presented at pre-test. The remaining 60% of respondents either exhibited a less scientific reason for endorsement at both testing opportunities (53%), or exhibited an unfavourable change in that the post-test reason for endorsement was less scientific than that presented at pre-test (7%). For comparative purposes, the results of the control group were discussed. The data suggests that while 11% of the respondents in the control group exhibited a similarly favourable change in their presented reason for endorsement, the majority, 89% did not. Considerations as to why it was unlikely that the more scientific reasons for endorsement were maintained across both testing opportunities, will be discussed in the following section.

#### **4.3 More Scientific Reasons for Endorsement at Both Testing Opportunities**

It has been alluded to earlier that a more scientific reason for endorsement at pre-test does not necessarily equate to a similarly sound reason for endorsement at post-test. Across both the experimental and control groups, 59% (10 out of 17) of those respondents who exhibited a MSR for endorsement at pre-test, failed to exhibit a similarly sound reason for endorsement at post-test. An

analysis of these reasons for endorsement failed to yield an identifiable theme as to why a more scientific pre-test reason for endorsement was unlikely to be maintained. It was originally hypothesized that the majority of respondents in the experimental group who comprised the negative change category, did so via the incorrect use of technical terms associated with research methodology. This notion proved unfounded in that only one of the five respondents in the experimental group who exhibited a negative change, did so with reference to terms associated with research methodology. The example below is representative. The remaining four respondents simply endorsed Rider, or Walker in a less scientifically sound manner at post-test.

Pre-test: I agree with Walker because;

*All the variables that can contribute to the productivity have not been taken into account.*

(W70)

Post-test: I agree with Rider because;

*Volume is an independent variable that influences productivity, the dependent variable, in a positive manner. Correlation is present. (R60iv)*

The data therefore suggests that, while the presence of a course in RM has seemingly proved effective at facilitating a MSR for endorsement amongst the experimental group, the absence, or presence of a course in RM, has no direct effect on the likelihood of those respondents maintaining a more scientific line of reasoning with respect to the vignette. A course in RM therefore seems to be independent of the extent to which the MSR for endorsement are maintained across both testing opportunities. However, the notion that a course in psychology may prove detrimental to critical reasoning ability of tertiary students receives support elsewhere (see Standing & Huber, 2003). Such an inference is of course limited by the discrepancy in sample size between the experimental

and control groups. A similar method of analysis is necessary to determine whether or not this notion is generic to the remaining vignettes.

#### 4.4 Quantitative Analysis

In addition to the qualitative analysis a Chi-square analysis was undertaken. The contingency table is presented in Table 6, which follows. The Chi-square analysis indicates highly significant differences between the critical reasoning ability of the control and experimental groups at pre compared to post-test,  $\chi^2(1) = 10.73, p < .01$ . Table 7, on the following page, is representative. As a means to calculate an effect size, an odds ratio measure was computed (Field, 2005). The data suggests that, based on the odds ratio, a respondent was 5.06 times more likely to exhibit a more critically based reason for endorsement at post-test if they received a course in research methodology, than those who did not receive the course.

Table 6

*Contingency Table Representing the Frequency of MSR for Endorsement between the Experimental and Control Groups at Post-test*

Post-test	Course in RM	Control Group	Total
MSR	34	5	39
LSR	42	32	74
Total	76	37	113

Table 7

*Chi-square Analysis Representing the Frequency of MSR for Endorsement between the Experimental and Control Groups at Post-test*

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson chi-square	10.734 <sup>a</sup>	1	.001		
Continuity correction <sup>b</sup>	9.397	1	.002		
Likelihood ratio	11.809	1	.001		
Fisher's exact test				.001	.001
Linear-by-linear association	10.639	1	.001		
N of valid cases	113				

*Notes:*

<sup>a</sup>0 cells (.0%) have expected count less than 5. The minimum expected count is 12.77.

<sup>b</sup>Computed only for a 2x2 table.

## CHAPTER 5

### Conclusion and Discussion

#### 5.1 Summary

Saville, Zinn, Kerr Lawrence, Barron, & Andre (2008) highlight barriers to the teaching of critical thinking through the suggestion that the practical consideration that an increase in the number of topics, and no concurrent increase in the number of days in which to cover those topics, lead many teachers to believe that time would be better spent focusing on course content and less on critical thinking. Similarly, Douglas (2000), while examining the contribution of social psychology to critical thinking, highlights two 'enemies' of critical thinking. Firstly, Douglas highlights the notion of human gullibility through his suggestion that humans tend to believe any idea that is presented to them, unless that idea is blatantly contradictory. Secondly, Douglas highlights the human tendency to stubbornly resist new information. While these two tendencies seem to be contradictory, the author contends that they operate within us at one time or another. These cognitive considerations find support elsewhere. Gambrell (1990) for example, highlights the fact that individuals show a remarkable tendency to believe in their initial judgements. Similarly, Wild and Pfannkuch (1999) suggest that, what we "know" is not only our greatest asset but also our biggest curse. The authors go on to suggest that what we "know" determines where we look, and by desensitizing us to important information, the foundations of what we "know" are often not soundly based. The tendency for people to evaluate and generate evidence, as well as test hypothesis in a manner biased toward their personal opinion has been consistently demonstrated in numerous studies (Stanovich & West, 2007).

Others such as Denfeld-Wood and Petriglieri (2005) provide a discussion of the neural underpinnings of binary thinking. The authors contend that the reduction of alternatives into

simplistic binary options is a fundamental mechanism deeply engraved in our neural tissue, and as such, is an inherent part of human nature. This archaic mode of thinking comes to the fore most clearly through topics that elicit a strong emotional reaction. From this premise, scientific thinking becomes the ability to transcend such a dualistic, reductionist approach, and is instead defined as the capacity to make ever finer distinctions of one's reality. Indeed; Chinn and Bruwer (1993) present a framework that highlights the manner in which students integrate anomalous data to their existing beliefs. Of the seven forms of psychological response, only one represents a full theory change from the beliefs held. The implications are obvious in that changing adherence to pre-instructional beliefs is no easy task. Indeed, some authors such as Zeidler (1997) suggest that the pre-instructional beliefs students hold amounts to intellectual baggage. For further discussion on barriers to the teaching of critical thinking, see Buskist and Irons (2008), Nosich (2005), Stanovich and West (2007).

In essence, teaching for the promotion of scientific thinking skills is a difficult task, with (a) numerous barriers undermining the pedagogical process. (see above); and (b) the assessment of scientific thinking skills not being without limitations (see chapter 2). Educators are therefore placed in a difficult position. As such, the present research sought to characterize the reasons respondents gave when asked to elaborate on their selected binary answer. The construed code scheme proved a useful means by which to characterize those reasons, and allowed for the conceptualization of a scientific thinking framework. The framework in turn allowed for commentary on point (a), as outlined above, and illustrated related themes that collectively undermine the adoption of a more scientific reasoning process among respondents (discussed in chapter four). The present thesis will therefore conclude with a discussion of Proxy values; and thereby point (b), as above.



## 5.2 Conclusion

As a reminder, any single respondent was able to select either Walker, the more scientific option (MSO), or Rider, the less scientific option (LSO). Respondents were then asked to elaborate on their reason for selection. A qualitative analysis revealed that certain reasons for endorsement were more scientific in nature. These reasons were termed as more scientific reasons (MSR). Similarly, reasons for selection that were less scientific in nature, were termed as less scientific reasons (LSR). Figure 1, presented earlier, highlights the means by which reasons for endorsement were characterized as either MSR or LSR. As such, any respondent was able to endorse either Walker or Rider, for reasons that were either less scientific, or more scientific in nature. The MSO, or LSO, was assigned as a consequence of the binary answer selected. The MSR or LSR were derived from the alphanumeric code scheme as outlined in chapter three. The Proxy diagram below is representative.

		Binary answer selected	
		<i>LSO</i>	<i>MSO</i>
Reason Presented	<i>MSR</i>	Rider	Walker
	<i>LSR</i>	Rider	Walker




Figure 2. A Proxy Table Outlining the Relationship between the Binary Answer Endorsed and the Underlying Reason for Selection. The arrow represents the goal of instruction.

### 5.2.1 Proxy analysis

It was hoped that the selection of the MSO, indicates the presence of a MSR. As such, the proxy value would equate to:  $1 - (\text{MSO}/\text{MSR})$ . Ideally the equation would come to 0, and one could therefore

conclude that the binary answer selected can be perceived as a valid proxy for the underlying reason for selection.

Those respondents whose provided reasons for endorsement were more scientific in nature, and who endorsed Walker, were considered as consistent in that the declared decision, that is an endorsement of Walker, may be regarded as a valid proxy of the underlying reason for endorsement. The proxy value is derived from the equation below.

$$\text{Proxy}_{\text{Experimental-Pre-test}} = 1 - \frac{\text{MSR}}{\text{MSO}} = 1 - \frac{11}{40} = 1 - .28 = .72$$

$$\text{Proxy}_{\text{Experimental-Post-test}} = 1 - \frac{\text{MSR}}{\text{MSO}} = 1 - \frac{34}{57} = 1 - .60 = .40$$

The results of the experimental group are presented in Table 8, on the following page. Table 8 highlights the fact that the number of respondents, who endorsed Walker in a manner that is more scientifically sound was 28% at pre-test. This figure, following a course in RM, doubles to 60%. Further, the number of more scientifically sound reasons for endorsement, as a percentage of the total number of endorsements in the experimental group, increased from 14% to 45%. The data therefore suggests that while respondents were more likely to endorse the MSO, for MSR at post-test compared to pre-test, an analysis of the binary answer alone does not accurately reflect the underlying reason for selection.

Table 8

*Extent to which the Selection of the MSO Acts as a Valid Proxy for a MSR for Endorsement, in the Experimental Group, at Pre- and Post-Test (n=76)*

<u>Experimental Group (n=76)</u>				
	MSO	MSR	%	Proxy value
Pre-test	40	11	28%	.72
Post-test	57	34	60%	.40

For comparative purposes, the number of more scientifically sound reasons for endorsement in the control group, as a percentage of total reasons for endorsement, declined from 16% to 14%. The data for the control group is presented below.

Table 9

*Extent to which the Selection of the MSO Acts as a Valid Proxy for a MSR for Endorsement in the Control Group at Pre- and Post-Test (n=37)*

<u>Control Group (n=37)</u>				
	MSO	MSR	%	Proxy value
Pre-test	16	6	38%	.62
Post-test	11	5	45%	.55

In summary, at pre-test the binary answer endorsed acted as a proxy for a MSR for endorsement, across both the experimental and control groups, 17 out of 56 occasions (30%). At post-test, this figure rose to 37 out of 68 (54%). The remaining 31 (46%) respondents who endorsed the MSO at post-test across both the experimental and control groups, were however likely to do so for LSR

As such, an analysis that fails to take into account the qualitative data (for example Kagee et al., 2010) fails to accurately measure the construct under investigation. Assigning a numeric value to a simple binary answer does not accurately account for the absence, or presence, of an underlying

thought process that is congruent with the selected binary answer. Future research across the remaining vignettes would do well to bear this in mind.

### **5.3 Limitations**

While a pre-test comparison between the control and experimental groups indicate a great deal of similarity, in terms of both the reasons for endorsement, and the relative frequency with which those reasons emerge, potential sources of variation that are extraneous to the independent variable cannot be dismissed. More specifically, courses in addition to RM that form part of the second year syllabus might account for part of the increased critical reasoning exhibited by the experimental group. Additionally, the cognitive maturity resulting as a consequence of exposure to at least one academic year at a University might also account for some of the variation in increased critical thinking ability among the experimental group. The data presented in Appendix D however goes some way to mediating this concern. The data presented therein suggests that a comparison of pre-test reasons for endorsement amongst experimental group respondents, and post-test control group respondents, reveal non-significant differences in the frequency of MSR for endorsement. As such, the data suggest the ‘improvement’ in scientific reasoning exhibited by experimental group respondents is more likely a function of a course in RM – or additional courses comprising the second year syllabus - than it is from generic concerns associated with a first year course in introductory psychology, or from the cognitive maturity that may result from at least one year’s exposure to tertiary education.

As only one coder was responsible for the coding procedure, the bias that may result from varying interpretations of the data is a threat to the developed code scheme. It was sought to negate this threat to the validity of the research by undertaking a methodical and repetitious approach to the coding procedure.

The discrepancy in sample size between the experimental and control groups was noted, and any inferences drawn from the results presented above should bear this in mind. Additionally, many studies examining the acquisition of critical thinking skills make use of standardized assessment measures. As suggested earlier, future research might well focus on establishing the validity of the developed instrument.

Finally, while constructs pertaining to scientific thinking were presented, the extent to which the constructs in the developed instrument match those have not to date been verified. As such, the developed instrument cannot be considered a valid and reliable measure of critical thinking in the sense that psychometric confirmation necessitates. The extent to which the developed instrument adheres to expert group validation should form the basis of subsequent research.

## REFERENCES

- American Association for the Advancement of Science (1989). *Project 2061: Science for all Americans*. Washington, D.C.: Author.
- Aegisdottir, S., White, M.J., Spengler, P.M., Maugherman, A.S., Anderson, L.A. Cook, R.S., *et al.* (2006). The meta-analysis of clinical judgment project: Fifty six years of accumulated research on clinical versus statistical prediction. *The Counselling Psychologist*, *34*, 341-382.
- Allie, S., Buffler, A., Kaunda, L., Campbell, B., & Lubben, F. (1998). First-year physics students' perceptions of the quality of experimental measurements. *International Journal of Science Education*, *20*, 447-459.
- Alvarez-Ortiz, C.M. (2007). *Does philosophy improve critical thinking skills?* (Unpublished masters thesis). University of Melbourne, Australia.
- Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science*, *29*(1), 1-32.
- Angeli, C., & Valanides, N. (2009). Instructional effects on critical thinking: Performance on ill-defined issues. *Learning and Instruction*, *19*, 322-334.
- Bailin, S. (2002). Critical thinking and science education. *Science and Education*, *11*, 361-375.
- Bailin, S., Case, R., Coombs, J.R., & Daniels, L.B. (1999). Conceptualizing critical thinking. *Journal of Curriculum Studies*, *31*, 285-302.
- Baron, J. (2008). *Thinking and deciding*. Cambridge: Cambridge University Press.
- Beins, B.C. (2008). Why we believe: Fostering critical thought and scientific literacy in research methods. In D.S. Dunn, J.S. Halonen, & R.A. Smith (Eds.), *Teaching critical thinking in psychology: A handbook of best practices* (pp. 199-210). Sussex, UK: Wiley Blackwell.
- Bernard, R.M., Zhang, D., Abrami, P.C., Sicoly, F., Borokhovski, E., & Surkes, M.A. (2008). Exploring the structure of the Watson–Glaser Critical Thinking Appraisal: One scale or many subscales? *Thinking Skills and Creativity* *3*, 15-22.

- Bessick, S.C. (2008). *Improved critical thinking skills as a result of direct instruction and their relationship to academic achievement*. Unpublished doctoral dissertation, Indiana University, Pennsylvania.
- Bloom, B.S., Bertram, B.M., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York, N.Y: McKay.
- Bondy, K., Koenigseder, L., Ishee, J., & Williams, B. (2001). Psychometric properties of the California Critical Thinking Tests. *Journal of Nursing Measurement, 9*, 309-328.
- Buskist, W., & Irons, J.G. (2008). Simple strategies for teaching your students to think critically. In D.S. Dunn, J. S. Halonen, & R.A. Smith (Eds.), *Teaching critical thinking in psychology: A handbook of best practices* (pp. 49-58). Sussex, UK: Wiley Blackwell.
- Casscells, W., Schoenberger, A., & Graboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine, 299*, 999-1001.
- Chinn, C.A., & Bruwer, W.F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63*(1), 1-49.
- Commeyras, M. (1993). Promoting critical thinking through dialogical-thinking reading lessons. *The Reading Teacher, 46*, 486-494.
- Denfeld-Wood D.J., & Petrigleri, G. (2005). Transcending polarization: Beyond binary thinking. *Transactional Analysis Journal, 35*(1), 31-39.
- Douglas, N. (2000). Enemies of critical thinking: Lessons from social psychology research. *Reading Psychology, 21*, 129-144.
- Dunbar, K. (2000). What scientific thinking reveals about the nature of cognition. In K. Crowley, C.D. Schunn, & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom and professional settings* (pp. 115-140). Mahwah, N.J: Lawrence Erlbaum Associates.

- Edman, L.R.O. (2008). Are they ready yet? Developmental issues in teaching critical thinking. In D.S. Dunn, J.S. Halonen, & R.A. Smith (Eds.), *Teaching critical thinking in psychology: A handbook of best practices* (pp. 35-49). Sussex, UK: Wiley Blackwell.
- El Hassan, K., & Madhum, G. (2007). Validating the Watson Glaser Critical Thinking Appraisal. *Higher Education, 54*, 361-383.
- Ellsworth, P.C., & Ross, L. (1983). Public opinion and capital punishment: A close examination of the views of abolitionists and retentionists. *Crime and Delinquency, 29*(1), 16-69.
- Ennis, R.H. (1987). A taxonomy of critical thinking dispositions and abilities. In J.B. Baron & R.J. Sternberg (Eds.), *Teaching for thinking* (pp. 9-26). New York: Freeman.
- Ennis, R.H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher, 18*(3), 4-10.
- Ennis, R.H. (1993). Critical thinking assessment. *Theory into Practice, 32*, 179-186.
- Ennis, R.H., & Weir, E. (1985). *The Ennis-Weir Critical Thinking Essay Test*. Pacific Grove, CA: Midwest Publications.
- Eric Digests (1988). *Critical thinking skills and teacher education*. ERIC Clearinghouse on Teacher Education. Washington, DC. ERIC Document No. 297033.
- Facione, P.A. (1990a). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction: The Delphi Report*. Milbrae, C.A: California Academic Press.
- Facione, P.A. (1990b). *The California Critical Thinking Skills Test: College level*. Retrieved April 25, 2009, from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED327549>
- Facione, P.A. (1992). *Manual, the California Critical Thinking Skills Test, Form A and Form B*. Millbrae, CA: California Academic Press.



- Field, A. (2005). *Discovering statistics using SPSS*. London: Sage Publications.
- Fisher, A. (2001). *Critical thinking: An introduction*. Cambridge: Cambridge University Press.
- Frijters, S., Ten Dam, G., & Rijlaarsdam, G. (2008). Effects of dialogic learning on value-loaded critical thinking. *Learning and Instruction, 18*, 66-82.
- Galician, M., & Merskin, D. L. (2007). *Critical thinking about sex, love and romance in the mass media*. London: Lawrence Erlbaum Associates.
- Gallagher, S.A. (1998). The road to critical thinking: The Perry scheme and meaningful differentiation. *NASSP Bulletin, 82*(595), 12-20.
- Gambrill, E. (1990). *Critical thinking in clinical practice: Improving the accuracy of judgments and decisions about clients*. Oxford: Jossey-Bass.
- Garb, H.N., & Boyle, P.A. (2004). Understanding why some clinicians use pseudoscientific methods. In S. O. Lilienfeld S. J. Lynn & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 17 –39). New York: Guilford.
- Geertsen, H.R. (2003). Rethinking thinking about higher-level thinking. *Teaching Sociology, 32*(1), 1-19.
- Gieves, S. (1998). A reader reacts. *Tesol Quarterly, 32*, 123-129.
- Glaser, B.G. (1978). *Advances in the methodology of Grounded Theory; Theoretical sensitivity*. University of California, San Francisco.
- Glaser, B.G., & Strauss, A. (1967). *The discovery of Grounded Theory: Strategies for qualitative research*. London: Weidenfeld and Nicolson.
- Goldacre, B. (2009). *Bad science*. London: Fourth Estate.
- Griggs, R.A., Jackson, S.L., Marek, P., & Christopher, A.N. (1998). Critical thinking in introductory psychology texts and supplements. *Teaching of Psychology, 25*, 254-265.
- Halpern, D.F. (2003). *Thought and language: An introduction to critical thinking*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Hawker, S., & Kerr, C. (2007). Doing Grounded Theory. In E. Lyons & A. Coyle (Eds.), *Analyzing qualitative data in psychology* (pp. 87-97). Thousand Oaks, California: Sage.
- Hawkins, S.A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin* 107, 311-327.
- Heath, H., & Cowley, S. (2004). Developing a Grounded Theory approach: A comparison of Glaser and Strauss. *International Journal of Nursing Studies*, 41, 141-150.
- Hopson, M.H., Simms, R.L., & Knezek, G.A. (2001). Using a technology-enriched environment to improve higher-order thinking skills. *Journal of Research on Technology in Education*, 34, 109-119.
- Hummel, J., & Huitt, W. (1994). What you measure is what you get. *GaASCD Newsletter: The Reporter*, 10-11. Retrieved May 20, 2009, from <http://chiron.valdosta.edu/whuitt/papers/wymiwyg.html>
- Jackson, S.L. (2003). *Research methods and statistics; A critical thinking approach*. Victoria: Thomson/Wadsworth.
- Jacobs, S.S. (1995). Technical characteristics and some correlates of the California Critical Thinking Skills Test, forms A and B. *Research in Higher Education*, 36(1), 89-108.
- Jakoubek, J. (1995). Developing critical-thinking skills in Psychology content courses. *Teaching of Psychology*, 22(1), 57-59.
- Kagee, A. (2007, March). *Training psychology students to think critically: Revisiting psychology as science*. Paper presented at the Teaching of Psychology in South Africa Conference, Bloemfontein, South Africa.
- Kagee, A. (2009). South African psychologists' judgments of widely held beliefs about psychological phenomena. *South African Journal of Psychology*, 39, 195-201.

- Kagee, A., Harper, M., & Giorcos, G. (2008). Judgments of widely held beliefs about psychological phenomena among South African postgraduate Psychology students. *South African Journal of Higher Education*, 22, 789-798.
- Kagee, A., Allie, S., & Lesch, A. (2010). Effect of a research methods course on scientific thinking among psychology students. *South African Journal of Psychology*, 40, 272-281.
- Kerlinger, F., & Lee, H. (2000). *Foundations of behavioral research*. Orlando, FL: Harcourt College Publishers.
- Korzybski, A. (2003). *Manhood of humanity: The Science and art of human engineering*. New York: E.P. Dutton and Company.
- Ku, K.Y.L. (2008). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity*, 4(1), 70-76.  
doi:10.1016/j.tsc.2009.02.001.
- Ku, K.Y.L., & Ho, I.T. (2009). Dispositional factors predicting Chinese students' critical thinking performance. *Personality and Individual Differences*, 48(1), 54-58.  
doi:10.1016/j.paid.2009.08.015
- Kuebli, J.E., Harvey R.D., & Korn, J.H. (2008). In D.S., Dunn, J.S., Halonen, & R.A. Smith (Eds.), *Teaching critical thinking in psychology: A handbook of best practices* (pp. 137-148). Sussex, UK: Wiley Blackwell.
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Lam, C. (2007). Is Popper's falsificationist heuristic a helpful resource for developing critical thinking? *Educational Philosophy and Theory*, 39, 432-448.
- Lampert, N. (2006). Critical thinking dispositions as an outcome of arts education. *Studies in Art Education*, 47, 215-228.
- Lawson, T.J. (1999). Assessing psychological critical thinking as a learning outcome for psychology majors. *Teaching of Psychology*, 26, 207-209.

- Leech, N.L., & Onwugbuezie, A.J. (2008). Qualitative data analysis: A compendium of techniques and a framework for selection for school psychology research and beyond. *School Psychology Quarterly*, 23, 587-604.
- Lehman, D.R., & Nisbett, R.E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology*, 26, 952-960.
- Lilienfeld, S.O. (1998). Pseudoscience in contemporary clinical psychology: What it is and what we can do about it. *Clinical Psychologists*, 51(4), 3-9.
- Lilienfeld, S.O., Lynn, S.J., & Lohr, J.M. (2004). Science and pseudoscience in clinical psychology: Initial thoughts reflections and considerations. In S.O.Lilienfeld, S.J. Lynn & J.M. Lohr (Eds.), *Science and pseudoscience in Clinical Psychology* (pp.1-16). New York: Guilford.
- Lipman, M. (2003). *Thinking in education*. Cambridge: Cambridge University Press.
- Macpherson, R., & Stanovich, K.E. (2007). Cognitive ability, thinking dispositions and instructional set as predictors of critical thinking. *Learning and individual differences*, 17, 115-127.
- Makgoba, M.W. (2000). HIV/AIDS: The peril of pseudoscience. *Science*, 288, 1171.
- Makgoba, M.W. (2002). Politics, the media and science in HIV/AIDS: The peril of pseudoscience. *Vaccine*, 20, 1899-1904.
- McKay, T. (2000). *Reasons, explanations and decisions: Guidelines for critical thinking*. Wadsworth, Canada.
- McKelvie, P., & Low, J. (2002). Listening to Mozart does not improve children's spatial ability: Final curtains for the Mozart effect. *British Journal of Developmental Psychology*, 20, 241-258.
- McPeck, H. (1981). *Critical thinking and education*. New York: St. Martin's Press.

- Meredith, F. (2009). Baby Einstein boasts prove not so smart for Disney. *Irish Times*. Retrieved January 12, 2010, from <http://www.irishtimes.com/newspaper/weekend/2009/1031/1224257741654.html>
- Michael, J.J., Devaney, R L., & Michael, W.B. (1980). The factorial validity of the Cornell Critical Thinking Test for a junior high school sample. *Education and Psychological Measurement*, 40, 437-450.
- Miri, B., David, B., & Uri, Z. (2007). Purposely teaching for the promotion of higher order thinking skills: A case of critical thinking. *Research Science Education*, 37, 353-369.
- Nosich, G.M. (2005). Problems with two standard models for teaching critical thinking. *New Directions for Community Colleges*, 150, 59-67.
- O'Donnel, S.L., Francis, A.L., & Mahurin, S.L. (2008). Critical thinking on contemporary issues. In D.S. Dunn, J.S. Halonen, & R.A. Smith (Eds.), *Teaching critical thinking in psychology: A handbook of best practices* (pp. 117-126). Sussex, UK: Wiley Blackwell.
- Osborne, R.E., Kriese, P., Tobey, H., & Johnson, E. (2009). Putting it all together: Incorporating SOTL practices for teaching interpersonal and critical thinking skills in an online course. *Insight: A Journal of Scholarly Teaching*, 4, 45-55.
- Paul, R. (1995). *Critical thinking: How to prepare students for a rapidly changing world*. Santa Rosa, California: Foundation for Critical Thinking.
- Paul, R. (2005). The state of critical thinking today. *New Directions for Community Colleges*, 130, 27-38.
- Paul, R., & Elder, L. (2005). *The thinkers guide to the nature and functions of critical and creative thought*. Dillon Beach, California: Foundation for Critical Thinking.
- Paul, R., Elder, L., & Bartell, T. (1997). *California teacher preparation for instruction in critical thinking: Research findings and policy recommendations*. Dillon Beach, California: The Foundation for Critical Thinking.

- Payne, S. (2007). Grounded Theory. In E. Lyons, & A. Coyle (Eds.), *Analysing qualitative data in psychology* (pp. 65-86). Thousand Oaks, California: Sage.
- Phillips, V., & Bond, C. (2004) Undergraduates' experiences of critical thinking. *Higher Education Research and Development*, 23, 277-294.
- Pinto, R.C. (1995). Post hoc ergo propter hoc. In H. V. Hansen, & R. C. Pinto (Eds.), *Fallacies: Classical and contemporary readings* (pp. 302-314). Penn State Press: University Park.
- Pirie, M. (2006). *How to win every argument: The use and abuse of logic*. London: MPG Books Ltd.
- Popper, K.R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. London: Routledge & Paul.
- Renaud, R. D., & Murray, H. G. (2008). A comparison of a subject-specific and a general measure of critical thinking. *Thinking Skills and Creativity* 3(2), 85–93.
- Resnick, L. (1987). *Education and learning to think*. Washington, DC: National Academy.
- Ruminski, H.J., & Hanks, W.E. (1995). Critical thinking lacks definition and uniform evaluation criteria. *Journalism and Mass Education Educator*, 50(3), 4-11.
- Rusco, J. (2000). Risky business: Vividness availability and the media paradox. *Skeptical Inquirer* 24(2), 22-26.
- Sadler, T.D., & Zeidler, D.L. (2005). Patterns of informal reasoning in the context of socioscientific decision making. *Journal of research in Science Teaching*, 42, 112-138.
- Saville, B.K., Zinn, T.E., Kerr Lawrence, N., Barron, K.E., & Andre, J. (2008). Teaching critical thinking in statistics and research methods. In D.S. Dunn, J.S. Halonen, & R.A. Smith (Eds.), *Teaching critical thinking in psychology: A handbook of best practices* (pp. 149-160). Sussex, UK: Wiley Blackwell.
- Shermer, M. (1997). *Why people believe weird things: Pseudoscience, superstition and other confusions of our time*. New York: Freeman Press.

- Siegel, H. (1998). *Education reason: Rationality, critical thinking, and education*. New York: Routledge, Chapman & Hall.
- Solon, T. (2006). Generic critical thinking infusion and course content learning in introductory psychology. *Journal of Instructional Psychology*, 34(2), 95-109.
- Standing, L.G., & Huber, H. (2003). Do psychology courses reduce belief in psychological myths? *Social Behaviour and Personality*, 31, 585-592.
- Stanovich, K.E. (1998). *How to think straight about psychology* (5<sup>th</sup>ed.). New York: Addison Wesley Longman.
- Stanovich, K E. (2004). *How to think straight about psychology*. Boston: Allyn and Bacon.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing Grounded Theory* (2<sup>nd</sup>ed.). London: Sage.
- Sumner, W.G. (1906). *Folkways a study of the Sociological importance of usages, manners, customs, mores, and morals*. New York: Dover.
- Taube, K.T. (1997). Critical thinking ability and disposition as factors of performance on a written critical thinking test. *Journal of General Education*, 46, 129-164.
- Tavris, C. (2003). The widening scientist-practitioner gap. In S.O. Lilienfeld, S.J. Lynn, & J.M. Lohr (Eds.), *Science and pseudoscience in clinical psychology*. New York: Guilford.
- Thompson, W.F., Schellenberg, E.G., & Husain, G. (2001). Arousal, mood and the Mozart effect. *Psychological Science*, 12, 245-251.
- Toplak, M.E., & Stanovich, K.E. (2003). Associations between myside bias on an informal reasoning task and amount of post-secondary education. *Applied Cognitive Psychology*, 17, 851-860.
- Tyler, M.A. (2006). *Critical spirit manifestations in TAFE teachers and their work*. Unpublished doctoral dissertation, University of Southern Queensland, Toowoomba, Australia.
- Van den Brink-Budgen, R. (2000). *Critical thinking for students*. Oxford: Cromwell Press.

- Van Gelder, T. (2000). *The efficacy of informal reasoning courses*. Retrieved April 10, 2009, from <http://www.philosophy.unimelb.edu.au/reason>.
- Van Gelder, T. (2001). *How to improve critical thinking using educational technology*. Retrieved April 14, 2009, from <http://www.ascilite.org.au/conferences/melbourne01/pdf/.../vangeldert.pdf>.
- Van Gelder, T. (2007). *Rationale: Making people smarter through argument mapping*. Retrieved April 14, 2009, from: <http://tillers.net/timvangelderdraftpaper.pdf>.
- Volkwyn, T.S. (2005). *First year students' understanding of measurement in physics laboratory work*. Unpublished master's thesis, University of Cape Town.
- Wade, C. (2008). Critical thinking: Needed now more than ever. In D.S. Dunn, J.S. Halonen, & R.A. Smith (Eds.), *Teaching critical thinking in psychology: A handbook of best practices* (pp. 11-22). Sussex, UK: Wiley Blackwell.
- Wagner, T.A., & Harvey, R.J. (2003). *Developing a new critical thinking test using item response theory*. Paper presented at the 2003 Annual Conference of the Society for Industrial and Organizational Psychology, Orlando.
- Walters, K.S. (1994). *Re-thinking reason: New perspectives in critical thinking*. Albany: State University of New York Press.
- Wang, S.Y., Tsai, J.C., Chiang, H C., Lai S.C., & Lin, H.J. (2008). Socrates, problem-based learning and critical thinking; a philosophic point of view. *Kaohsiung Journal of Medical Science*, 24(3), 6-13.
- Watson, G., & Glaser, E.M. (1980). *Watson–Glaser Critical Thinking Appraisal: Forms A and B*. San Antonio, TX: PsychCorp.
- West, R.F., & Stanovich, K.E. (2007). Natural myside bias is independent of cognitive ability. *Thinking and Reasoning*, 13, 225-247.



- West, R.F., Stanovich, K.E., & Toplak, M.E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Education Psychology, 100*, 930-941.
- Wild, C.J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*, 223-248.
- Willingham, D.T. (2008). Critical thinking: Why is it so hard to teach? *Arts Education Policy Review, 109*(4), 21-29.
- Wu, Y., & Tsai, C. (2007). High school students' informal reasoning on a Socio-scientific issue: Qualitative and quantitative analyses. *International Journal of Science Education, 29*, 1163-1187.
- Yanchar, S.C., Slife, B.D., & Warne, R. (2008). Critical thinking as disciplinary practice. *Review of General Psychology, 12*, 265-281.
- Yang, F.Y., & Anderson, O.R. (2003). Senior high school students' preference and reasoning modes about nuclear energy use. *Journal of Science Education, 25*, 221-244.
- Yeh, M. (2002). Assessing the reliability and validity of the Chinese version of the California Critical Thinking Disposition Inventory. *International Journal of Nursing Studies, 39*, 123-132.
- Yeh, S.S. (2001). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher, 30*(9), 12-17.
- Zeidler, D.L. (1997). The central role of fallacious thinking in science education. *Science Education, 81* 483-496.

## **APPENDIX A**

### Cognitive Skills Associated with Critical Thinking – The Delphi Report

#### 1) Interpretation,

Interpretation is the ability to ‘comprehend and express the meaning or significance of a wide variety of experiences, situations, data, events, judgments, conventions, beliefs, rules, procedures or criteria." Interpretation includes the sub-skills of categorization, decoding significance, and clarifying meaning.

#### (2) Analysis,

Involves the ability to ‘identify the intended and actual inferential relationships among statements, questions, concepts, descriptions or other forms of representation intended to express beliefs, judgments, experiences, reasons, information or opinions." Analysis includes the sub-skills of examining ideas, detecting arguments, and analyzing arguments into their component elements.

#### (3) Evaluation,

Evaluation is representative of the ability to, ‘assess the credibility of statements or other representations which are accounts or descriptions of a person's perception, experience, situation, judgment, belief or opinion; and to assess the logical strength of the actual or intended inferential relationships among statements, descriptions, questions, or other forms of representations."

Evaluation includes the sub-skills of assessing claims and assessing arguments.

(4) Inference,

Inference is the ability to, 'identify and secure elements needed to draw reasonable conclusions; to form conjectures and hypotheses, to consider relevant information and to educe the consequences flowing from data, statements, principles, evidence, judgments, beliefs, opinions, concepts, descriptions, questions, or other forms of representation." Inference includes the sub-skills of querying evidence, conjecturing alternatives, and drawing conclusions.

(5) Explanation,

Involves the ability to state the results of one's reasoning; to justify that reasoning in terms of the evidential, conceptual, methodological, criteriological and contextual considerations upon which one's results were based; and to present one's reasoning in the form of cogent arguments.

Explanation includes the sub-skills of stating results, justifying procedures, and presenting arguments.

6) Self-regulation,

Self-regulation is defined as "self-consciously to monitor one's cognitive activities, the elements used in those activities, and the results educed, particularly by applying skills in analysis and evaluation to one's own inferential judgments with a view toward questioning, confirming, validating, or correcting either one's reasoning or one's results." Self-regulation includes the sub-skills of self-examination and self-correction.

**APPENDIX B**Geertsen's (2003) List of Revised Micro-thinking SkillsHighest Degree of Abstraction

---

Extrapolation	to discover how what applies to known situations/data might apply to less familiar situations/data
Evaluation	to determine the value of something
Explanation	to deduce/show how something logically follows
Synthesis	to put together to form some coherent whole Analysis - to break down into various parts (compare and contrast)

---

(low transfer)

---

Application	to use with familiar situations/data
Classification	to arrange according to common characteristics
Interpretation	to give the essential meaning of something
Paraphrase	to restate the same thing in different words
Recall	to recollect or remember something

---

Lowest Degree of Abstraction

## APPENDIX C

### Vignettes Comprising the *Scientific Thinking in Psychology Scale*

Carefully read each of the scenarios presented below and answer the questions that follow. In many cases you may not have enough information or you may not have a strong preference for the choices that are given to you or perhaps the choice you really want is not an option. However, please choose the option that **most closely** fits your opinion. Please indicate your choice by making an X in the box next to your choice. In all cases please provide a detailed explanation of your choice.

1. Professor Rider announces to a group of workers that he wants to study the effect of playing music on their productivity. After a week of playing to them he finds that productivity has indeed increased. He then turns up the volume and after another week finds that the productivity has increased further. He thus concludes that music causes an increase in productivity. His colleague Professor Walker however, says this is not a correct conclusion.

With whom do you most strongly agree, Professor Rider or Professor Walker?

Professor Rider

Professor Walker

Please explain your choice.

2. Dr. Humperdinck conducts a study on a group of motor vehicle accident victims to test the effectiveness of a new treatment on reducing the symptoms of post-traumatic stress disorder

(PTSD) among them. His study results show that most patients who receive the new treatment report a reduction in their symptoms. Based on the results Dr. Humperdinck concludes that the treatment is effective in reducing symptoms of PTSD. Do you agree or disagree with Dr. Humperdinck's conclusion?

- Agree
- Disagree

Please explain your choice.

3. Dr. Johnson is treating an autistic child with play therapy and finds after a period of time that the child expresses himself more emotionally than at the beginning of therapy. He therefore recommends this approach to Dr. Fritz who has also recently started treating autistic children in his practice. Dr. Fritz conducts a review of the scientific articles about autism and concludes that there is no evidence that play therapy enhances emotional expression among autistic children. He decides not to use play therapy techniques with his patients. With whom do you more strongly agree, Dr. Johnson or Dr. Fritz?

- Dr. Johnson
- Dr. Fritz

Please explain your choice.

4. Professor Beano claims that the low self-esteem drives aggressive behaviour in children. His colleague, Professor Dandy, states that this statement is untrue. Professor Beano responds by challenging Professor Dandy to provide evidence that his statement is untrue.

Please indicate which of the following statements you agree with more strongly.

- It is Professor Beano's responsibility to present evidence to show that his statement is correct.
- It is Professor Dandy's responsibility to present evidence that Professor Beano's statement is incorrect.

Please explain your choice.

5. The following is stated in a report:

*"After video games became popular in the 1970s, national surveys have shown an increase in the incidence of aggressive behaviour among boys aged 10-16 in South Africa, the United States and the United Kingdom. These data show that watching video games causes aggressive behaviour among boys."*

Do you agree or disagree with the last sentence above?

- Agree
- Disagree

Please explain your choice.

6. Professor Claasen presents data showing that students who have high scores on a self esteem test also do well in psychology tests and examinations. He therefore concludes that self esteem causes academic success in psychology. Do you agree or disagree with Professor Claasen's conclusion?

- Agree
- Disagree

Please explain your choice.

7. Dr. Bailey specializes in treating patients with major depression. He observes that all his patients reported that their mothers also suffered from depression. He concludes that depression is inherited from the mother. Do you agree or disagree with Dr. Bailey's conclusion?

Agree

Disagree

Please explain your choice.

8. Psychologist Dr. Phillips treats patients who present with a variety of psychological disorders such as obsessive-compulsive disorder and bulimia nervosa. He uses his intuition to decide which techniques to use to treat his patients. His colleague, Dr. Frank, uses only techniques that are supported by the results of clinical trials.

With whose approach do you agree?

Dr. Phillips.

Dr. Frank.

Please explain your choice.

9. Psychologist Dr. Rice treats Susan in her private practice for several weeks. During this time Susan talks about her intense feelings of anger towards her mother and often mentions that she has thoughts of murdering her. One morning Dr. Rice reads a newspaper report that Susan has shot and killed her mother. Dr. Rice tells her colleague, Dr. Stew, that she had been convinced all along that



Susan would murder her mother. Dr. Stew, however, suggests that it was only **after** the murder occurred that Dr. Rice actually became convinced that Susan would kill her mother.

With whom do you agree, Dr. Rice or Dr. Stew?

Dr. Rice

Dr. Stew

Please explain your choice.

10. Over the period of a year, a clinic treats a total of 10 000 infants under two years of age.

Researchers at the clinic find that 2 500 of these infants under two are infected with HIV. They therefore conclude that 25% of infants under two in the area served by the clinic have HIV. The study is widely reported in the newspapers and causes lots of controversy with some experts saying that the figure is too high. With which group do you agree?

The researchers who conducted the study.

Critics who say the estimate is too high.

Please explain your choice.

11. Jack and Bob examine the results of a study that indicates that as the severity of depression increases, alcohol consumption also increases. Bob says “there are only two possible conclusions. Either depression causes alcohol consumption or alcohol consumption causes depression.” Jack says this is not true, as he can think of at least one other possibility. With whom do you agree?

Jack

Bob

Please explain your choice.

**APPENDIX D**

*Chi-square Table Representing the Relationship between the Frequency of MSR for Endorsement between the Control Group at Post-test and the Experimental Group at Pre-test*

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson chi-square	.019 <sup>a</sup>	1	.891		
Continuity correction <sup>b</sup>	.000	1	1.000		
Likelihood ratio	.019	1	.890		
Fisher's exact test				1.000	.569
Linear-by-linear Association	.019	1	.891		
N of valid cases	113				

*Notes:*

<sup>a</sup>0 cells (.0%) have expected count less than 5. The minimum expected count is 5.24.

<sup>b</sup>Computed only for a 2x2 table.

**APPENDIX E**

Summary of Code Descriptors, and Frequency of Occurrence, for the Music as the Catalyst Reasons for Endorsement at Pre- and Post-test – Experimental Group

Code descriptor	Code	Pre-test	Post-test
It clearly shows that music had a positive outcome on the workers' productivity	R/W10	10	6
If it is music that the workers like, they will feel happier and work better	R/W11	3	1
We do not know what music was played - type of music will be of importance in this study	R/W12	13	4
Turning up the volume means that it is the volume and not only the mere presence of music that has an effect	R/W13	4	8
When he changed the volume he actually changed the question – how does music volume influence productivity	W14	1	1
There can be more things than the music that had an influence, like the type of music	R/W15	6	1
It is already proven that music, e.g. Baroque music, stimulates the brain and helps productivity	R16	1	0
Music definitely played a role but there may be other factors	W17	2	2
Total		40(40%)	23(25%)

**APPENDIX F**

Summary of Code Descriptors, and Frequency of Occurrence, for the Type of Activities Reasons for Endorsement at Pre- and Post-test – Experimental Group

Code descriptor	Code	Pre-test	Post-test
It depends on the type of activities the workers were busy with	R/W20	7	0
Total		7 (7%)	0 (0%)

**APPENDIX G**

Summary of Code Descriptors, and Frequency of Occurrence, for the More Research Reasons for Endorsement at Pre- and Post-test – Experimental Group

Code descriptor	Code	Pre-test	Post-test
More research necessary	R/W30	3	3
More research – hypothetical considerations	R/W31	1	1
More research – sample concerns	W32	3	2
More research – more time needed	R/W33	2	1
Total		9(9%)	7(9%)

**APPENDIX H**

Summary of Code Descriptors, and Frequency of Occurrence, for the Endorsement Based on Personal Experience Reason for Endorsement at Pre- and Post-test – Experimental Group

Code descriptor	Code	Pre-test	Post-test
Endorsement based on personal experience	R/W40	11	1
Total		11 (11%)	1 (1%)

**APPENDIX I**

Summary of Code Descriptors, and Frequency of Occurrence, for the Individual Considerations

Reason for Endorsement at Pre- and Post-test – Experimental Group

Code descriptor	Code	Pre-test	Post-test
Not every person is able to produce under same the situations	R/W50	4	2
Total		4 (4%)	2 (2%)

**APPENDIX J**

Summary of Code Descriptors, and Frequency of Occurrence, for the Reliance on the Research Process Reasons for Endorsement at Pre- and Post-test – Experimental Group

Code descriptor	Code	Pre-test	Post-test
He made a decision to comply with the necessary resource to be able to get the result that his study aimed to achieve	R/W60	4	7
Prof. Walker does not give a reason why he says it does not increase productivity	R61	3	3
Prof. Rider has tested this hypothesis and with every test productivity increased	R62	2	1
Total		9 (9%)	11 (12%)



**APPENDIX K**

Summary of Code Descriptors, and Frequency of Occurrence, for the Confounding Variables

Reasons for Endorsement at Pre- and Post-test – Experimental Group

Code descriptor	Code	Pre-test	Post-test
There are other factors to consider	W70	7	7
He did not have a control group –there are other factors that could have contributed	W71	1	5
It is possible but are there other factors that could have increased productivity, e.g. Hawthorne effect	W72	0	1
The productivity may have increased because the workers may have started to bond with each other better	W73	1	0
One can't just assume that after such a short period in time the results can be directly related to the fact that music played	W74	0	6
Total		9 (9%)	19 (21%)

**APPENDIX L**

Summary of Code Descriptors, and Frequency of Occurrence, for the Identification of Potential Bias Reasons for Endorsement at Pre- and Post-test – Experimental Group

Code descriptor	Code	Pre-test	Post-test
Workers know they are being studied	R/W80	5	7
Workers know they are being studied – Rider should make use of a control group	W81	0	3
Hawthorne effect may be taking place	W82	0	3
Total		5 (5%)	13 (14%)

**APPENDIX M**

Summary of Code Descriptors, and Frequency of Occurrence, for the Methodological Concerns

Reasons for Endorsement at Pre- and Post-test – Experimental Group

Code descriptor	Code	Pre-test	Post-test
Methodological concerns – Generalized	W90	1	0
Methodological concerns –Counterfactual condition	W91	1	1
Methodological concerns – Sample considerations	W92	1	2
Methodological concerns – Control/comparison group	W93	1	2
Methodological concerns – Time to practice	W94	0	1
Methodological concerns – Construct/operational definition	W95	1	0
Methodological concerns – Pre-test	W96	0	2
Methodological concerns – Maturation	W97	0	1
Total		5 (5%)	9 (10%)

**APPENDIX N**

Summary of Code Descriptors, and Frequency of Occurrence, for the Reference to Technical Terms  
Reasons for Endorsement at Pre- and Post-test – Experimental Group

Code descriptor	Code	Pre-test	Post-test
More research to prove hypothesis correct	W30hy	-	2
Rider has proved his hypothesis correct	R60-61hy	1	3
There is a pre-post test so Rider can draw these conclusions	W100pr	-	1
Called the Hawthorne effect – workers work hard because they like the music	W100ha	-	1
Total		1 (1%)	7 (7%)