

# ON CONCEPTUAL ANALYSIS AS THE PRIMARY QUALITATIVE APPROACH TO STATISTICS EDUCATION RESEARCH IN PSYCHOLOGY

AGNES PETOCZ

*University of Western Sydney*  
*a.petocz@uws.edu.au*

GLENN NEWBERY

*University of Western Sydney*  
*g.newbery@uws.edu.au*

## ABSTRACT

*Statistics education in psychology often falls disappointingly short of its goals. The increasing use of qualitative approaches in statistics education research has extended and enriched our understanding of statistical cognition processes, and thus facilitated improvements in statistical education and practices. Yet conceptual analysis, a fundamental part of the scientific method and arguably the primary qualitative method insofar as it is logically prior and equally applicable to all other empirical research methods—quantitative, qualitative, and mixed—has been largely overlooked. In this paper we present the case for this approach, and then report results from a conceptual analysis of statistics education in psychology. The results highlight a number of major problems that have received little attention in standard statistics education research.*

**Keywords:** *Scientific method; Critical inquiry; Qualitative and quantitative research; Statistics for psychology*

## 1. INTRODUCTION: PSYCHOLOGY AND STATISTICS—A MARRIAGE MADE IN HELL

When Benjamin Franklin remarked that the only two certainties in life are death and taxes, little did he know that, two centuries later, the undergraduate psychology student aspiring to become a properly-trained practitioner would have to face a third, much more frightening, certainty: statistics. Indeed, the existential angst arising from awareness of our finitude pales into insignificance (in the colloquial sense) when compared with the mixture of panic and despair generated by the dreaded statistics and research methods courses. Despite reassurances that psychologists are not required to be experts in statistics (most of the hard work is done for us by sophisticated computer software packages, backed up, where necessary, by experts available for consultation), many students struggle both with the content and with the relevance of statistics. For them, the marriage of psychology and statistics can come to feel like a marriage made in hell.

As a result, statistics education in psychology remains a challenge for those on both sides of the teaching relation, and it is widely acknowledged that it often falls disappointingly short of its goals (Chiesi & Primi, 2010; Lalonde & Gardner, 1993). This is the case even for those enlightened teachers of research methods who have kept up with

pedagogical developments promoting student “engagement” and have joined the educational shift in focus from the merely “epistemological” (what students will *know*) to including the “ontological” (what students will *become*) (cf. Barnett, 2007; Petocz & Reid, 2010). One major problem is that teachers and students diverge on the specific ontological goal; they both want to focus on what students will become, but, whereas students want to become psychologists and help people, their teachers want them to become competent researchers and evaluators of the psychological research of others, so as to become competent scientist-practitioners.

Teaching in psychology follows the North American *scientist-practitioner model* (established at the Boulder conference in 1949, cf. Belar & Perry, 1991). Psychology today is defined as the science of human behaviour and mental life, and is distinguished from pop psychology and its ubiquitous extensions and applications by the latter’s pseudo-scientific grounding. The discipline’s success as an applied field rests on the putative scientific status of the research findings from which are derived its various practical applications (psychological tests, measurement scales, types of clinical intervention, etc.). Accordingly, students are taught to be scientists first, and then practitioners; their practice must be scientifically-based; it must implement only those interventions that are evidence-based and derived from sound theory and rigorous empirical research. And the hallmarks of rigorous empirical research are held in psychology to be experimentation, measurement, and quantitative statistical data analysis. The use of statistics in psychology arose concurrently with the use of measurement and quantitative techniques. The latter became consolidated into the post-second world war methodological consensus (Michell, 2002, 2010), which was fuelled by the introduction of government funding (and subsequent monitoring) of scientific research (Leahy, 2004; Solovey, 2004), in the context of a rapidly expanding identification of quantification with objectivity (Porter, 1995). Since then, quantitative methods have become entrenched within mainstream psychology, and almost universally regarded as constitutive of science.

Given the challenging situation of trying to teach something that is considered essential in science to students who struggle with its content and relevance, it is not surprising that research in this area is burgeoning. Until recently most of this research has used quantitative methods. Nevertheless, because statistics education continues to need all the help it can get, it is understandable—even encouraging—that there has been a move towards qualitative and mixed-methods approaches to statistics education research.

## **2. THE RISE OF QUALITATIVE APPROACHES TO STATISTICS EDUCATION RESEARCH**

The recent move towards qualitative and mixed methods approaches to statistics education research has occurred in the context of a general reinvigoration of qualitative methods within the social sciences, described as a “quiet methodological revolution” (Denzin & Lincoln, 1994, p. ix). Qualitative methods draw upon a variety of broad philosophical approaches identified as post-Kuhnian “paradigms” (Lincoln & Guba, 2000) or “axiomatic positions” (Potter, 1996), such as social constructionism, phenomenology, symbolic interactionism, hermeneutics, structuralism, post-structuralism, and ethogenics (Richardson, 1996). The specific techniques or methods (grounded theory, discourse analysis, ethnography, phenomenography, participant observation, protocol analysis, case study, narrative analysis, etc.) are typically concerned with “the meaning of experience, actions and events as these are interpreted through the eyes of particular participants, researchers and (sub)cultures” (Henwood, 1996, p. 27), and their aim is to

“do more justice to the objects of research than is possible in quantitative research” (Flick, 2002, p. 8; cf. also, Janesick, 2000).

There are two good reasons for statistics education researchers to adopt qualitative (and mixed) methods. First, as Petocz & Reid (2010) remind us, “while statistics is essentially a quantitative discipline, it contains a necessary core of qualitative components” (p. 272), such as definitions, categories, and forms of linguistic structure (cf. Bowker & Star, 2000; Petocz & Sowey, 2010). Secondly, qualitative methods can allow educators to discover and address students’ conceptions of statistics. Petocz and Reid argue that, despite some pressure to use statistics to investigate statistics education, qualitative approaches can give us something that cannot be acquired from quantitative approaches. For instance, they can clarify the variety of ways in which statistics cognition processes can go awry. Kalinowski, Lai, Fidler, and Cumming (this issue) point out that, where quantitative methods may reveal, say, frequencies of false beliefs about some statistical concept, qualitative methods can illuminate how students have arrived at those false beliefs, helping us to access the reasoning processes and assumptions at work in the formation of misconceptions.

These views are justified. There is increasing evidence that the use of qualitative approaches to statistics education research has indeed extended and enriched our understanding of statistical cognition processes, and thus facilitated improvements in statistics education and practices. For example, Kalinowski et al. (this issue) report a series of studies that succeeded in capturing subtleties of statistical cognitive reasoning processes concerning conceptions (and misconceptions) of the difference between using confidence intervals (CIs) and null hypothesis significance tests (NHST). In one case, the results suggest that the seemingly successful statistical reform in medicine (in comparison to psychology), as evidenced by the increased reporting of CIs, is merely superficial in that the correct understanding and interpreting of CIs lags far behind the reporting of them. In another study, qualitative data allowed the researchers to test and confirm their hypotheses about specific misconceptions, and then to develop more effective methods of teaching the relevant statistical principles. Petocz and Reid (2010) review a series of their own studies that used phenomenography, a form of grounded theory furnishing descriptions of cognition processes from the learner’s perspective, derived from semi-structured interview data, and obtained via a Piagetian style of answer-contingent questioning. Synthesising the results of their research, they identify hierarchically structured qualitative variability in students’ understanding of statistical issues, from “narrow” conceptions to “broader” conceptions. Within that context, they identify six qualitatively distinct conceptions both of statistics and of learning statistics, ranging from the most limiting (mere techniques) to the most expansive (helping to make sense of the world and foster personal transformation). These findings inform their various recommendations concerning the pedagogical conditions in which student motivation can be fostered and movements towards the more expansive end of the continuum can occur. The recognition of the centrality of motivation is also found in the work of Bijker, Wynants and van Buuren (2006), who showed the value of an integrated teaching and learning design for fostering optimal motivation, more favourable attitudes and more adequate learning strategies in studying statistics (cf. also van Buuren, 2006; Wiberg, 2009).

More recently, Gal and Ograjenšek (2010) have extended the core message concerning the value of qualitative approaches by reporting results from a number of other studies (qualitative and mixed), such as naturalistic observational studies accompanied by follow-up structured interviews, protocol analysis of thinking-aloud-while-solving-problems, ethnographic participant observational studies, focus groups

combined with in-depth interviews, and content analysis of data derived from CIT (Critical Incident Technique) studies. They offer, amongst other conclusions, three recommendations. First, “qualitative thinking and knowledge of relevant qualitative methods should be considered part of the conceptual and methodological preparation of statisticians” (p. 293). Secondly, “students should be made aware of the existence of qualitative techniques during the introduction of the standard research process (e.g. problem formulation and research planning) at the beginning of a statistics course” (p. 293). Thirdly, “a solid exposure not only to advantages but also to limitations of quantitative methods should be provided” (p. 294).

We agree wholeheartedly with these three recommendations, and we find the general direction of qualitative approaches to be very encouraging, as far as it goes. But it needs to go much further. For there is one qualitative approach that, we feel, has been largely overlooked. This is the method of conceptual analysis, whose special place as primary within scientific methods in general has also been overlooked. As we hope to show, this is the only method that allows us to identify and tackle a number of issues in statistics education that have crucial theoretical and practical implications, but which are left unexamined by other methods, whether qualitative, quantitative or mixed.

### **3. CONCEPTUAL ANALYSIS AND ITS NEGLECT IN PSYCHOLOGY**

Conceptual analysis is analysis of concepts, terms, variables, constructs, definitions, assertions, hypotheses, and theories. It involves examining these for clarity and coherence, critically scrutinising their logical relations, and identifying assumptions and implications. Sometimes called *theoretical research*, and closely related to *critical thinking*, conceptual analysis is not merely a matter of language or language use (cf. Bennett & Hacker, 2003); it is also a matter of the *content* of our linguistic expressions, that is, what we claim to be thinking and talking *about*. And sometimes conceptual analysis serves to expose (typically unconscious) *practical inconsistency* (Bassham, Irwin, Nardone, & Wallace, 2008), such as when someone rejects logic by means of a valid deductive argument (cf. Triplett, 1988), or behaves like a realist in their research while explicitly claiming allegiance to antirealist perspectives (cf. Lambie, 1991; Levitt, 2001).

In a recent paper entitled “Toward a richer view of the scientific method: The role of conceptual analysis,” Machado and Silva (2007) point to the neglect of conceptual analysis in psychology, as evidenced by the general lack of space and attention given to it in teaching courses and curricula, textbooks, conference presentations, published papers, and book chapters. Scientific method, they say, has always included three discernible subsets or clusters of activity: *experimentation* (performing controlled experiments, systematic observations and correlational studies); *mathematisation* (framing mathematical or statistical laws and models on the basis of data collected via experimentation); and *conceptual analysis* (clarifying concepts, exposing conceptual problems in models, revealing unacknowledged assumptions and steps in arguments, evaluating the consistency of theoretical accounts). Psychology, however, regards only the first two of these three activities as central to science, and so operates with an impoverished view of scientific method. Moreover, psychologists continue to devalue conceptual analysis on various (ironically, inconsistent) grounds: that it is “obvious” and an inherent part of scientific activity; that it is a form of verbal tyranny, stifling scientific creativity; and that it is purely negative and destructive, failing to offer constructive avenues of progress.

To counter these misconceptions, Machado and Silva (2007) illustrate the intertwining of all three activities within the work of the quintessential scientist of the scientific revolution, Galileo. They then show the applicability and value of conceptual analysis to a number of research problems across various areas in contemporary psychology. For example, conceptual analysis solves the age-old “problem” of the inversion of retinal images, by clarifying that it rests on the unacknowledged incoherent assumption that the *object* of perception is the retinal image: “The problem of how objects are seen upright if their retinal images are inverted is not to be solved empirically but to be *dissolved conceptually*” (p. 675, emphasis added). Bennett and Hacker (2003) provide similar conceptual analyses of this and many other “problems” in current experimental cognitive neuroscience. Machado and Silva conclude that, because “Observations and experiments, on the one hand, and conceptual analysis, on the other hand, are filters through which all scientific hypotheses, models and theories must pass” (p. 679), it follows that “psychologists should replace the currently dominant view of the scientific method with one that assigns conceptual analysis its proper weight” (p. 680).

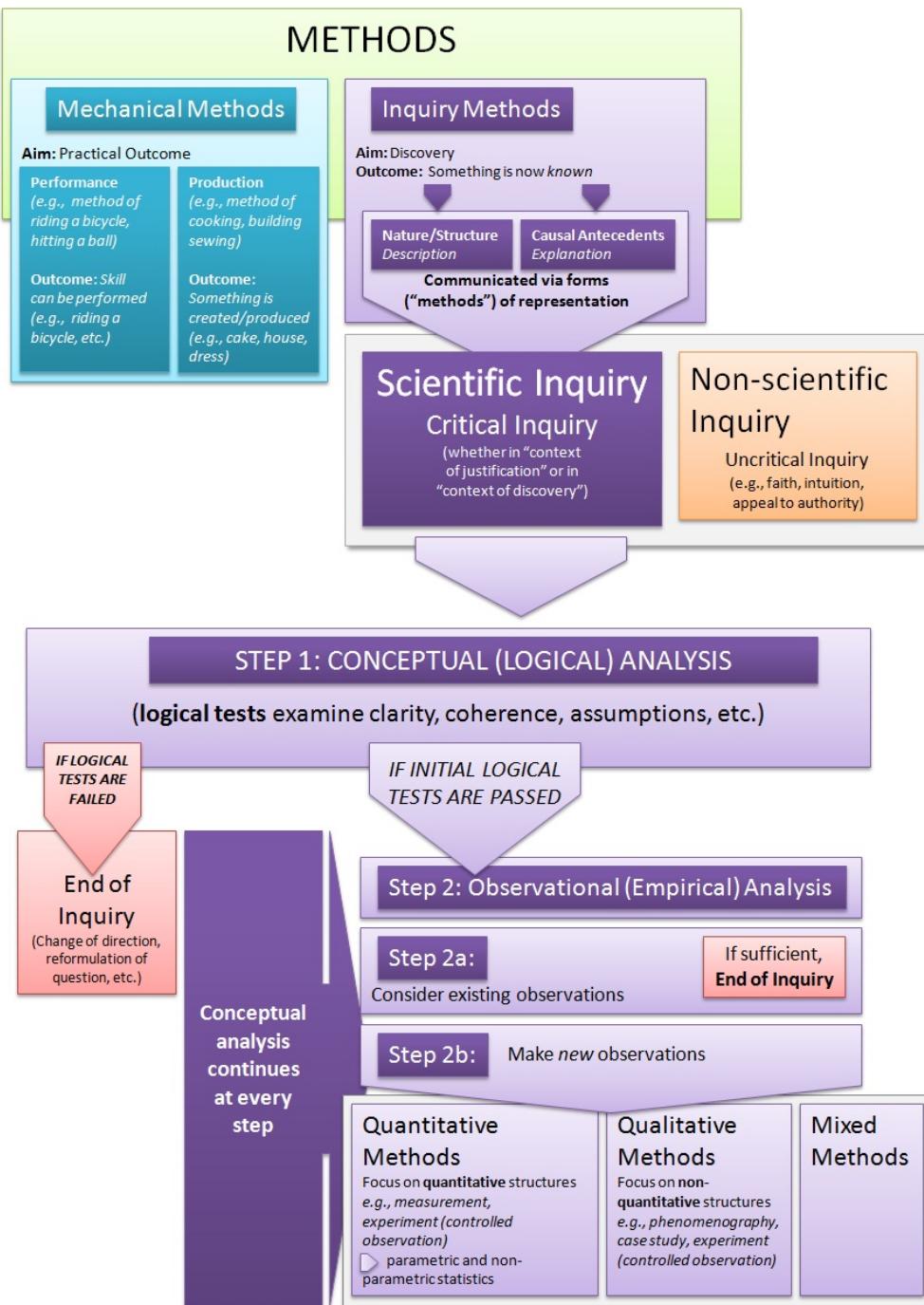
Here, too, we wholeheartedly agree. But again we feel that there is a stronger case to be made. This case emerges from a closer examination of scientific method. As we shall attempt to show, conceptual analysis, far from sitting alongside experimentation and mathematisation as one of three separate activities constituting science, is logically and temporally *prior to* the other two, is responsible on occasion for *precluding* the other two when it reveals them to be scientifically inappropriate, and is indispensable *during* the other two. Therefore, conceptual analysis is, in these three senses, the primary method.

#### **4. LOCATING CONCEPTUAL ANALYSIS WITHIN SCIENTIFIC METHOD AS CRITICAL INQUIRY**

In order to locate conceptual analysis clearly within scientific method, we begin with four points that are generally accepted within the scientific community, including mainstream psychology: (a) Science is a human social activity aimed at the investigation of natural systems (including, of course, human and psychological systems); (b) These systems are comprised (as is all reality) of complex, dynamic, interrelated and interacting spatio-temporal situations; (c) These situations *can* (in principle) be known by humans (or any other cognising organisms), and many of them *are* known (although, of course, given their infinite complexity, we cannot know *everything* about a situation); (d) Our cognitive and perceptual apparatus is limited and fallible; we can, and do, make mistakes.

These four facts underpin the broad conception of science that was introduced by the ancient Greeks and developed throughout the European tradition (cf. Crombie, 1994; Haack, 2003; Machado & Silva, 2007). This is the view of science as *critical inquiry*; that is, critical inquiry is the *core* method of science. This view contrasts with mainstream psychology’s much narrower view of science, one which (as will become clear from the analysis that follows) not only excludes conceptual analysis but which, as a consequence of that exclusion, consists of a scientific package of distortions and misconceptions, involving implicit lapses from explicit commitments to science and realism (Bickhard, 1992; Green, 1992; Mackay & Petocz, in press; Petocz, 2004, in press).

So, where exactly does conceptual analysis fit within scientific method as critical inquiry? This can be clarified via a hierarchical classification of methods, which includes the stages of scientific inquiry (see Figure 1 for a diagrammatic representation of the themes discussed here).



*Figure 1. Classification of methods showing the primary role of conceptual analysis in the stages of scientific inquiry*

Our initial division of methods is into two basic categories: *mechanical* methods and *inquiry* methods. Mechanical methods involve doing something to achieve a particular outcome, and they can be further divided into *performance* methods (e.g., how to ride a

bike, how to hit a tennis forehand) and *production* methods (e.g., recipes for cooking, procedures for building, techniques for sewing). The result of a successful performance method is that we can do something that we could not previously do (e.g., we can hit a tennis forehand). The result of a successful production method is that something which did not exist before exists now (e.g., we now have a cake, a house, a dress).

In contrast to mechanical methods, the fundamental aim of methods of inquiry is *discovery*. The result of a successful inquiry method is that something which was previously not known is now known. The something that is discovered may be the *nature* or *structure* of some situation (its properties, processes, etc.), such that we have discovered the answer to the “what?” or “how?” (including the “how does it work?”) questions, and we can therefore provide a *description* of that situation. Alternatively, the something that is discovered may be the situation’s *causal antecedents*, the necessary and sufficient conditions for the situation’s occurrence, such that we have discovered the answer to the “why?” question, and we can therefore provide an *explanation* of the situation. Hence, the goals of description and explanation are made possible by achieving the inquiry method’s aim of discovery. These findings are communicated by various forms (sometimes called “methods”) of representation.

Now, leaving aside mechanical methods and remaining with methods of inquiry, these can be divided into two basic categories: *scientific* and *non-scientific*. The difference between them is that scientific inquiry is *critical* inquiry in that it is premised on the recognition of our cognitive and perceptual fallibility. Because we might be wrong, we direct our efforts to applying the best available error-detection mechanisms, so that, if possible, we can identify and correct our errors. Science “bends over backwards” to use procedures that can be self-corrective. This attitude is what unites the various specific forms of scientific investigation. To repeat, then, *critical inquiry is the single core scientific method* (cf. Cohen & Nagel, 1934; Haack, 2003; Michell, 2010), whether it be in the “context of discovery” or in the “context of justification” (Herschel, 1987). In contrast, non-scientific inquiry is not critical or self-critical; it does not focus on the possibility of being wrong, and does not welcome criticism. Faith, intuition, appeal to authority—these may well lead to knowledge, but they are not concerned with challenging their own knowledge claims and subjecting them to critical testing.

Insofar as it rests on critical inquiry, scientific inquiry overlaps with ordinary everyday inquiry. In that respect, scientific method “does not differ in its nature from the normal activity of thought” (Freud, 1933/1975, p. 170), and is “merely a potentiation of common sense, exercised with a specially firm determination not to persist in error if any exertion of hand or mind can deliver us from it” (Medawar, 1969, p. 59). As such, it draws upon no special concepts of evidence or mode of inference:

Respect for evidence, care in weighing it, and persistence in seeking it out, so far from being exclusively scientific desiderata, are the standards by which we judge *all* inquirers, detectives, historians, investigative journalists, etc., as well as scientists ... Scientific inquiry is continuous with the most ordinary of everyday empirical inquiry. There is no mode of inference, no ‘scientific method,’ exclusive to the sciences and guaranteed to produce true, probably true, more nearly true, or more empirically adequate, results. (Haack, 2003, pp. 23–24)

The critical inquiry that is the core method of science can be further divided into two different methods, depending on its focus. These are *conceptual (or logical) analysis* and *observational (or empirical) analysis*. When we pose questions about the nature and ways of working of the systems that we are investigating, and when we offer answers to those questions, critical inquiry requires that we subject these to *testing*; that is, we consider whether we could be wrong.

In conceptual analysis, we are conducting theoretical research and applying logical tests. We examine logical structures, including assumptions and implications, and apply tests of clarity, intelligibility, coherence, and so on to our concepts, questions, hypotheses, and theories. This shows the *first* sense in which conceptual analysis is *primary*; it is the first step in scientific inquiry (Step 1 in Figure 1); it is that part of scientific research that must be applied *prior* to the choice of any specific observational method. For, if we are faced with a supposed field of inquiry or a research question, only conceptual analysis will establish whether or not our question is clear, logically coherent, and so on, and thus admits of further investigation. This is what is meant by saying that the *testing* of a scientific theory involves first subjecting it to *logical* tests. If the logical tests are failed, if our conceptual analysis reveals confusions, ambiguities, contradictions, implicit assumptions, and so forth, then we know without going any further (i.e., without taking the next step into any specific observational analysis) that the situation as envisaged is either *not yet clear enough* or *could not possibly be the case*. We are then constrained, in accordance with the requirements of scientific investigation, to reconsider the question, reformulate it, clarify it, adjust it, or abandon it. This reveals the *second* sense in which conceptual analysis is primary; it has the power to *preclude* observational inquiry, whereas observational analysis can never reveal that conceptual analysis is inappropriate.

If the results of conceptual analysis show that the initial logical tests *are* passed, then the next step (Step 2) is to move to the other form of scientific inquiry, observational (empirical) analysis. The first step in this venture (Step 2a) is obviously to consider *existing* observations, and, if those existing observations are judged to be sufficient, the inquiry can end there. In other words, from a scientific point of view (in contrast, perhaps, to a social or political point of view), there is no point in wasting time and resources making observations that have already been satisfactorily made, so it is incumbent upon scientific researchers to be familiar with research that has already been successfully conducted. If, however, existing observations are insufficient, or there are no existing observations, then the next step is to make new observations. These new observations are further categorised into the familiar “research methods”: *quantitative* methods, *qualitative* (i.e., *non-quantitative*) methods, and *mixed* methods.

And here we come to the *third* sense in which conceptual analysis is primary. Even if we have established via conceptual analysis that our theory or research question is logically clear and coherent in all relevant aspects, we do not then leave conceptual analysis behind as we move into the various empirical observational methods. Conceptual analysis continues to be required all along the way, up to and including the final interpretation of our results and suggestions for future research directions. For example, logic is needed in order to derive testable hypotheses from the theory via the standard hypothetico-deductive method. Then it proceeds via recognising the *logical* point that quantity and quality are not interchangeable, and that the grounds for choice between quantitative or qualitative methods is an empirical issue. That is, from a realist perspective (and most scientific psychologists claim to be realists) scientific observational inquiry requires attunement of the specific method or technique to the nature of what is being investigated or focused on. Therefore, the choice of quantitative or qualitative method must be determined not by a priori ideological commitment (e.g., “I subscribe to social constructionism, so I do qualitative research,” or “I am a scientist, so I do quantitative research”), nor by *imposing* one type of structure onto another (e.g., “If I apply numbers to this material, that will render the material quantitative”), but by the nature of the material under investigation. If we are investigating structures or aspects of structures that are quantitative in nature (whether discrete or continuous), then quantitative methods will be appropriate, whereas if we are investigating structures or aspects of structures that are

*not* quantitative in nature, then qualitative methods are scientifically demanded and quantitative methods will be scientifically *inappropriate*. And if we are interested in both quantitative and qualitative features, then mixed methods will be warranted. Next, once the specific observational method has been appropriately determined, conceptual analysis is required to identify appropriate methods of control, to determine the mode of data analysis, and to interpret the results in the light of the underlying assumptions of that analysis. For example, the value of *experimentation* (as opposed to other observational methods) will depend not only on how much control of variables is possible but also on how much control is appropriate under the circumstances (as is indicated by the justifiable concern in psychology with *ecological validity*). Similarly, the value of analysing data via parametric or non-parametric statistics is directly related to logical considerations of the fit between the nature of the collected data and the assumptions and implications of the statistical analysis. In these ways, conceptual analysis remains an essential accompaniment to observational (empirical) analysis, and indispensable if the conclusions drawn from the research findings are to be sound.

Against this background of locating conceptual analysis as primary within scientific method, we can now highlight four important implications that are uncomfortably at odds with current views and practices in mainstream psychology, and which are relevant to our subsequent scrutiny of statistics education.

## **5. IMPLICATIONS FOR PSYCHOLOGY OF THE PLACE OF CONCEPTUAL ANALYSIS AS PRIMARY WITHIN SCIENTIFIC METHOD**

To begin with, scientific psychology subscribes explicitly to realism. However, mistakenly believing that positivism and realism are equivalent, it has accepted positivism's version of the distinction between *logical* and *empirical* inquiry, according to which only empirical inquiry can be regarded as scientific, and logical inquiry is to be relegated to philosophy. Hence, psychology came to regard logical or conceptual analysis as part of the philosophy package that needed to be dismissed in the name of science. But here (as in other crucial respects) positivism is antithetical to realism (cf. Friedman, 1991, 1999; Hibberd, 2005). Thus, contrary to the received view in mainstream psychology, neither experimentation nor measurement nor the use of statistics (let alone all three combined) is necessary for science. Nor does scientific research begin with the quantitative or qualitative methods adopted to make new observations. Instead, according to realism, scientific research begins with conceptual analysis, and what *is* necessary is “the persistent application of logic as the common feature of all reasoned knowledge ... in essence scientific method is simply the pursuit of truth as determined by logical considerations” (Cohen & Nagel, 1934, p. 192).

Secondly, because the realism to which scientific psychology is explicitly committed entails that it is a question of the way things are in the world whether any particular attribute is quantitative or qualitative in structure, it follows that qualitative methods are just as scientifically legitimate (indeed, warranted) as are quantitative methods. Nature is full of non-quantitative structures (e.g., semantic, categorical, logical), all of which are appropriately investigated by non-quantitative methods, typically included under the umbrella term *qualitative*. Qualitative methods are scientific when they are appropriately applied and applied carefully and critically. Conversely, quantitative methods are *unscientific* when they are *misapplied*. Here, again, psychology has been misled by the antirealist aspects of positivism. The positivist conventionalism of mainstream psychological practices adopts the view that “the difference between quality and quantity is not a difference in nature, but a difference in our conceptual schema—in our language”

(Carnap, 1966, p. 58). But, again from the point of view of the realism to which psychology explicitly subscribes, it is a point of logic, discovered via conceptual analysis, that it is *not* possible that “data collection procedures are independent of data analysis techniques … [such that] collected data may be *transformed* at any point, and may be analysed both quantitatively and qualitatively” (Tashakkori & Teddlie, 2003a, p. 696, emphasis added). In other words, it is *not* possible to “quantitize” qualitative data or “qualitize” quantitative data (Teddlie & Tashakkori, 2003, p. 9).

This leads into the third implication. To believe that we can *make* something quantitative merely by throwing numbers at it is to conflate mechanical methods with inquiry methods; it is to treat a mechanical method (e.g., applying a recipe) as if it were an inquiry method. Though it is true that mechanical methods rest on facts about the world that have been discovered, and though it is true that in pursuing inquiry one may make use of mechanical procedures (operating a microscope, making calculations via a computer, following the “recipe” for performing a particular statistical analysis, etc.), these methods of performance and production cannot *replace* methods of inquiry. To think that they can is to engage in a form of scientific *practicalism*, which begins with subordinating the (realist) search for truth to the goal of achieving specific practical (performance or production) outcomes, and ends with believing that the essence of science lies in simply adopting quantitative methods and performing statistical analysis.

Finally, insofar as conceptual analysis is a method that focuses on *non*-quantitative structures, namely, logical, semantic and categorical structures, it is clearly not a quantitative method. As such, it belongs under the umbrella term *qualitative* on the same grounds as are all other so-called qualitative methods. When this is combined with the primacy of conceptual analysis within scientific research, according to which sometimes the inquiry will end without going on to the other specific observational methods, where at other times conceptual analysis will be needed alongside those other methods, it can be concluded that conceptual analysis is the *primary qualitative method* in scientific research.

## **6. APPLYING CONCEPTUAL ANALYSIS TO STATISTICS EDUCATION IN PSYCHOLOGY**

Previous qualitative research has focused on students’ perceptions of statistics and its relevance, on their statistical cognition processes, and, in particular, on their conceptions and misconceptions of various statistical issues. Our conceptual analysis of statistics education goes deeper; it addresses the *bases from which* we derive our evaluative criteria regarding students’ conceptions and misconceptions, and for this we turn the spotlight back onto ourselves as teachers, and our own understandings of statistical issues. We also adopt the scientist-practitioner model; we examine first the scientific *content* and then the various *practices*. As we hope to show, conceptual analysis highlights a number of major problems that have “flown under the radar” in standard statistics education research. With respect to scientific content, these problems include unacknowledged and conflicting assumptions about the nature of quantity, quality, measurement, probability, evidence, and inference. Many of these assumptions are controversial, unwarranted, likely to be false (on the basis of available evidence), or known to be false. With respect to practice, the problems include the existence of long-entrenched practices that are at odds with psychology’s stated scientific goals. That is, from a conceptual analysis perspective they are forms of practical inconsistency. Furthermore, these practices preclude access to the tools necessary for statistics education reform.

## 6.1. CONCEPTUAL ANALYSIS OF STATISTICS EDUCATION CONTENT

With respect to the *content* of statistics, we can divide it for purposes of discussion into two separate clusters of unexamined and conflicting assumptions. The first cluster concerns the nature of measurement, quantity, quality, the distinction between quantitative and qualitative methods, test scores, and test validity. The second cluster concerns the nature of probability, evidence, and the evidential role of statistical inference.

Taking the first cluster, there are three different uses of the term *measurement*. In loose colloquial language it can mean “assess,” “examine,” “observe” or even “describe.” Many psychology students have difficulty in letting go of this usage (e.g., “How can we ‘measure’ the mind?”). But, in science, the concept of measurement is more specific and technical. The physical sciences (which psychology has long attempted to emulate) follow the standard, classical realist definition of measurement, according to which measurement is “*the estimation or discovery of the ratio of some magnitude of a quantitative attribute to a unit of the same attribute*” (Michell, 1997, p. 358, emphasis in original). That is, in science, measurement is the assessment of (continuous) quantitative structure and so *requires* and *presupposes* quantitative structure. If the variable that we are considering is not by nature quantitative, then measurement is impossible; attempting measurement would be like trying to drink a loaf of bread or discover the width of one’s coughing; bread just doesn’t have drinkable properties and coughs just don’t come in widths. However, as *measure* is not coextensive with “describe,” “observe,” “investigate” or “assess,” we can still legitimately do any of these other things to our variable, and so it does not follow that, if we cannot measure it, we must ignore it (cf. Petocz & Sowey, 2010, p. 61); we can explore what kind of *non-quantitative* structure it has, and investigate it via whatever qualitative methods are appropriate.

Now, mainstream psychology’s treatment of measurement conflicts with this scientific view, and with psychology’s own explicit commitment to scientific realism. This theme has been most thoroughly and cogently explored in the work of Michell over the last two decades (e.g., Michell, 1990, 1999, 2010, and especially 1997, which is his seminal paper in this area and which includes peer responses and discussion). To summarise this work, psychology’s commitment to the *quantitative imperative* (the false belief that measurement is a necessary condition of science—see Michell, 2003) has led to a psychometric approach that constitutes a “pathology of science” (Michell, 2000, 2008), a form of “methodological thought-disorder” (Michell, 1997). This involves a two-stage breakdown in scientific practice. At the first stage, the underlying empirical hypothesis that the psychological attribute being assessed is quantitative is not subject to empirical test. In other words, psychologists merely *assume* that the variables of interest (abilities, intelligence, attitudes, personality traits, etc.) are indeed quantitative and hence *measurable*. At the second stage, this failure to test the empirical hypothesis is then *disguised*. The disguise is effected by psychology’s abandoning the classical realist definition of measurement and adopting its own special definition of measurement, following S. S. Stevens (Stevens, 1946; cf. Michell, 2002), according to which measurement is the “*assignment of numerals to objects or events according to rule*” (Michell, 1997, p. 360, emphasis in original). This nonrealist, operationist definition of measurement allows psychologists to claim to be measuring any variable of interest, regardless of whether or not it actually has a quantitative structure; they think they can *make* something quantitative simply by throwing numerals at it. And this is how measurement is (mis)taught in psychology (Michell, 2001).

The inference from order to quantity, which Michell (2009a) labels the *psychometricians' fallacy*, and the underlying assumption of quantitativity, lead consequentially to standard treatments of psychological test scores. In response to the objection that his arguments do not apply to probabilistic item response models, Michell (2004a) points out that the same untested assumption underlies them, insofar as the issue of whether the relevant psychological attribute is quantitative is *never* raised as a source of model misfit. Moreover, "the locus of quantitativity in probabilistic models derives not from anything known directly about the relevant attribute but from the postulated shape of 'error'" (pp. 125–6), with the result that "if the attribute is not quantitative, the supposed shape of 'error' only projects the image of a fictitious quantitativity" (p. 126). In general, then, psychology appears to be dealing with ordered attributes whose degrees of difference are intrinsically nonmetric because they are qualitatively heterogeneous (Michell, 2010). Therefore, "transforming" test scores (whether into factor scores via factor analysis or into ability scores via item response models) does not transform them into measures of anything when there is no evidence that the hypothesised factors or abilities are quantitative.

The same problem extends to psychology's concept of test validity, which was introduced nearly a century ago, and which is typically understood and taught to students as being the extent to which a test measures what it is supposed to measure. Being a crucial concept in psychology by virtue of the widespread use of tests and scales, and there being nothing analogous for measurement in the physical sciences (where the structure of the attribute is identified *before* the measuring instrument is devised), validity has been the subject of much discussion and many attempts at unification (cf. Lissitz, 2009). But here, again, as Michell (2009b) points out, the definition of validity rests on two *untested* assumptions: that the test measures *something*, and that whatever it is supposed to be measuring *can be measured*. However, "Few testers realize that knowledge of the structure of attributes is necessary to provide a scientific base for testing practice" (p. 123). As "no scientific evidence has been collected capable of sustaining the hypothesis that such attributes are quantitative" (Michell, 2010, p. 48), it follows that "the presumptions underlying the concept of validity are invalidly endorsed" (Michell, 2009b, p. 111).

Taking the second cluster, the problems of assumed quantitativity and illegitimate measurement are exacerbated by the application of statistical techniques that rest on confused assumptions concerning the nature of probability and the evidential role of statistical inference. Psychologists are generally unaware that there are competing views on the nature of probability, and that those who developed the statistical techniques now widely applied in psychology advised against their use and debated the implications of their disagreements. Psychologists are generally unaware that what is taught in their textbooks is the result of an "inference revolution" (Gigerenzer & Murray, 1987) that occurred in the middle of the twentieth century and "saw the widespread adoption of a hybrid of frequently contradictory Fisherian and Neyman-Pearson statistical concepts as a single, uncontroversial method of inductive inference" (Hubbard & Ryan, 2000, p. 666). Specifically, this hybrid cobbled together Fisherian NHST with the competing views of Neyman and Pearson, together with their ideas about an alternative hypothesis, Type 1 and Type 2 errors, and statistical power, in a manner that neither camp would have condoned. As Hubbard and Ryan summarise it, "Despite serious misgivings on both sides, this hybrid methodology, with its emphasis on the rejection of the null hypothesis at  $p < .05$  or better, gradually made its appearance in behavioral science textbooks on statistical methods" (p. 666). Since then, "four decades of increasingly trenchant criticism over the

use and abuse of SST [statistical significance testing] has failed to reduce its visibility in the psychological literature” (p. 672).

Others have gone deeper in their analyses. For example, Grayson (1998) argues convincingly that *the* major statistical confusion in psychology (and other social sciences) is the blurring of the distinction between the Neyman frequentist view of long-run probability (which disallows the evidential use of probabilistic information in the single case) and the *scientific issue of evidential inference*. According to Neyman and Pearson (1933), in terms of the nature of scientific evidence, there is nothing to be gained in a *single* piece of research by performing a statistical analysis and computing a *single* p-value; that is, there is no point at all to how we routinely perform statistical analyses following single experiments. Nevertheless, psychologists are unwilling to give up the objective (yet scientifically sterile) long-run frequency view of probability in exchange for one that seems too subjective, insofar as it treats probabilities as numbers that represent *degrees of belief* that can be quantified on single occasions (as in the Bayesian approach) (cf. also Gigerenzer, 1987). Yet the attraction of the Bayesian approach lies in its allowing for evidential inference in a single experiment. So, psychologists want to have their cake and eat it. Following Gigerenzer’s (1993) Freudian analogy, Grayson (1998) agrees that, in mainstream psychology’s statistical psyche, the superego is represented by rigorous, socially-desirable frequentism, the ego is the confused yet pragmatic Fisherian significance testing, and the id is the Bayesian lustful desire for evidential inference, allowing us to draw conclusions about the probability of our hypotheses given the data. Furthermore, Grayson demonstrates that the unavoidable issue of evidential inference in a single case is left unaddressed by the increasingly popular recommendations to avoid the problem by moving to CIs, effect sizes, meta-analyses, and the like (even if these do bring other genuine advantages). The current assumption is that replacing NHST with CIs, properly understood, constitutes statistical reform and progress, partly because it challenges the dichotomous thinking associated with the accept/reject decision rule in NHST. However, if the arguments presented by Gigerenzer and Grayson are correct, then this is a band-aid remedy for a life-threatening wound.

Psychology’s statistical psyche is populated by numerous other confusions and misconceptions (cf. also Grayson, 1988; Rosnow & Rosenthal, 1989; Rozeboom, 1960), such as those concerning “negative” results, against the reporting of which there is widespread scientific prejudice (Furedy, 1978; Rosenthal, 1979); the use of “protection” methods in multiple-inference procedures (Grayson & Oliphant, 1996); the misunderstanding of post-hoc testing (Oliphant & Grayson, 1996); and the widespread practice of *HARKing* (hypothesising after the results are known) by adjusting predictions in the light of results (Kerr, 1998).

Grayson’s (1998) conclusions can be generalised to all of these: The issues typically are *scientific*, not *statistical*; each of us has the responsibility of understanding and coming to a (possibly uncertain) point of view on them; the evaluation of statistical evidence is only one small part of “the whole scientific enterprise of reasoning about and evaluation of evidence” (p. 342); and, finally, “if nothing else comes from the present debate, it will be immensely valuable if it destroys the sanguine reliance on ‘statistical experts’ that has characterized the attitudes of many social scientists in the past” (p. 341).

## **6.2. CONCEPTUAL ANALYSIS OF STATISTICS EDUCATION PRACTICE**

From a conceptual analysis point of view, the nature of our reliance on “experts” in statistics education is part of a general package that can best be described as a case of practical inconsistency. Of course, practical inconsistency within psychology is not

restricted to the teaching of statistics. For example, it appears also in the discrepancy within *qualitative* methods between what is explicitly preached (antirealist relativism, including the rejection of the “procrustean” quantitative “psychometric trinity” of validity, generalisability, and reliability) and what is actually practised (implicit methodological realism, adherence to qualitative counterparts of the psychometric trinity—viz. credibility, transferability, and dependability) (Lincoln & Guba, 2000). Any brief survey of the various handbooks of qualitative or mixed research methods (e.g., Denzin & Lincoln, 2000; Flick, 2002; Potter, 1996; Richardson, 1996; Tashakkori & Teddlie, 2003b) will confirm the existence of this discrepancy (Hammersley, 1989; but see Bryman, 1988, for a comprehensive and perceptive discussion of the issues). This has contributed to the reluctance of scientific psychologists pursuing quantitative methods to embrace qualitative methods as being of equal scientific merit and to integrate them appropriately in their research methods courses. Within quantitative psychology in general and statistics education in particular a similar (albeit converse and more disguised) discrepancy exists. Here, the philosophy that is explicitly preached is scientific realism, whereas the methodology that is practised is an antirealist and unscientific form of scientific practicalism.

Scientific practicalism includes an instrumentalist view of science as directed towards practical ends, a concern with “what works.” In terms of our earlier classification of methods, practicalism blurs the distinction between methods of inquiry and mechanical methods. In statistics education in psychology, we pay lip service to the scientist-practitioner model and to the goal of developing students into critical consumers and conductors of psychological research. For example, the Australian Psychology Accreditation Council (APAC), the professional body that accredits psychology teaching programs within universities, bases its required standards on the scientist-practitioner model. Amongst these standards, critical thinking skills are emphasised (APAC, 2010; Tanner & Danielson, 2007). However, we do not follow this in practice. In addition to neglecting conceptual analysis in the ways we have discussed, we typically train and reinforce students to treat statistical and data-analytic methods as a set of practical recipes, the blind implementation of which ipso facto guarantees the scientific credentials of the researchers and the scientific value of any results.

Moreover, unfortunately, we become role models for our students. We do as we are told—by journal editors, psychological associations, accreditation bodies, and so on. We allow these masters to dictate the fashions that we follow, and make the decisions for us. For example, as Schmidt (1992) complains, “it has been remarkably difficult to wean social science researchers away from their entrancement with significance testing.... The psychology of addiction to significance testing would be a fascinating research area” (p. 1176). Ironically, the terms “wean” and “addiction” are appropriate, for they imply an infantile state of dependence. Accordingly, a task force was created in 1996 by the American Psychological Association (APA) Board of Scientific Affairs to determine whether to proscribe the use of NHST and replace it with CIs, that is, to “phase out” NHST from course texts and journal articles, thus constraining teachers to change their courses, authors to revise their textbooks, and journal article authors to modify their inference strategies (Hubbard, Parsa, & Luthy, 1997; Kirk, 1996). In the event, the task force (quite reasonably) refused to ban NHST or p-values. Nevertheless, the pressures continue.

For example, in a recent paper “Challenges for quantitative psychology and measurement in the 21st century,” Osbourne (2010) notes,

At this, the dawn of the 21st century, there are remarkably promising signs. Researchers are beginning to understand that strict null hypothesis statistical testing

(NHST) is limiting and provides an incomplete picture of results. More *journals now require* effect sizes, confidence intervals, and other practices one might argue are well overdue. (p. 2, emphasis added)

He urges that only by challenging assumptions can we be assured of following the path of “best practice”:

The world doesn’t need another journal promulgating 20th century thinking, genuflecting at the altar of  $p < 0.05$ . I challenge us to challenge tradition. Shrug off the shackles of 20th century methodology and thinking, and the next time you sit down to examine your hard-earned data, challenge yourself to implement *one new methodology* that represents best practice. Use Rasch measurement or IRT rather than averaging items to form scale scores. Calculate  $p_{(rep)}$  in addition to power and  $p$ . Use HLM to study change over time, or use propensity scores to create more sound comparison groups. Use meta-analysis to leverage the findings of dozens of studies rather than merely adding one more to the literature. Choose just one best practice, and use it. (p. 3)

This advice is proffered in the context of the author’s warning that “To blindly accept the dogma of the field without scholarly examination is to diminish what we do,” so “We must be vigilant, as researchers … to continue questioning and examining our tacit assumptions” (p. 1).

Here, practical inconsistency is enshrined in the message. Conceptual analysis does indeed require us to test our assumptions. But *which* assumptions? In this enthusiastic rallying of the quantitative troops, there is no awareness of the widespread failure to recognise, let alone test, the *basic* assumption underlying psychological measurement in general. Nor is there any apparent awareness that the “promising signs” at the dawn of the 21st century were already presented and discussed nearly a century earlier, sometimes even by those who invented the techniques. There is no awareness of the invalidity at the heart of psychology’s concept of psychometric test validity. There is no acknowledgment that the altars of effect sizes, confidence intervals, Rasch measurements, and the like are no more deserving of genuflection than that of  $p < 0.05$ .

Psychology’s failure to engage in conceptual analysis “all the way down” means that in the future it may become just as hard to “wean” researchers from their addiction to meta-analysis and confidence intervals, or from any other statistical technique considered at the time to be “best practice.” Those who *are* aware of the problems nevertheless continue to succumb to the pressure to engage in a flawed practice; it guarantees publication, continued employment, professional development, and the respect of the scientific psychological community. According to Hubbard and Ryan (2000), “This alone underscores the degree to which SST has attained the status of a methodological imprimatur that even the staunchest critics of its use must occasionally bow down before” (p. 673).

Hence, the replacement of inquiry methods with mechanical methods leads to the *courting of error*, in direct contrast to the aim of the core scientific method of critical inquiry to address the possibility of error in order to reduce it. The teaching of statistics, just like the teaching of psychometrics, “actually subverts the scientific method” (Michell, 2001, p. 211).

The teaching of statistics in psychology also subverts our explicit ethical aims. As Furedy (1978) points out in his discussion of the prejudice against “negative” or “null” results, “the effect of such scientific prejudice on the progress of knowledge can be … every bit as devastating as the effects of racial prejudice on the life of a political community” (p. 169). For example, it can lead to “the birth of seemingly indestructible empirical generalizations which actually have little or no evidential basis” (p. 177). The

result is unconscious double standards in our ethics. Fudging data is universally condemned; allowing our psychological “measures” to determine social or educational policies is not.

Part of the problem lies in the fact that the various contributions in the theory and history of statistics, including the controversial issues concerning measurement, probability, evidential inference, and so on, are omitted from our statistics and research methods textbooks, making it even more difficult to include them in the curriculum and to examine them critically. Psychologists in general are simply not exposed to the debates and criticisms relating to their own statistical practices (cf. Halpin & Stam, 2006). Of course, many of them may be thankful for this. But it precludes access to the tools necessary for implementing reform in the teaching of statistics.

## 7. RECOMMENDATIONS

In the light of the results of our conceptual analysis, we offer three recommendations for statistics education in psychology.

First, statistics education should include a much broader context, both conceptual and socio-historical. As indicated in our classification of methods, statistics must first be appropriately contextualised within scientific method in general, and specific empirical research methods (quantitative, qualitative, and mixed) in particular (cf. Bryman, 1988). It must also encompass not only the underlying core concepts but also some of the conditions under which particular concepts and practices (including the use of various statistical techniques) were adopted and others rejected, and their different implications. It might be objected that this is an unrealistic counsel of perfection, for we simply do not have the time or space to cover all the (admittedly useful) contextual factors. Just as it has been argued that students cannot reasonably be expected to absorb and understand the complex mathematical calculations that are involved in statistics, so it might equally be objected that they cannot reasonably be expected to understand the difficult conceptual content involved in the various controversies regarding the nature of measurement, probability, evidential inference, and so on. In any case, assumptions are always involved in any field of inquiry, and one cannot be expected to commit oneself to a potentially infinite regress of questioning and examining assumptions. Our reply to these objections is that *any* awareness is better than none, and that all assumptions are not equal. Some assumptions are warranted and uncontroversial, others are not; some are relevant, others are not. Where an assumption is controversial, relevant, and appears to be unwarranted, a scientific investigator cannot afford *not* to examine it. Just as Socrates claimed that, from an ethical point of view, an unexamined life is not worth leading, so, too, from a scientific point of view, unexamined assumptions are not worth holding, and unexamined practices not worth pursuing.

Our second recommendation flows from the first and from our discussion of scientific method as critical inquiry. Appropriate contextualising of statistics within scientific method means that neither quantitative methods nor statistics can be taught as having a privileged place within science or scientific research methods. As scientific methods must be attuned to the nature of the phenomena being investigated, and because the phenomena of interest will vary from one scientific field to another (e.g., physics deals more obviously with quantitative structures), one cannot legislate a priori on the relevance and value of quantitative methods. Clearly, as Michell (2004b) points out, “the fixation upon quantitative methods that characterizes modern psychology really has no justification given the realist understanding of science” (p. 307). Therefore, it is *antiscientific* for psychology to continue to insist that quantitative methods are somehow more scientific

than are qualitative methods. Instead, nature contains a potentially infinite number of possible alternatives to quantitative structure, and so the mainstream psychological researcher, no less than the student of psychology, should be prepared to explore novel empirical structures and appropriate ways of investigating them, including mathematically (Michell, 2001; cf. Resnick, 1997). It might be objected that, contrary to this new message, quantitative and statistical methods just *are* accorded a privileged place, and it is useless to protest. However, social privilege is not logical privilege. To conflate the sociology with the logic of science is to join postmodernism and scientific irrationalism and to retreat from the realist self-understanding of scientific psychology (Haack, 1996; Lambie, 1991; Stove, 1991). Where statistics and quantitative methods have been mistakenly allocated a privileged logical place, students will come to appreciate that this has arisen not from the requirements of science per se but from social, political, and other forces.

Our third recommendation is that the current prescriptive teaching and research practices, being unjustified, need to be replaced by practices more in keeping with the spirit of the core scientific method of critical inquiry. Even if we cannot prevent the prescriptions that appear in textbooks and are issued by task forces, professional associations, accreditation bodies, and the like, we can pursue the scientific job of treating them critically. We can follow Osbourne's (2010) call to challenge assumptions by challenging also the implicit assumptions on which *his* recommendations rest. Prescription in statistics education substitutes mechanical methods for methods of inquiry. Prescriptions are appropriate for cooking recipes and building guidelines (although, even in those cases, they are always conditional—if you want to produce a cake that rises, *then* you must use self-raising flour or baking powder). It might be objected that this would be too disillusioning for students, shaking their confidence in psychological science. However, as the term suggests, disillusioning students involves challenging and dispelling their illusions. Surely no honest scientist knowingly aims to instil or reinforce illusions, including the illusion that there is a royal road to scientific certainty that comes from doing as one is told. In any case, this move would be consistent with the “ontological” goal of education, helping students to gain the self-confidence that comes from becoming authentic. In our experience, students are quite resilient enough, and even enjoy iconoclasm and its wake of creative opportunities, especially if they perceive that their teachers are companions on the same journey.

It should be clear that we see our recommendations not as competing with some of the pedagogical conclusions that are emerging from recent qualitative approaches to statistics education research but as both supporting and supplementing them. Conceptual analysis reveals that the field of statistics is not as rigid and exact as it is typically portrayed, and its variability and uncertainty cannot be magically spirited away. Our recommendation to teach statistics within its broader context reinforces the three recommendations made by Gal and Ograjenšek (2010), and also sits comfortably alongside calls (e.g., Petocz & Reid, 2010) to broaden the learning context to include the student as person and as professional, and consider the value and meaning of statistics for the person and their life. We believe that, in a very profound way, exposure to the uncertainty, variability, and inconsistencies in assumptions about the *content* and *use* of statistics will mirror and validate the qualitative variability in the meanings of statistics for individuals. Furthermore, using conceptual analysis as a qualitative approach to statistics education research will result in promoting conceptual analysis as a learner strategy, and, thereby, contributing to the ontological goals of education, namely the transformation of the person into an autonomous thinker.

## 8. SOME CONCLUDING REMARKS

We are acutely aware that we have insufficient space here to detail and substantiate these claims and explore the underlying issues as fully as they warrant (that would be a task for our several statistics and research methods courses). It would be nice if we could, via some quick-and-easy summary, save the reader the time and effort required by the challenging task of examining all of the primary source material (e.g., the work of Michell) from which we have built our case. But, even if that were possible, it would contravene what we are proposing—that psychologists and their students should be in a position to examine the material carefully and critically for themselves, before coming to their own conclusions. Having said that, we hope our sketch has been sufficient to make the main point. The results of our conceptual analysis of statistics education in psychology suggest that, if we wish to do more than merely pay lip service to contemporary educational goals, we shall need to supplement our qualitative investigations into our students' conceptions and misconceptions of statistics by turning the research spotlight back onto ourselves, our own understandings, and our teaching practices.

We have restricted our discussion to statistics education in *psychology*, and leave it to others to consider whether or not aspects of our analysis might extend beyond psychology into other disciplines. Perhaps psychology is a special case; perhaps only in psychology is there “confusion and barrenness” that are attributable to the combination of “experimental methods and *conceptual confusion*” (Wittgenstein, 1953, p. 232); perhaps only in psychology does the “powerful impact of disciplinary socialization practices” (Good, 2007, p. 286) constitute an almost insurmountable obstacle to scientific progress; perhaps only in psychology does the aversion to examining assumptions mean that “a great deal of psychological research might well rest on philosophical quicksand” (Green, 1992, p. 292); perhaps only in psychology does it happen that “Galloping empiricism, which is our present occupational disease, dashes forth like a headless horseman” (Allport, 1966, p. 3), such that “We find ourselves confused by our intemperate empiricism, which often yields unnameable factors, arbitrary codes, unintelligible interaction effects, and sheer flatulence from our computers” (p. 8).

If we have given the impression that statistics education in psychology is sometimes a case of the unaware leading the naïve into error, then we make no apologies. For, on any thorough and dispassionate scrutiny, the evidence is overwhelming. Mainstream psychology’s obsession with being “scientific” has led, ironically, via its marriage-of-convenience with statistics and quantitative methods and its devaluing of qualitative approaches (including conceptual analysis) to an institutionalised disguising of the problems and misconceptions in its psychometric, data-analytic, and statistics education practices.

But we cannot honestly claim to be surprised by this. As psychologists, we are all too aware of the fact that the mind is not a unity but a society of competing impulses, and that genuinely scientific concerns can so easily be displaced or subverted by social, political, and economic interests, especially if the latter are indeed well-served by the current practices of the profit-driven academic institutions (Nussbaum, 2010). It is therefore quite understandable that so much of mainstream psychology is engaged in what Chomsky (1959/1964) identified as “a kind of play acting at science” (p. 559).

Nevertheless, we are cautiously optimistic. For, as Bennett and Hacker (2003) point out in their exhaustive critique of contemporary cognitive neuroscience, “Conceptual entanglement *can* coexist with flourishing science.... Hidden reefs do not imply that the reefs are not navigable, only that they are dangerous. The moot question is how running

on these reefs is manifest" (p. 5). Our positive message emerges from three of the many ironies in this story. First, conceptual analysis of statistics education in psychology produces evidence that one of the major problems has been the neglect of conceptual analysis within the research methods teaching programs. Because critical inquiry is the core method of science, conceptual analysis could perhaps be "re-marketed" by harnessing the motivations of psychologists; it could be gently reintroduced as a way of reinforcing and enriching psychology's scientific goal and its commitments to evidence-based practice (Machado & Silva, 2007). Next, if it is the case, as the evidence suggests, that methodological thinking in psychology has gone astray over the last 60 years (Toomela & Valsiner, 2010), it may turn out that, in bringing it back on track, we come to agree that the marriage between psychology and statistics was indeed made in hell and is best off heading for the divorce court—albeit not for the student-perception reasons we considered at the outset; rather, in terms of the subject matter of the two disciplines, the relationship is more appropriately that of acquaintances who are occasional fruitful collaborators than that of intimate long-term bedfellows. Finally, conceptual analysis reveals an additional way in which statistics education can move beyond the mere "epistemological" focus and achieve the deepest possible form of "ontological" goal. Here, the self-transformation of its students (no less than its teachers) will come in the shape of the familiar Socratic irony: The greater the field of our awareness and understanding, the more we can transform from mute followers of rigid authoritarian prescriptions into authentic and willing embracers of aporia and enlightened ignorance.

## REFERENCES

- Allport, G. (1966). Traits revisited. *American Psychologist*, 21(1), 1–10.
- Australian Psychology Accreditation Council (2010). Rules for accreditation and accreditation standards for psychology courses.  
 [Online:[http://www.apac.psychology.org.au/Assets/Files/APAC\\_Rules\\_for%20\\_Accreditation\\_and\\_Accreditation\\_Standards\\_for%20\\_Psychology\\_Courses\\_Ver\\_10\\_June\\_2010.pdf](http://www.apac.psychology.org.au/Assets/Files/APAC_Rules_for%20_Accreditation_and_Accreditation_Standards_for%20_Psychology_Courses_Ver_10_June_2010.pdf) ]
- Barnett, R. (2007). *A will to learn: Being a student in an age of uncertainty*. Buckingham, UK: Society for research in Higher Education, Open University Press.
- Bassham, G., Irwin, W., Nardone, H., & Wallace, J. M. (2008). *Critical thinking: A student's introduction* (3<sup>rd</sup> ed.). Boston: McGraw Hill.
- Belar, C. D., & Perry, N. W. (Eds.). (1991). *Proceedings: National conference on scientist-practitioner education*. Sarasota, FL: Professional Resource Exchange.
- Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Oxford: Blackwell.
- Bickhard, M. H. (1992). Myths of science. *Theory & Psychology*, 2(3), 321–337.
- Bijker, M., Wynants, G., & van Buuren, H. (2006). A comparative study of the effects of motivational and attitudinal factors on studying statistics. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.  
 [Online: [http://www.stat.auckland.ac.nz/~iase/publications/17/5H2\\_BIJK.pdf](http://www.stat.auckland.ac.nz/~iase/publications/17/5H2_BIJK.pdf) ]
- Bowker, G., & Star, S. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: The MIT Press.
- Bryman, A. (1988). *Quantity and quality in social research*. London: Unwin Hyman.
- Carnap, R. (1966). *Philosophical foundations of physics*. New York: Basic Books.
- Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students' statistics achievement. *Statistics Education Research Journal*, 9(1), 6–26.  
 [Online: [http://www.stat.auckland.ac.nz/~iase/sjerj/SERJ9%281%29\\_Chiesi\\_Primi.pdf](http://www.stat.auckland.ac.nz/~iase/sjerj/SERJ9%281%29_Chiesi_Primi.pdf) ]

- Chomsky, N. (1964). A review of B. F. Skinner's *Verbal behavior*. In J. A. Fodor & J. J. Katz (Eds.), *The structure of language: Readings in the philosophy of language* (pp. 547–578). Englewood Cliffs, NJ: Prentice-Hall. (Reprinted from *Language*, 1959, 35, 26–58)
- Cohen, M. R., & Nagel, E. (1934). *An introduction to logic and scientific method*. London: Routledge & Kegan Paul.
- Crombie, A. C. (1994). *Styles of scientific thinking in the European tradition*. London: Duckworth.
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (2000). *Handbook of qualitative research* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.
- Flick, U. (2002). *An introduction to qualitative research* (2<sup>nd</sup> ed.). London: Sage.
- Freud, S. (1975). Lecture XXXV. The question of a *Weltanschauung*. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 22, pp. 158–182). London: Hogarth Press. (Original work published 1933)
- Friedman, M. (1991). The re-evaluation of logical positivism. *Journal of Philosophy*, 88(10), 505–19.
- Friedman, M. (1999). *Reconsidering logical positivism*. Cambridge, MA: Cambridge University Press.
- Furedy, J. J. (1978). Negative results: Abolish the name but honour the same. In J. P. Sutcliffe (Ed.), *Conceptual analysis and method in psychology. Essays in honour of W. M. O'Neil*. Sydney: Sydney University Press.
- Gal, I., & Ograjenšek, I. (2010). Qualitative research in the service of understanding learners and users of statistics. *International Statistical Review*, 78(2), 287–296.
- Gigerenzer, G. (1987). Probabilistic thinking and the fight against subjectivity. In L. Krüger, G. Gigerenzer, & M. Morgan (Eds.), *The probabilistic revolution* (Vol. 2, pp. 11–36). Cambridge, MA: MIT Press.
- Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 311–339). Hillsdale, NJ: Erlbaum.
- Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Lawrence Erlbaum.
- Good, J. M. M. (2007). The affordances for social psychology of the ecological approach to social knowing. *Theory & Psychology*, 17(2), 265–295.
- Grayson, D. A. (1988). Limitation on the use of scales in psychiatry. *Australian & New Zealand Journal of Psychiatry*, 22(1), 99–108.
- Grayson, D. A. (1998). The frequentist façade and the flight from evidential inference. *British Journal of Psychology*, 83(2), 325–45.
- Grayson, D. A., & Oliphant, G. W. (1996). Multiple inference procedures: Problems and solutions. In C. R. Latimer & J. Michell (Eds.), *At once scientific and philosophic: A Festschrift for John Philip Sutcliffe* (pp. 23–46). Brisbane: Boombana Publications.
- Green, C. D. (1992). Of immortal mythological beasts. *Theory & Psychology*, 2(3), 291–320.
- Haack, S. (1996). Towards a sober sociology of science. In P. R. Gross (Ed.), *The flight from science and reason* (pp. 259–65). New York: Annals of the New York Academy of Sciences.
- Haack, S. (2003). *Defending science—within reason: Between scientism and cynicism*. New York: Prometheus Books.

- Halpin, P. F., & Stam, H. J. (2006). Inductive inference or inductive behavior: Fisher and Neyman-Pearson approaches to statistical testing in psychological research. *American Journal of Psychology*, 119(4), 625–653.
- Hammersley, M. (1989). *The dilemma of qualitative method: Herbert Blumer and the Chicago tradition*. London: Routledge.
- Henwood, K. (1996). Qualitative inquiry: perspectives, methods and psychology. In J. T. E. Richardson (Ed.), *Handbook of qualitative research methods* (pp. 25–40). Oxford: BPS Blackwell.
- Herschel, J. F. W. (1987). *A preliminary discourse on the study of natural philosophy*. Chicago: University of Chicago Press.
- Hibberd, F. J. (2005). *Unfolding social constructionism*. New York: Springer.
- Hubbard, R., Parsa, R. A., & Luthy, M. R. (1997). The spread of statistical significance testing in psychology. The case of the *Journal of Applied Psychology*, 1917–1994. *Theory & Psychology*, 7(4), 545–554.
- Hubbard, R., & Ryan, P. (2000). The historical growth of statistical significance testing in psychology—and its future prospects. *Educational and Psychological Measurement*, 60(5), 661–681.
- Janesick, V. J. (2000). The choreography of qualitative research design. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2<sup>nd</sup> ed.) (pp. 379–399). Thousand Oaks: Sage.
- Kalinowski, P., Lai, J., Fidler, F., & Cumming, G. (2010). Qualitative research: An essential part of statistical cognition. *Statistics Education Research Journal*, 9(2), 22–34.  
[Online: [http://www.stat.auckland.ac.nz/~iase/serj/SERJ9%282%29\\_Kalinowski.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ9%282%29_Kalinowski.pdf) ]
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196–217.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746–759.
- Lalonde, R. N., & Gardner, R. C. (1993). Statistics as a second language? A model for predicting performance in psychology students. *Canadian Journal of Behavioural Science*, 25(1), 108–125.
- Lambie, J. (1991). The misuse of Kuhn in psychology. *The Psychologist: Bulletin of the British Psychological Society*, 1, 6–11.
- Leahey, T. H. (2004). *A history of psychology. Main currents in psychological thought* (6<sup>th</sup> ed.). New Jersey: Pearson Education.
- Levitt, N. (2001). The sources and dangers of postmodern anti-science: Do these intellectual popguns matter? *Free Inquiry*, 21(2), 44–47.
- Lincoln, Y. S., & Guba, E. G. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2<sup>nd</sup> ed., pp. 163–188). Thousand Oaks, CA: Sage.
- Lissitz, R. W. (Ed.). (2009). *The concept of validity. Revisions, new directions, and applications*. Charlotte, NC: Information Age Publishing Inc.
- Machado, A., & Silva, F. J. (2007). Toward a richer view of the scientific method. The role of conceptual analysis. *American Psychologist*, 62(7), 671–681.
- Mackay, N., & Petocz, A. (Eds.). (in press). *Realism and psychology: Collected essays*. Leiden: Brill.
- Medawar, P. B. (1969). *Induction and intuition in scientific thought*. London: Methuen & Co Ltd.
- Michell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale, NJ: Erlbaum.

- Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88(3), 355–83.
- Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge: Cambridge University Press.
- Michell, J. (2000). Normal science, pathological science and psychometrics. *Theory & Psychology*, 10(5), 639–667.
- Michell, J. (2001). Teaching and misteaching measurement in psychology. *Australian Psychologist*, 36(3), 211–217.
- Michell, J. (2002). Stevens's theory of scales of measurement and its place in modern psychology. *Australian Journal of Psychology*, 54(2), 99–104.
- Michell, J. (2003). The quantitative imperative: Positivism, naïve realism and the place of qualitative methods in Psychology. *Theory & Psychology*, 13(1), 5–31.
- Michell, J. (2004a). Item response models, pathological science and the shape of error: Reply to Boorsboom and Mellenbergh. *Theory & Psychology*, 14(1), 121–129.
- Michell, J. (2004b). The place of qualitative research in psychology. *Qualitative Research in Psychology*, 1(4), 307–319.
- Michell, J. (2008). Is psychometrics pathological science? *Measurement: Interdisciplinary Research and Perspectives*, 6(1 & 2), 7–24.
- Michell, J. (2009a). The psychometricians' fallacy: Too clever by half? *British Journal of Mathematical and Statistical Psychology*, 62(1), 41–55.
- Michell, J. (2009b). Invalidity in validity. In R. W. Lissitz (Ed.), *The concept of validity. Revisions, new directions, and applications* (pp. 111–133). Charlotte, NC: Information Age Publishing Inc.
- Michell, J. (2010). The quantity/quality interchange. A blind spot on the highway of science. In A. Toomela & J. Valsiner (Eds.), *Methodological thinking in psychology: 60 years gone astray?* (pp. 45–68). Charlotte, NC: Information Age Publishing Inc.
- Neyman, J., & Pearson, E. S. (1933). On the problem of most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London*, A, 231, 289–337.
- Nussbaum, M. (2010, April 30). Skills for life. Why cuts in humanities teaching pose a threat to democracy itself. *Times Literary Supplement*, pp. 13–15.
- Oliphant, G. W., & Grayson, D. A. (1996). Some misconceptions concerning post hoc statistical hypothesis testing. In C. R. Latimer & J. Michell (Eds.), *At once scientific and philosophic: A Festschrift for John Philip Sutcliffe* (pp. 47–56). Brisbane: Boombana Publications.
- Osborne, J. W. (2010). Challenges for quantitative psychology and measurement in the 21st century. *Frontiers in Psychology*, 1(1), 1–3.
- Petocz, A. (2004). Science, meaning and the scientist-practitioner model of treatment. In D. Jones (Ed.), *Working with dangerous people* (pp. 25–42). Oxford: Radcliffe Medical Press.
- Petocz, A. (in press). Re-thinking the place of semiotics in psychology and its implications for psychological research. In S. C. Hamel (Ed.), *Semiotics: Theory and applications*. New York: Nova Science.
- Petocz, P., & Reid, A. (2010). On becoming a statistician: A qualitative view. *International Statistical Review*, 78(2), 271–286.
- Petocz, P., & Sowey, E. (2010). Statistical diversions. *Teaching Statistics*, 32(2), 60–64.
- Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton: Princeton University Press.
- Potter, W. J. (1996). *An analysis of thinking and research about qualitative methods*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Resnick, M. D. (1997). *Mathematics as a science of patterns*. Oxford: Oxford University Press.
- Richardson, J. T. E. (1996). *Handbook of qualitative research methods*. Oxford: BPS Blackwell.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641.
- Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, 44(10), 1276–84.
- Rozeboom, W. (1960). The fallacy of the null hypothesis significance test. *Psychological Bulletin*, 57(5), 416–428.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47(10), 1173–1181.
- Solovey, M. (2004). Riding natural scientists’ coattails onto the endless frontier: The SSRC and the quest for scientific legitimacy. *Journal of the History of the Behavioral Sciences*, 40(4), 393–422.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 667–680.
- Stove, D. C. (1991). *The Plato cult and other philosophical follies*. Oxford: Blackwell.
- Tanner, P. L., & Danielson, M. L. (2007). Components necessary for the preparation of the scientist-practitioner. *American Behavioral Scientist*, 50(6), 772–777.
- Tashakkori, A., & Teddlie, C. (2003a). The past and future of mixed methods research: From data triangulation to mixed model designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 671–701). Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. (Eds.). (2003b). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.
- Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioural sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 3–50). Thousand Oaks, CA: Sage.
- Toomela, A., & Valsiner, J. (Eds.). (2010). *Methodological thinking in psychology: 60 years gone astray?* Charlotte, NC: Information Age Publishing Inc.
- Triplett, T. (1988). Azande logic versus Western logic? *British Journal for the Philosophy of Science*, 39(3), 361–366.
- van Buuren, H. (2006). Teaching statistics and research methods: An integrated approach. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.  
[Online: [http://www.stat.auckland.ac.nz/~iase/publications/17/3H1\\_VANB.pdf](http://www.stat.auckland.ac.nz/~iase/publications/17/3H1_VANB.pdf) ]
- Wiberg, M. (2009). Teaching statistics in integration with psychology. *Journal of Statistics Education*, 17(1).  
[Online: <http://www.amstat.org/publications/jse/v17n1/wiberg.html> ]
- Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Oxford: Blackwell.

AGNES PETOCZ

School of Psychology, University of Western Sydney - Bankstown  
 Locked Bag 1797, South Penrith DC NSW 1797  
 Australia