School of Nursing

Exploring the construct validity of the Context Assessment Index (CAI) using the Rasch model of item response theory with data collected in a Western Australian tertiary hospital

Joanne Hardy

This thesis is presented for the Degree of

Master of Philosophy (Nursing)

of Curtin University

June 2011

"To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university."

Signature: _____ Date: _____

Acknowledgements

It is important that I take this opportunity to acknowledge the many people who have contributed in various ways to the completion of this thesis.

In particular I wish to thank:

Dr Paul Snider, my principal supervisor, for his humour, patience, guidance and support without whom I would not have fulfilled my destiny and completed a project based in quantitative analysis.

Dr Irene Styles, independent consultant, for her generous contribution to my understanding and use of the Rasch model for statistical analysis.

Ms Sunita McGowan, my associate supervisor and Head of Department, for her unwavering support and belief that I would one day complete my Masters degree.

My nursing colleagues, within the Nursing Research and Evaluation team, for their interest and support of the project during the various stages.

The Western Australian Nurses Memorial Charitable Trust, who graciously provided funding and expressed a genuine interest in the nature and outcome of the project.

The nursing staff that so kindly completed the questionnaire, without whom there would have been no data to analyse.

Ms Margot Townley, my first nursing mentor, for believing in me all those years ago and being instrumental in my early professional development.

My friends and family who thought I was mad but supported me anyway.

And finally, I would like to dedicate this work to my husband Tony and our children Jamie and Isobelle. The completion of this project required us to work as a team and all of us had to make sacrifices. I look forward to being able to repay them by providing whatever support they require in their future endeavours.

Abstract

The adoption and integration of new knowledge into nursing practice is inconsistent. This is despite the active promotion of evidence-based practice by health care organisations and the nursing profession. The nursing profession is aware that there is a problem transferring empirical findings into accepted practices; however, it is not armed with the tools to determine what variables within a given clinical setting have the most influence on whether a change in practice is adopted. Nursing is not alone. There is growing recognition within health that context, i.e., the surroundings, circumstances, environment or settings in which care is delivered, is a critical, but poorly understood, determinate of success for any desired change in clinical behaviours across a range of professional groups.

A structured quantitative approach for assessing the relative influence of specific factors within each context is therefore advocated. It is believed that the most appropriate selection of strategies to overcome the identified contextual barriers can then occur. To achieve this goal, measurement instruments that explore the role of context within health settings need to be created and tested in multiple settings. Ideally such tools evolve from within sound theoretical frameworks and are then extensively tested in the real world. Such a tool was recently created in the Republic of Ireland.

The Context Assessment Index (CAI) was developed to contextually assess the readiness of different health care settings to implement procedural changes that reflect evidence-based practice. The CAI was derived from the Promoting Action on Research Implementation in Health Services (PARiHS) conceptual framework,

which recognises the complexities of implementing change within health care organisations.

The current research contributes to the testing of the CAI, by assessing the validity of the measurement instrument for potential use in Western Australian hospitals. This validation study is necessary because of differences in health care systems and the potential influence of Ireland‟s proximity to Europe and Australia‟s proximity to Asia on the recruitment of migrant workers. It is likely that the cultural mix of health care workers in these two settings is different. Consequently it is important to establish that the items within the CAI are measuring the same variables in Western Australia as in Ireland.

Study participants were recruited over a 2 week period from the medical and surgical wards of a metropolitan tertiary hospital in Western Australia. Participants completed the CAI and a demographic questionnaire to assist in identifying cultural difference. A total of 210 research instruments were returned for a response rate of 61%.

The Rasch measurement model of item response theory was used to statistically analyse the construct validity of the CAI in this Western Australian hospital. As nursing has historically relied on classical test theory to test measurement tools, this thesis discusses the relative contributions of different statistical approaches to validate the CAI. While preliminary Rasch analysis of the sample data has confirmed that the CAI is a unidimensional instrument, it has also identified that one item may not be contributing to the overall measurement of context. Further study using a larger sample is recommended to confirm these findings and to validate the absence of differential item functioning based on the cultural backgrounds of the respondents.

Table of Contents

List of Figures and Tables

Figures

Tables

Chapter One

Introduction

"Without theory, practice is but routine born of habit." (Louis Pasteur, 1854)

Changing Human Behaviour

History indicates that the widespread acceptance and diffusion of new ideas has long been a challenge of any human endeavour. There is comfort in the familiar, and this feeds inertia. To illustrate this point, the routine inclusion of citrus in the diet of naval crew is given as an example of how difficult it can be to change human behaviour.

In the late 1400s it was suspected that diet contributed to the development of scurvy during long sea voyages. Despite being the cause of numerous deaths, the possible link with dietary intake was not formally investigated until 1601. During a particular voyage to India, the crew of one British navy ship was rationed a daily serve of lemon juice, while the crew of three other ships were not. At the half way point of this trip, 40% of the crew on the other three ships (110 sailors) had died of scurvy, while no deaths from scurvy had occurred on the ship where the sailors were given lemon juice. Regardless of this result, the diet of British sailors remained unchanged.

In 1747 a second British study confirmed that the consumption of citrus was effective in preventing scurvy. But it took another 48 years before citrus was a standard inclusion in the dietary provisions of the British navy. From that point onward, scurvy ceased to be a problem on these ships. A further 70 years passed

before the regulatory authority for the merchant navy also mandated that citrus be provided on long sea voyages. Thus the widespread adoption of the consumption of citrus as a standard approach to the prevention of scurvy did not happen quickly. In fact, it took 264 years after the initial findings were confirmed by a second study for the evidence to be translated into normal practice (Berwick, 2003).

Habits and Health Care

As the above example illustrates, it is often easier to maintain the status quo than to alter established behaviours. The successful repetition of previously learned behaviours is integral to many areas of human activity, including the provision of health care. The development of professional habits is important when complex situations arise during patient management to ensure that a known and consistent course of action is undertaken by all members of the heath care team. As a consequence, there are numerous opportunities to develop these types of habits within a hospital environment. While undergraduate education provides the theory, a variety of clinical experience shapes the way in which care is ultimately provided. The junior health professional is expected to continually seek new knowledge and skill, learning from the daily interaction with peers and patients. Habits learned at this early period of professional development tend to be flexible. Later, when more established in the specialized role, changing a particular way of doing things is much harder. Habits will have developed that allow the professional to automatically process large amounts of information in a certain way, make a clinical decision and communicate care requirements in a timely manner. To embrace a new approach to care, there then needs to be a systematic dissolution of old thoughts and behaviours and a concentrated effort to build new habits. This can mean that the speed of clinical

decision making is initially reduced and accounts for some of the resistance encountered when a change in practice is introduced.

The mid 1990s was a defining point in the provision of health care. There were reports in the medical literature that some patients were being put at risk from unnecessary tests and procedures, and that others were not receiving the most effective or efficient treatment available for a particular condition (Wilson & Van Der Weyden, 2005). By the early 2000s, following several patient complaints and adverse events, the quality of services provided at a number of health institutions was under formal investigation (Dunbar, Reddy, Beresford, Ramsey, & Lord, 2007; Scott, Poole, & Jayathissa, 2008). Independent inquiries confirmed the potential risk to patients when health professionals were allowed to practice without a recognised process of peer review. Literature emerged that called for the widespread development and use of evidence-based clinical guidelines to provide surety that a certain level of care was consistently being delivered.

This development was met with suspicion and criticism by many in the medical profession (Straus & McAlister, 2000). While the concept of evidence-based medicine itself was not new, clearly defined best practice did not exist for every condition and health problem. The randomised control trial (long believed to be the gold standard for the determination of robust evidence) was not a realistic or cost-effective research method for every clinical question. The strength of the evidence available for different aspects of practice therefore varied. To overcome this limitation, clinical guidelines that provided a summary and synthesis of the available knowledge and innovation that was available for a particular area of practice were promoted. Some doctors, however, considered this concept to be a form of „cook book" medicine that would ultimately reduce the professional autonomy of the

clinician. Others saw it as a way for those interested in financial outcomes to define the type of care that would be provided by a particular institution. The seminal work by Sackett, Rosenberg, Gray, Haynes and Richardson (1996, p. 72) tried to address these concerns by explaining the expected relationship between clinical guidelines and clinical judgement in the following way:

> *Without clinical expertise, practice risks becoming tyrannised by evidence, for even excellent external evidence may be inapplicable to or inappropriate for an individual patient. Without current best evidence, practice risks becoming rapidly out of date, to the detriment of patients.*

These statements aimed to promote the recognition that clinical guidelines were not to be considered absolute. The clinician‟s judgement about the specific needs of an individual patient remained integral to the quality of care delivery. It was expected, however, that for aspects of care where evidence-based clinical guidelines existed, the reasons for any variation in practice would be clearly documented.

Variation in Nursing Practice

It was not long before the nursing literature also began to discuss the implications of using evidence-based clinical guidelines to direct practice. Like the medical profession, nurses have a long history in exploring which aspects of practice provide the best outcome for the patient (Marvin, 1927). Initially, quantitative research projects using scientific methods defined what was known about a subject and shaped care decisions. Later, qualitative research methods using humanistic and social science principles were recognised as being more useful when describing the patient experience. Currently, a mixed methods approach that combines both quantitative and qualitative techniques is often considered when investigating the impact of nursing practice on patient outcomes (Doyle, Brady, & Byrne, 2009).

Thus, while nursing is aware of the historical value placed on evidence derived from quantitative research methods, the profession also recognises the value of using the most appropriate research method to address the specific focus of each clinical question (Mantzoukas, 2007).

The process of investigating nursing practice has identified that there can be wide variations in the way traditional nursing tasks are performed. Elements of nursing care are not uniform even when nurses are following a set of routine medical orders. For example, the skills and techniques associated with maintaining patient comfort can differ between individual nurses and the teams in which the work is performed. In an attempt to limit the opportunity for unnecessary variation, standards of nursing practice are routinely revised by either regional special interest groups or senior nursing staff within individual hospitals (Mallory, 2010). New guidelines are then introduced into the clinical setting as updated policy and procedure documents or multidisciplinary clinical pathways. Audits can later examine whether these guidelines have been translated into routine clinical behaviours. The problem is that numerous examples have been reported in the literature in which this is not the case. These include evidence-based hand hygiene practices, pressure ulcer prevention strategies and nurse-led interventions for smoking cessation among patients. A similar situation has also been identified when reviewing attempts to remove an out-dated practice from a particular environment, such as the use of patient restraints for confused patients and longer than necessary periods of pre-operative fasting (van Achterberg, Schoonhoven, & Grol, 2008).

These findings reflect the complexity of human interaction that occurs within a hospital environment, which is often difficult to capture and even harder to describe. Many factors contribute to the circumstances in which nurses‟ work.

Members of the nursing profession are not a homogenous group. There are differences in education levels and specialisations. The skills required by nurses are not the same for all patient populations and work settings. The number and skill mix of nursing staff rostered for each shift is not constant and the required nursing composition differs among clinical specialities. The time and duration spent in each clinical area by other health professionals and non-clinical support staff also varies. Specific types of interaction between different types of personnel are required for diverse situations. Individual patient attributes and a range of presenting problems add to the complexity (Dodek, Cahill, & Heyland, 2010).

Thus the observable elements of nursing work may not provide a complete picture of the interactions that are necessary for the provision of care. Modifying the routine practice of nurses within hospital environments often requires a collective change in behaviour by many people, and progress can be difficult to measure. The success of initiatives that are designed to alter clinical behaviour can therefore be limited by the high degree of sophistication necessary to ensure that any change is sustained beyond the life of a particular project (Ploeg, Davies, Edwards, Gifford, & Elliot Miller, 2009).

Study Catalyst

The inherent complexities of healthcare delivery add to the challenge for health professionals when publicly funded hospitals are expected to demonstrate adherence to clinical guidelines to provide surety that quality health care is being delivered (Government of Western Australia Department of Health, December 2009). Funding bodies often promote the development and use of clinical guidelines to reduce duplication, limit variation, promote evidence-based clinical practice and satisfy the need for transparency and accountability in health care delivery. These

types of expectations inevitably lead to regular scrutiny and the formalised reporting of the care that is being provided by health professionals. The Safety and Quality Investment for Reform (SQuIRe) initiative is such an example. This program commenced in 2007 and is funded by the Department of Health in Western Australia. There are currently eight Clinical Practice Improvement projects within the program. Specific aspects of clinical practice are audited and reported on a quarterly basis to demonstrate a hospital‟s level of compliance with evidence-based criteria.

Pressure ulcer prevention is one such area. There is reasonable evidence that the use of a risk assessment tool on admission and subsequent use of specific interventions for at-risk patients can significantly reduce the incidence of hospital acquired pressure ulcers (Comfort, 2008). A recent Canadian study by Gallant, Morin, St-Germain and Dallaire (2010), however, warned that there is a notable discrepancy between what nurses know about pressure ulcer prevention and what is put into practice. This was preceded by another study conducted across Belgium hospitals that found many pressure ulcer prevention guidelines were not routinely integrated into daily practice (van Herck, Sermeus, Jylha, Michiels, & van den Heede, 2009). These findings reflect my experience over the last 4 years of managing the ongoing pressure ulcer prevention program within a tertiary hospital in Western Australia.

Since 2007, hospitals throughout Western Australia that are funded by the Department of Health have been actively encouraged to promote and audit nurses‟ compliance with the recording of a patient‟s risk of pressure ulcer development on admission to hospital (Government of Western Australia Department of Health, December 2009). High risk patients are then assessed to determine if the appropriate prevention strategies have been implemented. The audit process at the tertiary

hospital where I work has collected clinical data from over 5,500 patients. The data clearly demonstrate that clinical areas within an organisation do not uniformly adopt mandated practice (Hardy, 2010). Some clinical areas have made a number of changes at the local level and recommended best practice has become routine. Other areas have been much less successful. This is despite similarities in access to project leaders, print and electronic media, and additional educational support promoting best practice behaviours.

Audits for these types of projects usually involve the review of nursing paperwork (Government of Western Australia Department of Health, December 2009). My experience to date has confirmed that while many nurses do not document all of the care that is routinely provided, entrenched routines also encourage certain tasks to be signed off before being completed. This may mean that a pressure ulcer risk assessment and all the appropriate preventative strategies are documented, but there is no evidence that the care has been implemented at the bedside. The reverse may also be true. Observation at the bedside may indicate that all necessary care has been given, but there is no supporting documentation in the patient notes. For example, a patient at high risk of pressure ulcer development may have been placed on a pressure relieving mattress, pillows placed for protection between the knees, a turning and skin inspection regime commenced and the dietician contacted to arrange nutritional supplements – none of which in theory should have happened as a pressure ulcer risk assessment score has not been documented in the patient‟s notes. Both scenarios illustrate the difficulty in accurately measuring compliance with evidence-based guidelines if nursing documentation is audited in isolation from what is actually observed at the bedside at the time of care delivery.

This difficulty is compounded by another problem that I have observed in the clinical area. After an audit or incident highlights an area of concern, nursing staff are asked to focus on and improve either the care itself or the documentation surrounding that care. It is not unusual for a follow up audit to indicate that the area of concern has improved. Not long after, another aspect of care is identified as a new problem, and the whole process starts again. Once this issue is resolved, another problem is soon identified. The ongoing nature of this pattern has led me to consider that the capacity of nurses to deliver all the aspects that are considered to be optimal for patient care is limited by the constraints of their work environment. Highlighting an area of deficit seems to help in the short term, but often takes resources or attention away from another aspect of care. So while individual problems are addressed, the underlying barriers that prevent nurses from completing all the required tasks appear to remain.

The Role of Context

Gathering audit data and listening to anecdotal evidence within the clinical environment suggests that contextual factors regularly influence the nurse"s ability to document or deliver care. In recent times, for example, the political emphasis on bed management has increased the speed in which patients move through the hospital system, while the amount of time that nurses have to provide the evidence-based care itself has decreased. This can lead to tension between the achievement of organisational targets and the expectations of the professional role. Team dynamics and individual performance can be adversely affected if the required level of support is not provided throughout this period.

Research exploring how the contextual factors within a work setting influence the implementation of nursing knowledge is relatively new (van

Achterberg, et al., 2008). In a relatively short time, numerous inter-related factors have been identified and reflect the evolution of the health care system. Medical science continues to expand the treatment options available for a range of conditions. People with multiple co-morbidities are now admitted for complex procedures. Over the last decade, cost pressures as a result of public funding have also seen the length of time most patients spend in the acute hospital phase significantly reduced. Patients with higher care requirements are in hospital for shorter periods. The technology available to monitor and treat sicker patients has also increased in complexity. All these factors influence the amount and intensity of nursing care that is required (Duffield, Gardner, & Catling-Paul, 2008).

In 2007, Duffield et al. reported the findings of a large scale, Australian study that examined the impact of hospital work environments on nursing and patient-centred outcomes. These researchers found that a "typical" ward did not exist, with the level of variability within each area producing a demanding and unpredictable work environment. Demonstrated increases in patient acuity and movement into and out of nursing units added to the complexity. During the study period, patients moved 2.26 times in an average length of stay of about 4 days (not including intra-ward transfers), with an average of 1.25 patients passing through each medical-surgical bed every day. Each movement of the patient required the completion of nursing documentation, organisation of care, handover, checking of room equipment, assessment of the patient and planning of care. Shorter hospital stays also concentrated nursing actions that centred on the delivery of care into shorter time periods. Duffield et al. (2007, p. 151), thus concluded that in order to evaluate the impact of nursing care on patient outcomes "understanding the work environment is critical, including the composition of and relationship between team members."

A growing awareness of litigation issues has also increased the amount of record keeping and report writing expected by health professionals. Having a constant presence in the clinical area over each 24 hour period, with higher staffing numbers across each shift, means that the amount of care that nurses are required to document is greater than other disciplines. During each shift, there are numerous documents that nurses are required to complete that reflect the different phases of a patient‟s admission, as well as a summarised entry in the integrated patient record. Not surprisingly, the duplication of information is a common complaint, particularly when completing documentation can take up to 20% of the time available on a nurse‟s shift (Duffield, et al., 2008; Thompson, Wolf, & Spear, 2003).

While these types of environmental stressors are likely to occur to some degree in all large hospital environments, not all hospitals experience the same level of difficulty getting nursing staff to adopt evidence-based practices. Research has indicated that specific aspects of the working environment determine whether or not a suggested practice is successfully introduced and maintained over time. Successful hospitals have higher number of professionally satisfied nursing staff, and lower levels of nursing turnover and vacancy rates. Due to the relative stability of nursing staff in these environments, there are greater opportunities to explore the relationship between the nursing care that is delivered and the actual patient outcome (Gleason, Sochalski, & Aiken, 1991). The key characteristics that have been found in these types of working environments include a participatory and supportive management style, flexible nursing rosters, professional autonomy and responsibility, availability of specialist advice, an emphasis on continuing education, and career and management opportunities (Currie, Harvey, West, & Keeney, 2005). It would

therefore seem prudent to understand the relative presence of these types of factors whenever a recommended change in clinical practice is required.

Change management strategies often recognise the influence of contextual factors in determining the capacity of a nursing team to alter an aspect of clinical practice. Hence it can be hypothesised that being able to measure the relative strengths and weaknesses of a particular clinical context would be useful prior to introducing a change in practice, as it would provide insight into what elements need to be strengthened to ensure success. Strategies that are tailored to the local needs, opportunities and constraints of an environment are more likely to succeed. Thus this approach can be more cost effective in terms of time and effort (Plsek & Greenhalgh, 2001). Alternatively, acknowledging the constraints of the current working environment would help to establish a credible link between the required change and the context in which implementation is to take place. To do any of this, however, nursing needs a measurement tool that accurately assesses contextual factors within a range of clinical environments.

Development and Testing of Measurement Tools

The recent creation of the Context Assessment Index (CAI) by McCormack, McCarthy, Wright, Slater & Coffey (2009) has attempted to meet this need. The intent of the CAI is to measure which aspects of a clinical environment need to be strengthened to increase the likelihood that a change in clinical practice will be successfully adopted. The focus of this study is to test the CAI outside of the environment in which it was developed and to contribute to the discussion about its value to the nursing profession.

The quality of the responses elicited by any measurement instrument reflects both the strengths and weaknesses of the development process. Nursing has historically employed classical statistical techniques for the development and testing of measurement tools. A growing number of disciplines, however, are recognising the value of an alternative statistical approach (Hagquist, Bruce, & Gustavsson, 2008). For example, the fields of education and psychology have been applying item response theory to the development and testing of measurement tools for over 40 years (Tennant & Conaghan, 2007). Measurement tools which are developed for use in health care settings would also benefit from the additional testing and validation offered by this perspective.

Item response theory offers a more rigorous approach to statistical analysis than classical test theory and the differences will be discussed in more detail in later chapters. Importantly, the influence of demographic variables on the interpretation and meaning of the individual questions by different sub-groups within each sample can be investigated and determined. This has implications for the strength of findings that relate to the reliability and validity of the measurement instrument. A statistical approach based in item response theory such as that provided by the Rasch measurement model can also result in a bank of valid questions. These questions can then be used interchangeably when the same construct or idea needs to be assessed in the same or similar population over time. In addition, the statistical techniques of item response theory often identify that the required understanding of the captured data can be obtained while using fewer questions than originally anticipated (Streiner, 2010). The principles of item response theory and the Rasch measurement model are explained in *Chapter Three*.

Research Purpose and Objectives

This study tests the validity of the CAI outside of the environment in which it was developed. The CAI was developed in Ireland and as yet there are no other reports of its use. The implementation and interpretation of CAI data obtained from nurses working in a variety of clinical environments therefore needs to be investigated. The possibility that the tool may not measure the same thing in a sample drawn from a different cultural setting also needs to be explored. This study will use statistical techniques from classical test theory and item response theory to examine the construct validity of the CAI using data collected from a sample of Registered and Enrolled Nurses employed at a metropolitan tertiary hospital in Western Australia. Permission to implement the CAI in an Australian setting was granted by Professor Brendan McCormack, University of Ulster, Ireland (refer to Appendix A).

Although the CAI was developed by nursing researchers to assist the nursing profession to understand the context into which a change needs to be introduced, it is imperative that the overall measurement instrument and the individual items within the tool are subjected to vigorous scrutiny. While the primary intent is to test the applicability of the CAI outside of Ireland, it will also be used to demonstrate how the application of Rasch analysis can contribute to the development and implementation of clinical tools such as the CAI. This is important because it gives nursing the opportunity to develop experience and expertise in employing statistical techniques that are based in item response theory.

The purpose of this study is therefore to replicate and extend the psychometric testing process of the CAI using data collected from a preliminary

study within the organisational context of a Western Australian tertiary hospital. The specific objectives are to:

a) Assess the content validity of the CAI for a Western Australian setting using a convenience sample of registered and enrolled nurses working within traditional ward environments in a tertiary hospital;
b) Replicate certain aspects of the classical psychometric testing performed during the initial development of the CAI;
c) Gain experience in applying the Rasch model of modern item response theory at an exploratory level to determine if the CAI meets the requirements for objective measurement.

Significance

This particular study is significant for two reasons. The first is implementing the CAI in a Western Australian nursing population at a tertiary hospital which will add to what is known internationally about the validity, reliability and usability of the CAI within and across cultures. In doing so, this study will address one of the limitations identified by the team of nursing researchers that initially developed and tested the CAI (McCormack, et al., 2009). If the statistical methods employed in this study validate the CAI for use in Australian hospitals, researchers within Australia and elsewhere will be able to use it with increased confidence. This may lead to further use and development of the tool in a variety of clinical settings, ultimately providing data that could shape how nurses manage and influence change processes within the healthcare environment.

The second reason is the point of difference for this study. The statistical techniques usually employed by nursing researchers when examining the construct

validity of a measurement instrument will be supplemented by analysis based upon the Rasch model of item response theory. Undertaking this process will provide a valuable opportunity to gain new knowledge and entry level experience in a statistical approach that is slowly being embraced by the many professional disciplines that constitute the health sciences. An understanding of what Rasch analysis can contribute to the development and testing of measurement tools will also assist in the synthesis of relevant research outcomes from other health disciplines into nursing practice.

Thesis Outline

This thesis consists of six chapters, with *Chapter One* providing an introduction to the study, the objectives, significance of the study, and the structure of this report.

Chapter Two presents the challenge of changing clinical behaviours to reflect evidence-based practice. A discussion based on the tenets of knowledge translation and implementation science is provided. Research studies that investigate the influence of context on the adoption of evidence-based practice are critically reviewed. The emergence of the PARiHS framework to understand the complexity of interrelationships in the health environment is discussed and the subsequent development of the Context Assessment Index is outlined. The potential for culture to influence the assessment of context is included.

Chapter Three discusses the testing and evolution of new measurement instruments. The traditional use of statistical techniques based in classical test theory in nursing research is explained. The evolution of item response theory and the rationale for using the Rasch measurement model is then outlined. A discussion on the potential role of culture in the interpretation and response to a measurement tool

follows. The implications for the testing of the Context Assessment Index prior to use in Western Australia is then provided along with additional detail on the structure of the tool itself.

Chapter Four outlines the research methodology. The research approach is explained and access to the participants described. The chapter provides an outline of how the questionnaire was administered and what the data collection process involved. The characteristics of the preliminary study participants are described.

Chapter Five reports the findings drawn from the statistical analysis procedures based in classical test theory and item response theory. The results of the exploratory factor analysis are compared to that reported in the literature during the initial testing phase of the instrument‟s development. A preliminary application of the collected data to the clinical environment is then provided. Statistical procedures based in item response theory are used to examine the item and person parameters of the Context Assessment Index when used in the sample.

Chapter Six presents a discussion of the findings and recommendations that emerge from the research. Limitations are presented as well as the implications for nursing practice and future research.

Chapter Two

Literature Review

This chapter provides a review of the literature about the challenge of changing clinical behaviours so that evidence-based practice can be demonstrated. The review begins with an explanation of what is meant by evidence-based practice within hospital environments and the evolution of systems that synthesise large amounts of information on behalf of the clinician. This is followed by a discussion on the observed difficulty in changing clinical practice to reflect the evidence-based guidelines that are produced by this method. The emerging role of knowledge translation and implementation science in understanding how to promote the routine uptake of research outcomes into healthcare environments is then explored. A review of the research studies that investigate the influence of context on the adoption of evidence-based practice is presented, which includes the development of the PARiHS theoretical framework and subsequent creation of the Context Assessment Index. The chapter concludes by acknowledging the potential impact of culture on context and the construct validity of measurement instruments.

Practice in Hospital Environments

Hospitals provide an array of health services to meet a variety of patient needs within a complex environment. Service provision requires the interaction of many professional groups and seeks to balance the demands of increasing patient acuity within a diminishing time frame. In addition, prudent fiscal management is actively encouraged. Against this background, hospitals are increasingly being required to demonstrate to funding and regulatory agencies that the health care

professionals within their facility are following clinical practice guidelines that reflect the best available evidence when caring for patients (Kitson, 2009; Titler, Everett, & Adams, 2007).

Evidence-based clinical practice is considered to be in the best interest of the patient in terms of efficacy and consequently in the best interest of public and private funding bodies. As explained by Sackett, Rosenberg, Gray, Haynes and Richardson (1996, p. 71),

> *Evidence-based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research.*

The process involves consideration of what aspect of patient care requires clarification, asking the right question, retrieving the most appropriate information and critiquing the strength and quality of the available evidence (Jones, 2010).

The amount of research information available to clinicians to support evidence-based practice is vast and continues to grow. Mantzoukas (2009) highlighted the challenge this presents for clinicians in a review of the top 10 nursing journals published between the years 2000 and 2006. In total, 2574 research articles were identified as contributing to the understanding of nursing practice. Recognising the time constraints on clinicians, specialised groups such as the Joanna Briggs Institute have therefore been established to continuously evaluate and synthesise the evidence that is available for a particular aspect of patient care. This then leads to the development of best practice guidelines that are published for distribution and implementation within health care settings.

The problem for hospitals is that although this process results in solid evidence that can guide the delivery of health care in a number of clinical areas, there is no guarantee that best practice guidelines are routinely or uniformly implemented (Marchionni & Ritchie, 2008; Titler, 2010). The literature strongly suggests that even when health professionals are aware that best practice guidelines exist and are supported by robust evidence, it is not enough to change clinical behaviours (Currie & Loftus-Hills, 2002; Currie & Watterson, 2007; FitzGerald, Ferlie, & Hawkins, 2003; Plsek & Wilson, 2001; Wallin, 2009). For example, FitzGerald et al. (2003) found that a professional group first needs to agree that the evidence is transferrable to a clinical setting other than that of the initial research. This process often requires time and dedicated resources to enable extensive consultation, explanation and negotiation. Only then can targeted strategies be used to overcome social, organisational and managerial factors that may influence the widespread acceptance of a new practice by the entire health care team.

Improving the delivery of health care by changing a clinician‟s behaviour is therefore a daunting task. Changes in knowledge outpace changes in behaviour. The literature suggests that the spread of new knowledge into practice is a slow and unpredictable process and that it can take years for an idea to become embedded in routine clinical practice (Eccles, Grimshaw, Walker, Johnston, & Pitts, 2005; Tansella & Thornicroft, 2009; Wallin, 2009). Even then, 30-40% of patients may not receive “treatments of proven effectiveness” (Marchionni & Ritchie, 2008, p. 267). Wallin (2009) notes that this phenomenon appears to be universal despite health care taking place in a variety of settings, communities and cultures and the influence of different economic, social, political, fiscal, historical and psychosocial factors.

The key to understanding this reality lies in recognising that health care is delivered within a complex dynamic system that is constantly subjected to changing practices against competing demands for limited resources (Hughes & Mackay, 2006; Kitson, 2009; Plsek & Greenhalgh, 2001). The highly developed relationships that are necessary within such a system tend to be more important than the individual parts (Plsek & Wilson, 2001). Many members of professional groups, such as nurses, become resilient to change imposed from those outside the group. Experience and expert opinion can be more highly regarded than new evidence based on empirical findings (Dopson, FitzGerald, Ferlie, Gabbay, & Locock, 2002; Scott, et al., 2008). This presents a challenge when attempting to modify a discrete aspect of clinical practice because although individual clinicians often recognise the validity of particular practice guidelines or initiatives, it can be difficult to alter well established patterns of care (Grol & Grimshaw, 2003). As a result, although many studies have explored how to transfer research into practice, very few have produced favourable results that can be replicated in other settings. This is despite the implementation of similar strategies and support processes. The main problem seems to be that certain aspects of the clinical environment are exerting a stronger influence on the clinician than the evidence for change.

This possibility initially led nursing researchers to focus on identifying the barriers influencing research use within specific clinical contexts. Thus the use of the Barriers to Research Utilisation Scale (widely known as the BARRIERS scale) has been extensively reported in the nursing literature as part of a structured approach to bridging the gap between research outcomes and clinical practice (Kajermo et al., 2010). During development, each item on the BARRIERS scale was assigned to one of four factors that reflected the characteristics of either the nurse (research values,

skills and awareness), the setting (organisational barriers and limitations), the research (quality) or the presentation (communication and accessibility) (Funk, Champagne, Wiese, & Torniquist, 1991). The participant rates the extent to which each of 28 statements (items) is perceived to be a barrier to the implementation of research findings. The BARRIERS scale was designed to be self-administered by clinicians, academics and administrators to assist each group to understand the perspective of the other when collaborating on strategies for practice change. The premise was that the resultant strategies would be tailored so that the barriers to practice change within each group were either overcome or their impact reduced.

However, a recent review by Kajermo et al. (2010) of the 63 studies published between 1991 and 2009 that included the use of the BARRIERS scale, failed to find any that used the gathered data to tailor interventions to overcome the identified barriers. This was despite the reviewed studies having taken place in a variety of clinical environments in different countries (n = 14), with a wide range of sample sizes (from n = 18 to n = 2009) and response rates (9% - 92%). The authors therefore recommended that "no further descriptive studies using the BARRIERS scale be undertaken… instead we recommend examination of various contextual and human factors for enhancing research use in a given organizational context" (p. 20).

This view is supported in the contemporary nursing literature with many authors arguing that more research is needed that focuses on either the development or use of theory as the basis for the design and implementation of any measures intended to identify how knowledge can be better translated into practice (Eccles, et al., 2005; Foy et al., 2011; Kitson, 2009; Titler, et al., 2007; van Achterberg, et al., 2008; Wallin, 2009). For example, it is asserted by van Achterberg, Schoonhoven and Grol (2008) that the contextual factors that influence nursing implementation of

new practices need to be formally identified and then used to guide in the selection of appropriate implementation strategies to change the way in which care is delivered. Kitson (2009) goes one step further and contends that it is imperative that the principles of knowledge translation and implementation science are incorporated into any nursing research that is intended to ultimately change clinical practice.

Transferring Evidence into Practice

The evolving fields of knowledge translation and implementation science reflect the growing concern across a number of professional disciplines about the gap between what is known and what is actually done (van Herck, et al., 2009; Wallin, 2009). The concept of knowledge translation originated in the field of adult education in the 1950s and was adopted by the health sciences in the 1990s. The term is now commonly used across a multitude of health disciplines within a range of countries (McKibbon et al., 2010). Such wide spread use has lead to knowledge translation being used to describe a variety of interrelated activities including "evidence-based decision making, research utilisation, innovation diffusion, knowledge transfer, research dissemination, research implementation and research uptake" (Estabrooks, Thompson, Lovely, & Hofmeyer, 2006, p. 28). Implementation science is concerned with similar principles but has a broader focus. It is defined by Titler, et al. (2007) as "the investigation of methods, interventions, and variables that influence adoption of healthcare practices by individuals and organisations to improve clinical and operational decision making, and includes testing the effectiveness of interventions to promote and sustain the use of healthcare practices" (p.S53).

When discussing the contemporary literature on knowledge translation and organisational change in nursing, Kitson (2009, p. 219) notes, "it is evident that

context plays a major part in getting research into practice." This assertion is supported by Grol and Grimshaw (2003) who reviewed 235 articles on guideline dissemination and implementation strategies. Twelve barriers that influenced the implementation of evidence into clinical practice were identified. After further analysis these barriers were all considered to be contextual in nature and were assigned at either the organisational, social or professional level (refer to Table 2.1).

Table 2.1 Barriers to Implementation of Evidence

Context Category	Barrier	Example
Practice environment (organisational)	Financial disincentives	Lack of reimbursement
	Organisation constraints	Lack of time
	Perception of liability	Risk of formal complaint
	Patient's expectations	Expressed wishes related to care
Prevailing opinion (social)	Standards of practice	Usual routines
	Opinion leaders	Not agreeing with evidence
	Medical training	Obsolete knowledge
	Advocacy	By pharmaceutical companies
Knowledge/attitudes (professional)	Clinical uncertainty	Tests for vague symptoms
	Sense of competence	Self confidence in skills
	Compulsion to act	Need to do something
	Information overload	Inability to appraise evidence

(Grol & Grimshaw, 2003, p.1227)

Greenhalgh, Robert, MacFarlane, Bate and Kyriakidou (2004, p. 615) came to a similar conclusion following an extensive review of 495 articles that focussed on the diffusion of innovations, stating that "the multiple (and often unpredictable) interactions that arise in particular contexts and settings are precisely what determine the success or failure of a dissemination initiative."

Thus, it appears that the likelihood that a nursing team will adopt a best practice guideline is strongly influenced by the interplay of specific contextual factors within the clinical environment. The challenge is that the elements of a particular context are not easy to clearly define, are often dynamic, and can include, but not be limited to, the interactions between "leadership, resources, time, support

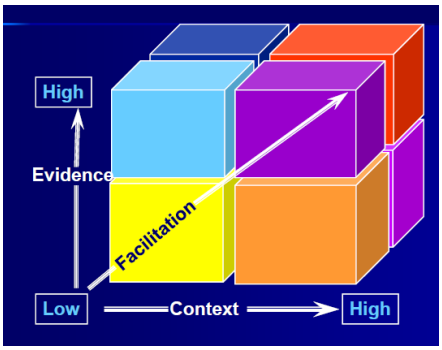
functions, staff development, interpersonal relationships, job pressure and organizational culture and climate" (Wallin, 2009, p. 582). As a result, context itself has been a poorly understood concept (Marchionni & Ritchie, 2008).

Wallin (2009) provides the development and refinement of the Promoting Action on Research Implementation in Health Services (PARiHS) conceptual framework as an example of how evolving theory can guide research in this area. PARiHS was developed to reflect the complexity of change within health care environments, taking into account the interplay and interdependence of the many factors that influence the uptake of evidence into practice. The original framework was published by Kitson, Harvey and McCormack in (1998) after being developed from collective nursing experience and wisdom. At the time, the authors observed that there was often the implicit assumption in the literature that the level and rigour of the available evidence was the most important factor influencing the uptake of new knowledge by clinicians. PARiHS challenged this assumption by arguing that it was actually the relationship between the evidence, the context and the facilitator (or implementation strategies) that accounted for whether a required change in clinical practice was adopted and sustained over time. Kitson et al. (1998, p. 152) therefore proposed that:

> *for the implementation of research into practice to be successful, there needs to be a clear understanding of the nature of evidence being used, the quality of context in terms of its ability to cope with change and type of facilitation needed to ensure a successful change process.*

This led the authors to express the PARiHS framework as SI = f (E, C, F), where SI indicates successful implementation and E, C and F represent evidence, context and facilitation respectively. The interplay between these elements was represented by a three dimensional continuum (refer to Figure 2.1).

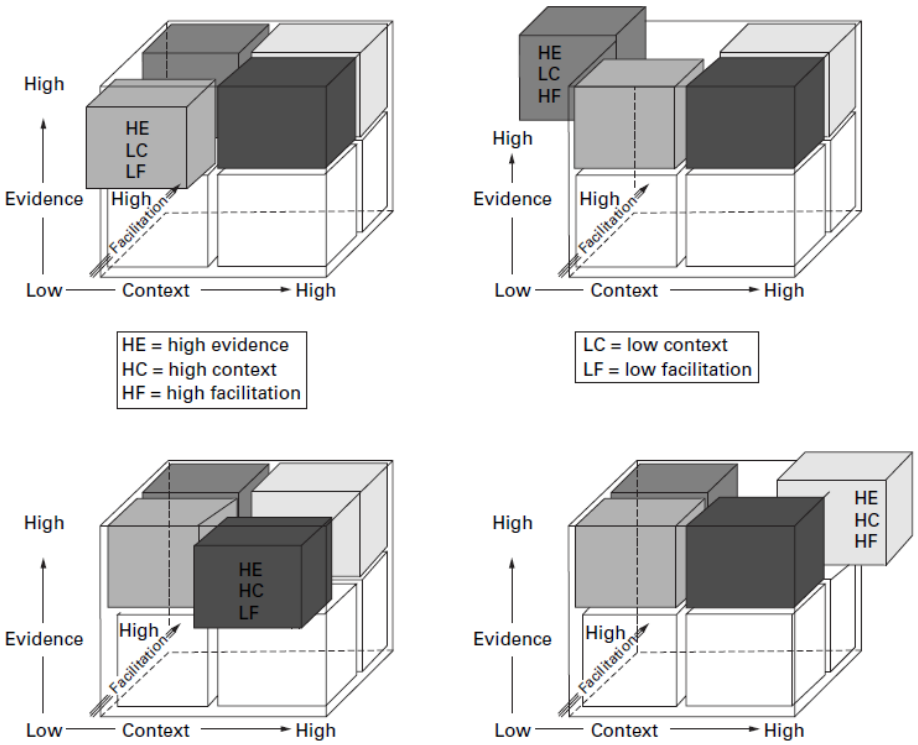
Figure 2.1 The PARiHS framework



(Harvey et al., 2008, p. 11)

To demonstrate the application of the conceptual framework, Kitson et al. (1998) examined four published studies that had investigated the implementation of research into practice. This process helped to illustrate the different relationships that could exist between evidence, clinical context and facilitation strategies within the PARiHS framework (refer to Figure 2.2).

Figure 2.2 Examples of possible interplay between PARiHS elements



(Kitson, et al., 1998, p. 153)

For example, the first study evaluated the effectiveness of facilitation strategies used to encourage the adoption of national guidelines by multidisciplinary teams in the rehabilitation of cardiac patients. Using the PARiHS framework, Kitson and colleagues identified that although a high level of evidence existed, the ability of the working context to adopt the practice was low and the strength of the methods used to facilitate this change was inadequate. Kitson et al. (1998) therefore surmised that despite the national guidelines being based on strong evidence, the change in practice was not successfully introduced because the facilitation strategies were unable to overcome the entrenched barriers that existed within the clinical area. In the second study, Kitson and colleagues identified that while the clinical evidence on how best to manage constipation in rehabilitation and respite settings was strong, there was limited support to alter practice because the ward culture was not patient-centred. Unlike the first study, however, the facilitation strategies in the second study were gradually tailored to address the specific challenges within this context, and the change in practice was eventually adopted. Kitson et al. (1998, p.155) therefore concluded that the "nature, focus and duration of facilitation can overcome and indeed alter poor contextual conditions to successfully implement research findings."

Since 1998, the PARiHS framework has undergone some refinement following a concept analysis of each of the dimensions: evidence, context, and facilitation, and a research study to assess its content validity (Helfrich et al., 2010). Each of these elements now includes a number of identified factors that can be arranged along a continuum of influence from weak to strong. Kitson et al. (2008) proposes that research implementation is more likely to occur when scientific evidence is robust and matches clinician experience and patient needs and clinical settings are supportive of the change. Additionally a positive culture and strong

leadership, relevant monitoring and feedback, and skilled internal and external facilitators involved in the process are needed for success. During the next phase of development, Kitson and colleagues plan to test the hypothesis that the PARiHS framework is best utilised as a two-stage process: firstly as a preliminary (diagnostic and evaluative) measure of the elements and sub-elements of evidence (E) and context (C), and then by using the aggregated data from these measures to determine the most appropriate facilitation method.

In 2010, Helfrich and colleagues published the results of a qualitative critical synthesis of peer reviewed PARiHS literature. Following a search in the CINAHL (Cumulative Index of Nursing and Allied Health Literature) and PubMed databases, the authors found 24 articles published in English before March 2009 that specifically made reference to the PARiHS framework. Six were concept articles from the original PARiHS authors, and 18 were empirical articles ranging from case reports to quantitative studies. According to Helfrich et al. (2010), while none of the empirical articles used PARiHS to prospectively tailor implementation strategies, there was a tendency to use the framework to organise the analyses and subsequent report of statistical findings. Despite this, Helfrich et al. (2010) found that the empirical studies generally supported the underlying tenets of the PARiHS framework by demonstrating the stability of the constructs when applied to specific projects investigating the translation of knowledge to practice. The authors also note that since March 2009, at least another five articles that refer to the use of the PARiHS have been published, with two of these detailing the validation process for measurement instruments that have been derived from the framework. It is surmised that these measurement instruments are the Context Assessment Index and the Alberta Context Tool which were both published in 2009 by members of the

PARiHS team (Estabrooks, Squires, Cummings, Birdsell, & Norton, 2009; McCormack, et al., 2009).

The Assessment of Context

Clinical areas that manage to successfully assimilate and diffuse changes into practice often have processes in place that recognise the role of such elusive factors as leadership, culture and team dynamics (Stevens, 2008). The challenge for researchers is how to measure these elements of context so that the data can be used to ensure the transfer of evidence into practice within a particular clinical setting.

McCormack et al. (2002) sought to address this problem by conducting a concept analysis to refine what context means when used in reference to clinical settings. This led to the central characteristics of the clinical environment being identified as the culture, leadership and the associated support system, which can include tangible and non-tangible resources. Context within the PARiHS framework is therefore defined by McCormack et al. (2002, p. 96) as "the environment or setting in which people receive health care services" and consists of three specific dimensions: culture, leadership and evaluation.

Culture recognises the prevailing beliefs and values, as well as consistency in these values, and the openness to change among the members of an organisation. Leadership is reflected by the nature of human relationships, role definition, teamwork (including involvement in decision making and learning) and organisational structures. Evaluation consists of the strategies used by an organisation to measure and provide feedback on the services being delivered.

Thus a change in practice is more likely to be successful in contexts where there is clarity of roles, a reliance on multiple sources of information on

performance, decentralised decision making and transformational leadership (Cummings, Estabrooks, Midodzi, Wallin, & Hayduk, 2007; Kitson, et al., 2008; Rycroft-Malone et al., 2002). Transformational leadership is enacted through charismatic role modelling, inspirational motivation to achieve a shared vision, intellectual stimulation to approach old problems in new ways, and consideration of the individual through mentoring and coaching (Eisenbeiss, van Knippenberg, & Boerner, 2008).

The PARiHS framework was used by Wright, et al. (2006) to investigate the specific factors influencing the use of evidence-based practice in continence care within rehabilitation settings for older people across two health care jurisdictions in Ireland. The initial phase of the project determined the contextual indicators that influenced continence promotion and management, which were then used to develop and test a contextual assessment instrument for the target population (Wright, et al., 2006). Measurement items that could be applied to a wider patient population were later identified. Further testing and refinement by McCormack et al. (2008) led to the creation of the Context Assessment Index (CAI). Additional information on the development of the CAI is provided in *Chapter Three*.

The CAI was designed to gather data that identifies the characteristics that are unique to the setting being assessed (McCormack & McCarthy, 2008). This was considered important because it is these unique characteristics that have been found to have the greatest influence on the perceptions and clinical behaviours of the staff who work in any given area (McCormack & McCarthy, 2008; Wallin, 2009). The CAI was therefore targeted at the team that works alongside patients within healthcare settings. Development and testing of the CAI across two jurisdictions in Ireland resulted in a series of statements (37 items) to which the respondent is

required to indicate a level of endorsement using a 4 point Likert scale: *strongly agree, agree, disagree,* and *strongly disagree* (refer to Appendix B) (Likert, 1932). The final tool emerged after an extensive period of development and testing using techniques and statistical analysis based in classical test theory. The principles underpinning this approach are explained in *Chapter Three*.

It was envisaged by McCormack and McCarthy (2008) that the CAI would be useful in a broad range of clinical settings whenever the acceptance and adoption of evidence-based practice needed to be encouraged. McCormack and McCarthy (2008) argue that the data collected with the CAI could be used to determine the relative strengths and weaknesses of the health delivery context in terms of culture, leadership and evaluation. The specific contextual characteristics that either enhance or hinder the receptiveness of a particular clinical area to a proposed change can then be identified and arranged along a continuum of influence from weak to strong (refer to Table 2.2). McCormack and McCarthy (2008) further propose that by addressing any contextual problems identified by the CAI, prior to introducing a new guideline or procedure, will increase the likelihood that a future change in practice will be adopted and maintained over time.

The potential therefore exists for the CAI to become recognised as a change management tool that is routinely used to collect data from clinical areas whenever a change in clinical practice needs to be introduced into a health care setting. Before this can happen, however, the CAI would require validation for use in the Australian context, as it has not yet been tested outside of Ireland.

Table 2.2 The PARiHS characteristics of context

Element	Sub-element	Indicators		
		Weak	Strong	
Context		Lack of clarity around boundaries	Physical Social Cultural Structural	Boundaries clearly defined
		Lack of appropriateness and transparency	Appropriate and transparent decision making processes	
		Lack of power and authority	Power and authority understood	
		Lack of resources	Appropriate resources	
		Lack of information and feedback	Information and feedback systems in place	
		Not receptive to change	Receptiveness to change	
	Culture	Unclear values and beliefs	Able to define culture(s) in terms of prevailing values/beliefs	
		Low regard for individuals	Values individual staff and clients	
		Task driven organisation	Promotes learning organisation	
		Lack of consistency	Consistency of individuals role/experience to value: - relationship with others - teamwork - power and authority - rewards/recognition	
	Leadership	Traditional, command and control leadership	Transformational leadership	
		Lack of role clarity	Role clarity	
		Lack of teamwork	Effective teamwork	
		Poor organisational structures	Effective organisation structures	
		Autocratic decision making processes	Democratic inclusive decision making processes	
		Didactic approaches to learning, teaching, managing	Enabling/empowering approach to teaching, learning, managing	
	Evaluation	Absence of any form of feedback	Feedback on: - individual - team - system	Performance
		Narrow use of performance information sources	Use of multiple sources of information on performance	
		Evaluations rely on single rather than multiple methods	Use of multiple methods: - clinical -performance -economic -experience	Evaluations

(McCormack et al., 2002, p. 102)

The Influence of Culture on Context

According to Prince (2008), it is often assumed that measurement instruments developed in one cultural setting are directly transferable to another, particularly if the sample participants are predominantly from English speaking backgrounds. The process of establishing construct validity, however, would not be complete without considering the possibility that a measurement instrument reflects the attributes of the cultural setting in which it was developed. Although there are similarities in the funding arrangements between the Irish and Australian public health care systems, it is possible that there are differences at the local level. This supposition is supported by Tracey (2006) who describes how Irish nurses work in hospital environments in which few experience positional power. Authority as supervision is often provided by general managers with no clinical background. In contrast, most Australian nurses are directly accountable to managers with nursing experience. These and other differences between Irish and Australian working environments may mean that both groups are not equal in terms of values, beliefs and social mores.

It is also likely that the underlying demographic and cultural immigrant composition of the two nursing populations is different. A key component of past and current international migration patterns is the recruitment of skilled workers from developing countries to industrialised countries that are attempting to solve internal nursing staff shortages (Buchan, Parkin, & Sochalski, 2003; Humphries, Brugha, & McGee, 2008). Ireland, a member of the European Union and relatively close in proximity to Central and Eastern Europe, is apt historically to have had a higher proportion of immigrant nurses from these areas. In contrast, it is likely that Australia‟s immigrant nursing population reflects both its colonial past and proximity to its Asian neighbours.

Australia recorded the immigration of 7,822 Registered Nurses between July 2006 and June 2009 (Commonwealth Government Department of Immigration and Citizenship, 2010). The proportion of nurses from different countries of origin is difficult to establish as this information is not centrally located due to the existence of different entry routes for temporary and permanent migrants between the Australian States and Territories (Buchan, et al., 2003). What is known is that there has been an increase in nursing migration in recent years. The traditional sources of the United Kingdom, Ireland and New Zealand have been supplemented by a growing number of nurses from Asia and Africa (Commonwealth Government Department of Immigration and Citizenship, 2010; Parker & McMillan, 2007).

The situation in Ireland is somewhat different. Ireland has traditionally been an exporter of nurses, due to the ability of the nurse to earn a higher wage or standard of living elsewhere. Shortages in Ireland over the last 10 years, however, have resulted in up to two-thirds of new registrations for the nursing workforce being comprised of immigrant nurses (Buchan, et al., 2003). In Ireland there is only one entry point, the national registration body, making it easier to estimate the number of international recruits that make up the total new number of nurses in any given year. Until the last decade, the main sources of immigrant nurses for Ireland were the United Kingdom, the Philippines and South Africa (Buchan, et al., 2003). Changes to the working visa scheme between June 2000 and December 2006, however, resulted in 50% of the 9,441 working visas being issued to nurses from the Philippines and 40% to nurses from India (Humphries, et al., 2008).

Countries may differ substantially in norms, values and accepted behaviour. Importantly, these differences reflect the diversity of values that exist between cultural groups. Every person internalises multiple cultural aspects of society in order

to function appropriately and automatically in any given situation (Meeuwesen, van den Brink-Muinen, & Hofstede, 2009). Hofstede (1997), a cross cultural researcher, describes values as broad tendencies to prefer certain state of affairs over others. While exploring cultural diversity, Hofstede (1997) identified that people in different countries use different strategies to deal with four common problems: (1) social inequalities, (2) dealing with the unpredictable, (3) relationship between individuals and groups, and (4) the emotional role division between the genders. Based on these findings and further research, Hofstede then formulated five dimensions of national culture to explain how different countries dealt with these types of issues. The first four reflect the previously identified common problems, while the fifth recognises the tendency for people to either look to the future or hold on to the past. Thus the five dimensions of national culture identified by Hofstede (1997) are power distance, uncertainty avoidance, individualism–collectivism, masculinity–femininity and long-term orientation (refer to Table 2.3).

These dimensions can be used to explain the observed differences between cultural groups when interacting with the health system. For example, in countries such as Romania where the power distance is large, and the level of uncertainty avoidance and collectivism is high, the role of both the health professional and the patient are clearly defined and understood. Following a brief consultation that is limited to the presenting problem, the health professional is expected to direct the care of the patient. All members of the extended family are involved in ensuring that the treatment regime is administered and that the patient is compliant. Alternatively, countries such as Sweden exhibit a more individualistic approach, with a smaller level of power distance and uncertainty avoidance. The relationship between the patient and the health professional is less clearly defined, consultations are not as

brief and there is a greater emphasis on the psychosocial aspects of illness. However, the onus of responsibility on the patient is also higher, with the underlying belief that the maintenance of health and successful treatment of disease is dependent on the characteristics and behaviours of the individual (Meeuwesen, et al., 2009).

Table 2.3 Dimensions of National Culture

Name	Description
Power distance	The extent to which the less powerful members of institutions and organizations within a country expect and accept that power is distributed unequally (from small to large).
Uncertainty avoidance	The extent to which the members of a culture feel threatened by ambiguous or unknown situations (from weak to strong).
Individualism vs. collectivism	Individualism stands for a society in which the ties between individuals are loose: everyone is expected to look after himself or herself and his or her immediate family only. Its opposite, collectivism, stands for a society in which people from birth onwards are integrated into strong, cohesive in-groups, which throughout people's lifetime continue to protect them in exchange for unquestioning loyalty.
Masculinity vs. femininity	Masculinity stands for a society in which emotional gender roles are clearly distinct: men are supposed to be assertive, tough, and focused on material success; women are supposed to be more modest, tender, and concerned with the quality of life. Its opposite, femininity, stands for a society in which emotional gender roles overlap: both men and women are supposed to be modest, tender, and concerned with the quality of life.
Long-term vs. short-term orientation	Long-term orientation stands for the fostering of virtues oriented towards future rewards, in particular perseverance, thrift, and adapting to changing circumstances. Is opposite, short-term orientation, stands for the fostering of virtues related to the past and present, in particular respect for tradition, preservation of 'face', and fulfilling social obligations.

(Meeuwesen, et al., 2009, p. 59)

Such actual and potential disparity between the cultural beliefs of patients and health professionals may also affect the respondents" interpretations of the 37 items which make up the CAI. Berry (2004), another cross cultural psychologist, argues that inherent cultural differences between individuals can play an important role in the development and maintenance of team dynamics within the work setting. Berry identifies the six most influential dimensions of cultural variation as diversity (how

people define themselves in relation to other cultural groups), equality (the importance of informal and formal hierarchies), conformity (the influence of social norms and obligations on behaviour), wealth (access to resources), space (closeness in interpersonal relationships) and time (the value of punctuality). According to Berry, these six dimensions play a defining role in group cohesiveness and homogeneity and can have a greater influence on work practice than the individual‟s professional role. This has implications for both the interpretation of the items on the CAI and the implementation of any strategies to address the contextual deficits identified in a given working environment.

As a consequence, the current study includes an exploration of the possible influence that culture may have had in the way each CAI item was answered by the respondents. Statistical analyses using the Rasch measurement model can establish that each item on the CAI measures the same construct for any particular subgroup that is identified within the total sample of respondents. To acquire the information necessary for this type of analysis, study participants were asked to complete a set of demographic questions that covered specific aspects of culture, such as the country of birth and first nursing qualification, and the language currently spoken at home. The tenets of item response theory that provide the foundation upon which the Rasch approach to this type of analysis is based are explained in *Chapter Three*.

This chapter has provided a brief overview of the challenges inherent in changing clinical practices within the complex environment of a hospital setting. The need to understand the influence of contextual factors in the transfer and implementation of new knowledge in clinical settings has been explained. The initial development of the CAI in response to this need has been outlined, as well as the

potential for culture to influence the interpretation of items on measurement instruments.

Acceptance of the CAI as a useful tool in the Australian setting therefore requires the testing of the instrument‟s construct validity and each of the individual items so that any potential limitations related to cultural specificity can be addressed (McCormack, et al., 2009). Quantitative procedures for establishing construct validity have historically included such tests as inter-item correlations and factor analysis. More recent developments have seen the use of statistical analysis derived from the Rasch measurement model based in item response theory (Prince, 2008). *Chapter Three* presents an overview of the statistical techniques that are available to test whether a measurement tool measures what it is intended to measure across a variety of samples. This includes an explanation of classical test theory and item response theory. The role of Rasch analysis in determining the cultural sensitivity of measurement instruments is also explained.

Chapter Three

Measurement Tools

This chapter discusses the principles of measurement that form the basis for the testing and refinement of new measurement instruments. The traditional use of statistical techniques based in classical test theory in nursing research is explained. The advantages and disadvantages of this approach are then outlined. This is followed by a description of the evolution of item response theory and the rationale for using the Rasch measurement model in the current study. The factors that limit the use and understanding of Rasch analysis in nursing are included. The advantage of applying Rasch analysis when exploring the potential role of culture in the interpretation and response to a measurement tool is then explained. The chapter concludes with a discussion on the implications for the testing of the Context Assessment Index (CAI) prior to general use in Western Australia along with additional detail on the structure of the tool itself.

Validation of New Instruments

The idea that objects, persons or entities have properties which can be thought of in terms of more or less, larger or smaller, stronger or weaker, and so on, forms the underlying principle of measurement (Wright & Linacre, 1989). Sometimes the property of interest is measured directly, e.g., the length of a piece of string, while for others it is only through the manifestation of that property that it is known to exist. For example, the ability of a person to complete a task can only be measured by observing the actions of the person while completing the task.

Many studies in health are interested in measuring phenomena that are not directly observable. These phenomena are known as latent traits. The testing of new

and evolving instruments in health therefore seeks to understand what observable aspects of the clinical environment can be used to indirectly measure these types of phenomenon. The aim of the development process is to demonstrate that although the instrument collects only observable measures, when viewed as a whole, the data provide practical information about the underlying phenomenon. These types of measurement instruments often rely on self reports from the subjects themselves and capture information about multiple elements which, when examined together, provide a level of understanding for things such as the work satisfaction of nurses or the pain perception of patients (Bond & Fox, 2007; DeVellis, 2006; Wright & Linacre, 1989).

Regardless of whether the variable of interest requires direct or indirect measurement, the process itself implies the existence of a calibrated well defined instrument that has previously been shown to work well enough to be useful. It is not enough to assume that counts of observed events are measurements in themselves, but rather, that they and others like them have formed the basis of a carefully constructed measurement instrument which reflects the relational nature of the observed counts to the variable of interest (Wright & Linacre, 1989). The key premise in the measurement of both observable and latent traits is that the trait can be mapped onto a line which forms a linear continuum. The line itself can be divided into equal units, which can be made greater or smaller from some point of origin. The process of mapping the amount of a trait onto a line thus involves the use of numbers. For such measurements to be meaningful, the units and the origin have to be agreed upon by those who use them.

Numerical values. The use of numbers in this way can lead to problems in terms of what the numbers mean in relation to the property being measured (Bond &

Fox, 2007; DeVellis, 2006). This is because counts of observed events are often assigned a numerical label for the purpose of measurement which can then imply a specific numerical relationship between the observed events (Wright & Linacre, 1989). For example, a measurement instrument may consist of 4 items that need to be rated by the respondent using a 4 point Likert scale with response categories ranging from *strongly agree* to *strongly disagree*. Each response category is then assigned a numerical value, with a response of *strongly agree* being assigned a value of 4, *agree* a value of 3, *disagree* a value of 2 and *strongly disagree* a value of 1. Although a necessary step in the conversion of observed counts into measurement, this arbitrary assignment of a numerical value implies that the interval between each of the response categories is equal without taking into account that it may be much harder for the respondent to choose to *disagree* with an item statement than it is to choose to either *agree* or *strongly agree* (Wright & Linacre, 1989). A second problem is also created. The sum of the assigned scores for each item is often assumed to be a reliable indicator of the respondent‟s level of the underlying trait being measured (van Alphen, Halfens, Hasman, & Imbos, 1994). The problem with this assumption is that a respondent who selects the *strongly agree* option for each of four items obtains a summed score across all 4 items that is 4 times higher than that of another respondent who selected the *strongly disagree* option for the same 4 items (Bond & Fox, 2007). Despite the 4 fold difference in the two summed scores, it does not necessarily reflect the size of the difference in the respondents‟ underlying trait. The testing of the reliability and validity of a measurement instrument therefore attempts to address these issues.

Reliability and validity. Reliability is concerned with determining that an instrument functions consistently between and within samples. Applied to the CAI,

this means that each of the 37 items in the instrument should remain similar in their function and meaning from person to person and group to group. Reliability is therefore established by investigating aspects of correlation, variation and error within and between respondent groups. In contrast, validity seeks to confirm that an instrument actually measures what it is intended to measure (DeVellis, 2006). In the case of the CAI, it will only be useful as a tool to measure the variable context if it has been confirmed that it does in fact measure the many different facets that are thought to influence context. As discussed in *Chapter Two*, the issue that then arises is that context itself has been acknowledged as a multifactorial and challenging concept. It is therefore essential that the construct validity of a measurement tool such as the CAI is established in the early stages of development before it is applied more widely.

Preliminary studies of new instruments such as the CAI tend to be exploratory, testing both the reliability and validity, with further studies required to confirm their usefulness in a variety of settings. The process usually requires investigating how each item in an instrument contributes to the understanding of the variable being measured, which ones are more likely to provide a consistent result, and the minimum number of items required for the instrument to be meaningful. While there are a number of statistical analyses available to assess the robustness of measurement instruments, nursing continues to favour the traditional statistical techniques based in classical test theory (DeVellis, 2006; Hagquist, et al., 2008; van Alphen, et al., 1994).

Classical Test Theory

In classical test theory the observed response of a person to an item on a questionnaire consists of two parts. The first is the level at which the respondents

understand what is being asked, and the second is the degree to which they are influenced by other factors during the completion of the questionnaire. Although an instrument is composed of separate items, the focus is at the test level (DeVellis, 2006; Hagquist, et al., 2008). For example, the results of a maths test are routinely expressed as the percentage of correct answers obtained by any given student, or as the average result for an entire class.

Classical test theory (CTT) consists of a set of principles that guide the determination of how successful a measurement instrument is at providing useful information about a latent trait from observable data. The basic premise of CTT is that an observed score is determined by the actual state of the latent trait plus the error attributed to other influences that affect the way in which a person responds to any given item. The observed score for each item by each person is thus a mixture of pertinent information and error. The actual state of the unobserved latent trait is its hypothetical true score (De Champlain, 2010; DeVellis, 2006). The probability that the observed score is obtained by chance is expressed on a linear scale from 0 (impossibility) to 1 (certainty) and is known as the *p*-value. The error associated with each item is assumed to be unique to that item, as well as being random in its effect, i.e., factors influencing the way in which a person answers any given item are assumed to be just as likely to increase or decrease the observed score. The aim is therefore to determine the accuracy of the item in measuring the latent trait and the magnitude of the error (De Champlain, 2010; DeVellis, 2006).

Statistical techniques. A well constructed item should result in a score that is relatively close to the true score. Because the true score will vary within the same individuals over time, as well as between different individuals, an observed score obtained from an item that is functioning properly will mirror this variation. The

association between the true score and the observed score therefore provides an important indication as to how accurate the observed score actually is (De Champlain, 2010). It is also assumed that every item is as good an indicator of the true score as every other item on the instrument, and that the magnitude of the error contribution to the total variance of each item is equal. Under these assumptions, the correlation between any two items is equal to the product of the correlations of each item with the true score. The square of the correlation coefficient is used to represent the proportion of variance shared between the true and observed score (Field, 2005). Thus the correlation of the observed scores for any two items provides a way of estimating the reliabilities of the items themselves to measure what they are intended to measure (DeVellis, 2006).

The process of establishing the reliability of each item results in the understanding that items that are more strongly correlated with each other are also more strongly correlated with the true score (Field, 2005). Similarly, an item that correlates strongly with the group of items on the instrument can more accurately discriminate between those respondents who demonstrate a low or a high total score over the entire set. These are therefore considered to be better items with which to measure the underlying latent trait (DeVellis, 2006).

The level of difficulty for a respondent to endorse/support/agree with an item is influenced by such things as knowledge, perception, understanding and experience. To explore the relative importance of these factors on the respondent‟s ability to answer each item, the number of respondents who choose a particular answer from an ordered set of response options is compared to how they have answered other similar items (Field, 2005). This notion explores the probability of respondents choosing particular response options based on how other items have

been answered in the instrument (DeVellis, 2006). This process in CTT is known as factor analysis.

Factor analysis is the primary means of assessing if the set of items meets the assumption that they are unidimensional, i.e., they only measure one underlying latent trait. It is usual to perform factor analysis to examine the variability within a construct and determine if certain items can be grouped into subscales that together help to explain different aspects of the phenomenon under study (Field, 2005). During the development and testing process of the CAI, for example, McCormack et al. (2009) used a combination of factor analysis and expert consensus to identify five subscales: *collaborative practice*, *evidence informed practice*, *respect for persons*, *practice boundaries* and *evaluation*.

Although CTT is concerned with certain properties of individual items, its primary interest is on the set of items and how they operate as a group. As more items are added to an instrument, the errors associated with each question are more likely to balance each other out and thus have a smaller effect on the sum or average score of each question (Streiner, 2010; van Alphen, et al., 1994). As a result, CTT emphasises that measurement instruments achieve strength through the number of items included, and supports the view that instruments with more items are more reliable. Increasing either the number of items or the strength of the correlations among those items will usually increase an instrument‟s reliability (DeVellis, 2006).

Use and limitations. CTT is frequently employed by nursing researchers for a number of reasons. It is common for nurse researchers to gather information using Likert scale questionnaires such as the CAI, and Likert scales are often analysed using techniques based in classical test theory (van Alphen, et al., 1994). The

statistical processes are therefore familiar to the nursing profession and are routinely taught in undergraduate and postgraduate courses. The methods of data analysis are also readily supported in statistical packages that are widely available, relatively cheap to purchase and fairly easy to use (Streiner, 2010).

Although popular, it is important to note that the use of CTT has some specific disadvantages. According to Streiner (2010), one concern is the assumption that "the amount of error associated with any particular item is unrelated to the true score, and that if we add up the error terms for all the items, the sum will...be equal to zero" (p.180). If the correlation of an individual item is not optimal, the researcher simply remedies the situation by progressively increasing the overall number of items. This works because the higher the number of items, the lower the overall error associated with the total score. As a result, researchers find it easier to increase the reliability of a measurement tool by increasing the number of items it contains rather than constructing better items with stronger correlations (Streiner, 2010). The final set of items often appears to work reasonably well together to provide information about the underlying variable of interest, even though the tool may be quite lengthy and consist of a series of items that seem quite similar.

Despite these limitations, the overall outcome is often considered satisfactory for the type of research investigation for which the instrument was designed (DeVellis, 2006; Hagquist, et al., 2008). In practice, as the number of items increases so too does the likelihood that the respondents will become frustrated with them and this can affect the way in which the later questions are answered. It can also decrease the likelihood that an instrument will be adopted in a busy clinical setting (Prieto, Alonso, & Lamarca, 2003; Streiner, 2010).

The other main disadvantage of CTT is the assumption that "all items contribute equally to the total score or that the total score is simply the sum of the individual items" (Streiner, 2010, p. 181). This leads to the common practice of each of the responses being summed to produce a total score, which is then assumed to be a reliable indicator of the latent trait measure for that respondent. That is, the respondents with the higher total scores are believed to be demonstrating a stronger presence of the latent trait (Thomas, 2010).

The implication behind this assumption is that all items on the tool are parallel measures of the same underlying construct, with any variation in the responses to the items being only attributed to the differences in the amount of the latent trait between the respondents (van Alphen, et al., 1994). This in turn relies on the assumption that the relative difference in the total scores between respondents is equal, because it is assumed that the relative value of each response category is also equal. The problem is that this approach does not promote rigorous scrutiny of the individual items, which may be particularly detrimental to measurement instruments that utilise Likert scales (Streiner, 2010). In reality, the capacity of the individual respondent to move between the Likert scale response categories of *strongly agree, agree, disagree* and *strongly disagree* as if they were equally spaced measures may vary between and within different sample groups for each item on the measurement tool (Bond & Fox, 2007; Wright, 1996).

The usefulness of summing ordinal data is therefore questioned in the literature (DeVellis, 2006; Prieto, et al., 2003; Streiner, 2010; van Alphen, et al., 1994). The problem mainly arises because many of the CTT statistics are sample dependent (De Champlain, 2010). For example, the *p*-value for a given test item will not represent the same level of difficulty if it is calculated from a sample with an

overall low level of ability than it would if calculated from a sample with an overall high level of ability. As a result, CTT is most useful when it is already known that the different sample groups to whom a measurement instrument will be applied are comparable in ability. It is least useful when the intent of the measurement instrument is to provide discrimination between individuals within a sample (De Champlain, 2010). In this instance, problems often arise because different samples with different variances do not yield equivalent data or data that can be compared across samples. This situation is particularly challenging when attempting to compare the usefulness of a tool across different populations (DeVellis, 2006).

Item Response Theory

Developments in the field of psychometrics suggest that item response theory (IRT) provides for a more robust interpretation than CTT of the internal construct validity of ordinal scales (Hagquist, et al., 2008). Importantly, the focus of IRT is on the item, and models the response of each participant of given ability to each item on the instrument. The term item is used because many measurement instruments use statements that require respondents to indicate a level of agreement along a continuum (such as a Likert scale) instead of a yes or no answer to a question. The basic premise in IRT is that the probability of a response to an item is a mathematical function of both person and item parameters (Baker, 2001; Bond & Fox, 2007). The latent trait or ability (for example, a person's intelligence or the strength of an attitude) is represented by the person parameter, while the difficulty or intensity of each statement is represented by the item parameter.

The information provided by statistical analysis based in IRT is more rigorous and sophisticated than that of CTT, and leads to greater opportunity for the researcher to improve the reliability and construct validity of the measurement

instrument. This occurs because IRT allows for the investigation of whether items on a scale are functioning properly and are making a valid contribution to the measurement of a single variable at a particular level on a scale. IRT assumes that at some level, i.e., either more or less, each person interacts with each item on the instrument. This interaction affects the interpretation of both the item by the person and the person by the item, with the unobservable trait being responsible for a person‟s response to any given item. This is why IRT is sometimes known as latent trait theory (Baker, 2001; Bond & Fox, 2007).

In addition, IRT supports the process of objective measurement by requiring that the total score obtained using a particular instrument has the same meaning across all relevant sample groups. Differential item functioning (DIF) is an IRT technique for studying item bias by identifying items that have different psychometric characteristics in two or more respondent groups within the sample. If differences exist, it may suggest that there is notable variation in the way the sub-groups of respondents are interpreting the content of the affected item (Streiner, 2010).

Thurstone Scale. Louis Thurstone played an important role in the development of IRT within the field of psychology. Thurstone (1927, p. 284) initially described the process of discrimination used by individuals when asked to evaluate the quality of such things as handwriting, and noted the existence of "ambiguity or qualitative variation with which one stimulus is perceived by the same observer on different occasions." This observation lead to the development of Thurstone‟s law of comparative judgment which recognises that a person‟s response to any given statement is influenced by the current perception of the relative value of that statement (Andrich, 1989; Thurstone, 1927).

In 1928 the Thurstone Scale was developed as a formal technique for the measurement of an attitude. Thurstone demonstrated that measurement instruments themselves are limited in being able to define an absolute value for latent traits, and are more useful when attempting to measure relative states such as happier or sadder (that is, more or less of an attribute) between and within groups of individuals. Thurstone argued that the measurement of attitude using psychological scales required items to occupy a point on a linear continuum, and that calibrating items on a linear scale should be independent of the sample of subjects responding to the items. This provided a statistical basis to check the consistency of a measurement instrument using a mathematical model (Andrich, 1989; Tenenbaum, 1999).

Guttman Structure. In the 1950s, Louis Guttman expanded upon Thurstone‟s contribution to the understanding of how to structure and interpret measurement instruments. Guttman, however, was interested in the measurement of ability rather than attitude. Guttman identified key requirements that a set of items should meet before the scores on the items could be meaningfully added to give a total score. The underlying principle being that if person A scores higher than person B on a given instrument, then person A should have endorsed all the items that person B has, and in addition, some other more difficult items (Bond & Fox, 2007). The ideal example in which this condition holds true for every pair of persons is known as the known as the Guttman pattern and is illustrated in Table 3.1.

Table 3.1

The Guttman Pattern for a Six Item Test

		Increasing Level of Item Difficulty						
	Person	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Score
Increasing Level of Ability (Strength of Opinion)	A	No	No	No	No	No	No	0
	B	Yes	No	No	No	No	No	1
	C	Yes	Yes	No	No	No	No	2
	D	Yes	Yes	Yes	No	No	No	3
	E	Yes	Yes	Yes	Yes	No	No	4
	F	Yes	Yes	Yes	Yes	Yes	No	5
	G	Yes	Yes	Yes	Yes	Yes	Yes	6

The Guttman pattern is considered to be deterministic, i.e., if the Guttman pattern exists in the data, the total score for each respondent determines which questions (in an ability test) were likely to be answered correctly and which ones were not. Each person's entire set of responses to all items can therefore be predicted from their cumulative score. For example, a person scoring a "4" on a 6 item test could be expected to agree with items 1-4 and disagree with items 5 and 6. In the current study this would mean that if some semblance of the Guttman pattern exists, then it could be surmised that the respondents with a higher total score have a more positive opinion of the context in which they work than those respondents with lower total scores. If, however, many people with the same score have entirely different patterns of item endorsement, then it would suggest that the variation of scores on individual items cannot be summarised by the total score, as it is possible that some of the items on the instrument are not measuring the underlying trait. Thus until the degree to which the Guttman pattern exists in the data is known, the outcome of the measurement for each person cannot be compared by simply looking at the total score.

Although the perfect deterministic Guttman pattern is a theoretical ideal, the principle upon which it is based provides an important function within the probabilistic framework of IRT (Andrich, 1982). Knowing in advance that the Guttman structure will be used as a frame of reference for later analysis encourages the construction of at least some relatively difficult items and some relatively easier items when developing a measurement tool. This can assist with both the initial development and testing of the construct validity of the measurement instrument. This is particularly true when using the Rasch measurement model to explore that the data conforms to the expectations of IRT.

Rasch analysis. The Rasch model of IRT incorporates the basic tenets from the work of both Thurstone and Guttman. It was named after the Danish mathematician Georg Rasch (1901-1980), and over the last 10 years has been used with increasing frequency in the fields of education and psychology to guide the development and the evaluation of measurement instruments (Streiner, 2010). Rasch contributed to the understanding of good measurement in a number of ways. Firstly, he realised that an instrument must retain its calibration or scale no matter what the setting in which it is being used. Secondly, the measurement instrument must maintain its level of difficulty regardless of who is responding to it. Thirdly, a person measured with a valid instrument will retain the same level of competence or ability regardless of which particular test items are used, so long as the items belong to a calibrated set of items which define the variable under study (Wright & Linacre, 1989).

The key feature of the Rasch model is that it allows for the characteristics of the item to be examined independently of the respondents using a set of properties that describe the relationship of the item with the underlying construct being

measured by the instrument. In the Rasch model, the only information required to characterise a measurement are the difficulty (or intensity) of an item and a person‟s ability or level. The focus of Rasch analysis is thus on item-level rather than test-level information (Hagquist, et al., 2008). This was explained by Rasch (1961, p. 332) in the following way:

> *The comparison between two stimuli should be independent of which particular individuals were instrumental for the comparison; ... (and) Symmetrically, a comparison between two individuals should be independent of which particular stimuli within the class considered were instrumental for comparisons… .*

The primary advantage of using the IRT Rasch model over CTT is that it allows the same standards of objective measurement to be applied to measurement instruments being used in the non-physical sciences as those used within the physical sciences. Scales or instruments developed according to the Rasch measurement model function the same way as a ruler on which the units of measurement are equidistant. While a ruler is limited to the measurement of physical properties, the Rasch model provides a means to measure abstract constructs such as latent traits using the same rigor (Hagquist, et al., 2008).

The Rasch model is considered to be a predictive mathematical model as it sits within a probabilistic framework. The predictive nature of the model implies that there must be a measurable relationship between the variables in the model and the observed phenomenon (Dougherty, 2009). Expected responses for each person-item interaction are produced, which makes it possible to examine the response of each individual in relation to the model‟s expectation (Finlayson, Peterson, Fujimoto, & Plow, 2009). As explained by Tennant & Conaghan (2007, p. 1359), using Rasch analysis enables the researcher to:

> *...test whether the response pattern observed in the data matches the theoretical pattern expected by the model (i.e., the probabilistic form of Guttman scaling). This difference (between observed and expected) is at the heart of the statistics used to test if the data fit the model.*

The Rasch model thus represents the structure which data should exhibit in order to obtain measurements from the data, i.e., it provides a criterion for successful measurement. Consequently, the Rasch model is not altered to suit data. Instead, the method of assessment and/or the design of the instrument should be changed so that the requirements of the model are met (Bond & Fox, 2007).

In the Rasch model, the Guttman response pattern is the most probable response pattern for a person when items are ordered from least difficult to most difficult and forms the basis for integer scoring used to analyse Likert scales. Importantly, the Rasch model takes into account that when measuring attributes other than ability (such as strength of an opinion), the overall response pattern (represented by an item characteristic curve) for a particular item may be too similar to the ideal Guttman pattern (ie: too perfect) or too random. This may indicate that the item is not functioning as a valid measure of the underlying construct. This concept will be explained in more detail in the analysis section.

Mathematical relationships. The Rasch model uses the item characteristic curve (ICC) to represent the theoretical relationship between the level of a respondent‟s skill (or, in the case of the CAI, strength of opinion) and the probability of choosing the correct answer (or, in the case of the CAI, agreeing with the item statement). Baker (2001) describes how the ICC is plotted along a numerical continuum with the mean of the skill level for all persons centred at *0*, the unit of measurement set at *1*, with a range of units from negative infinity to positive infinity

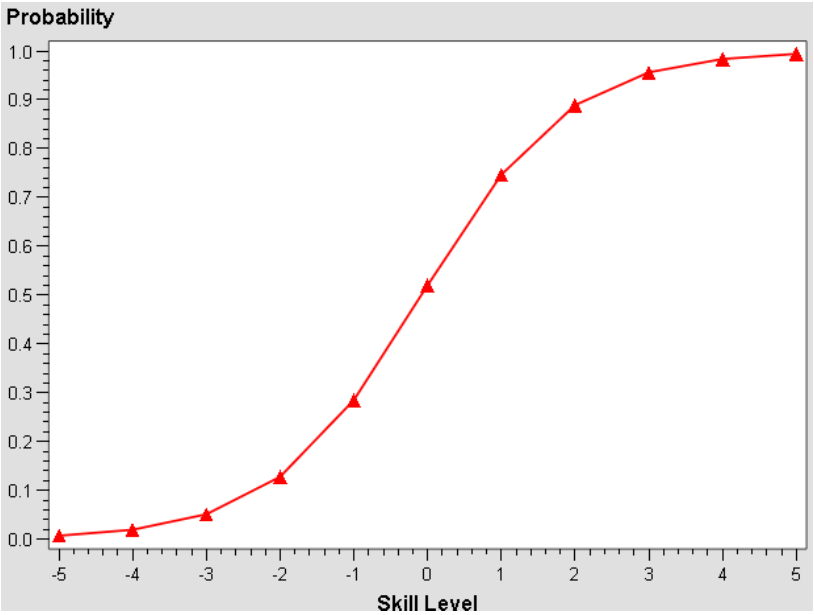
for all possible respondents. This type of measurement scale (numerical continuum) allows for a mathematical understanding of the relative ability (or strength of opinion) of any given respondent compared to any of the others. Each respondent can be assigned a numerical value (represented by the Greek letter theta θ), that represents the degree of trait being exhibited for each item by the respondent, allowing them to be placed at some point on the continuum. At each point θ on the continuum, there will be a specific probability denoted by P(θ), that the respondent will give a particular answer (which would be the correct answer in an ability test) to the item. For a typical item on an ability test, this probability will be smaller for respondents of lowest ability and higher for respondents of greatest ability. The ICC is the graphical representation of P(θ) when plotted against ability (Baker, 2001; Streiner, 2010; van Alphen, et al., 1994).

According to Baker (2001) the ICC plays an integral role in understanding the data as it illustrates two technical properties of an item. The first of these is the difficulty of the item, which describes where the item is located along the ability scale. The second is discrimination, which describes how well an item can differentiate between respondents whose abilities are above or below the item difficulty location. The steepness of the item characteristic curve is the key to understanding the discrimination of the item. The steeper the curve, the better the item can discriminate between respondents of similar ability (Baker, 2001).

The application of the ICC in the Rasch model is best illustrated with a series of examples. In Figure 3.1, the overall shape of the ICC suggests that the item being examined is of average difficulty, as a person with an average skill level (represented by the mean of *0*) has a 50% probability of agreeing with the item. For a person with

a relatively high skill level (i.e., around 4 on the *x* axis), the probability of agreeing with the item is close to 100%. Alternatively, the probability that a person with a lower skill level (i.e., around negative 3 on the *x* axis) will agree with the item is less than 10%. Thus in regards to the Context Assessment Index, the respondents with a stronger positive opinion (the underlying trait being measured) about a particular item (the context of their work environment), would be situated to the right of the mean, with those that are more likely to disagree being situated to the left. The overall shape of the ICC resembles an S-curve and suggests that the item discriminates evenly between respondents.

Figure 3.1. *The Item Characteristic Curve*

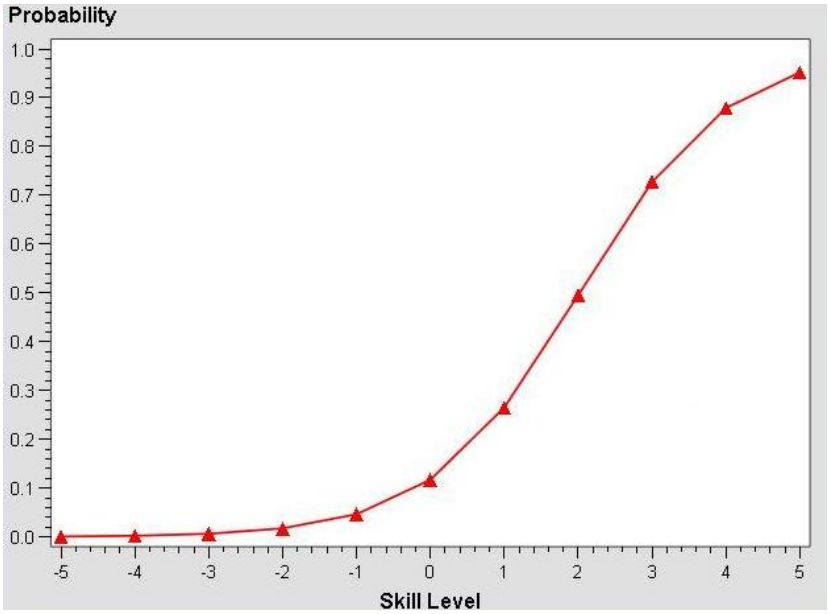


(Adapted from Ho Yu, 2009, p. 14)

The ICC looks different for items of different levels of difficulty (or intensity). For example, the ICC for an item that is harder for respondents to agree with will have a flatter slope below the mean. The ICC for this type of item suggests that it does not discriminate very well between persons with a lower ability, but does

help to discriminate those with higher ability. This is illustrated in Figure 3.2, in which the probability of respondents of average skill level (position denoted by the black arrow) agreeing with this item is less than 20%.

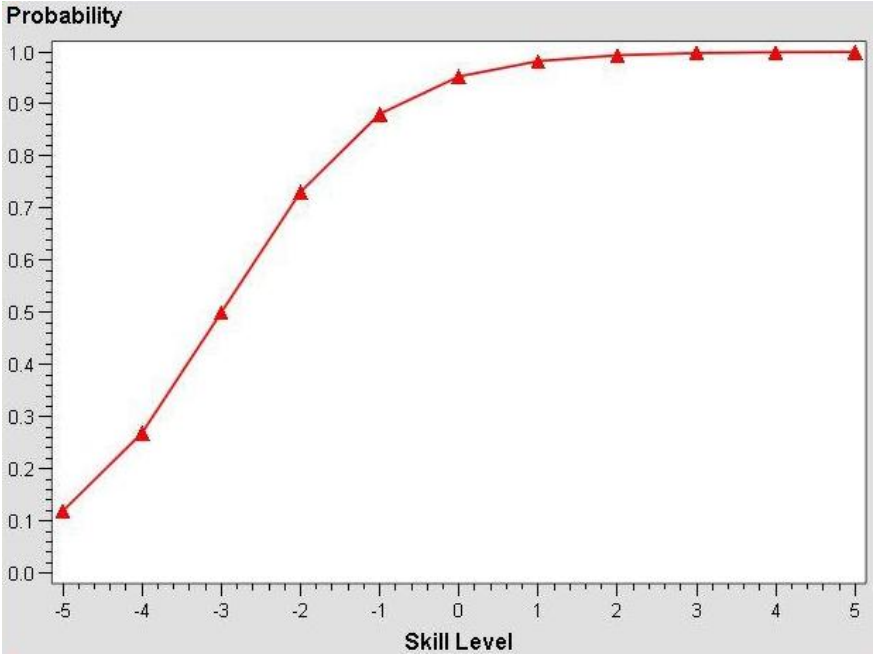
Figure 3.2. *The Item Characteristic Curve for a Higher Difficulty Item*



(Adapted from Ho Yu, 2009, p. 15)

Conversely, the ICC for an item that is easier for most of the respondents to agree with will have a steeper slope below the mean, and would be best used to discriminate between persons of lower ability. For example, Figure 3.3 illustrates the probability of respondents of average skill level (denoted by the position of the black arrow) agreeing with this item is around 95%.

Figure 3.3. The Item Characteristic Curve for a Lower Difficulty Item



(Adapted from Ho Yu, 2009, p. 16)

Application to Likert scales. The Rasch model can be applied when data is are either dichotomous (two response categories per item) or polytomous (more than two response categories per item). The polytomous Rasch model is suitable whenever the instrument‟s objective is to measure a trait or ability using items for which there is a set of ordinal responses that are scored with successive integers. Such is the case with Likert scales which are routinely used to collect attitudinal data (van Alphen, et al., 1994). When viewed from within the polytomous Rasch model, a score of *x* on a given item would imply that an individual has simultaneously surpassed the Likert thresholds below a certain region on the continuum, and failed to surpass the remaining thresholds above that region. In order for this to hold true, the thresholds must be in their natural order (Bond & Fox, 2007).

Unlike CTT which assumes that the intervals between the categories (*strongly agree*, *agree*, *disagree* and *strongly agree*) of a Likert scale are equal, the Rasch model takes into account the possibility that the probability for the choice of

each category will vary depending on the beliefs and attitudes of each individual, and that the influences that determine this choice will change between items. The model actually accepts the possibility of different thresholds between categories, thus allowing for the investigation of the assumption that the thresholds are equidistant and the same for all groups of persons and items (Tenenbaum, 1999). The statistical relationship between the latent trait and each response category can be mapped and is called a category characteristic curve (CCC).

The polytomous Rasch model uses both statistical and graphical data to identify any differences in the interpretation of Likert scale categories by testing the principle of invariance, i.e., is it likely that the tool consistently measures what it is intended to measure irrespective of differences within different samples (Andrich & Van Schoubroeck, 1989). Specifically, Rasch analysis allows the data to be investigated for any differential item functioning (often labelled item bias in CTT) to identify if the response categories of the Likert scale operated differently between existing subgroups of respondents for different items. The advantage of applying Rasch analysis in this study is that it will allow for the identification of any cultural differences in the participants‟ interpretations and responses to the CAI items. As explained by van de Vijer and Leung (1997), differential item functioning can occur when there are differences in the appropriateness of the item content (e.g., the underlying intent of the item is based upon a particular curriculum in one cultural group), the item design (e.g., the wording is complex), and the item translation (e.g., the interpretation of meaning is not universal among all respondents). The determination of differential item functioning is therefore important when investigating if a measurement instrument can be utilised outside of the country in which it was developed.

Advantages of the Rasch model. One of the assumptions inherent in the Rasch model is that for any given person the probability of endorsing one item is unrelated to the probability of endorsing another (Streiner, 2010). This is quite different from the CTT approach where items are often grouped together into subscales on the premise that respondents have provided a similar level of endorsement for these items and thus the items provide more information together than they do apart (DeVellis, 2006).

The Rasch model recognises that each response represents the observable outcome of a comparison between a person and item, with the probability of a person endorsing a statement depending on that person‟s ability to understand or interpret the statement and the inherent difficulty (or intensity) of the statement itself. The more able person is therefore more likely to endorse more difficult items, but due to the complexity inherent in measuring human traits, this outcome is not an absolute. As explained by Wright and Linacre (1989, p. 858) "Rasch recognized that the outcome of an interaction between an object-to-be-measured, such as a person, and a measuring-agent, such as a test item, cannot, in practice, be fully predetermined but must involve an additional, unavoidably unpredictable, component."

The Rasch model provides a method for nonlinear transformation of ordinal raw scores into equal interval measures. According to Ward and Linacre (1989), "Rasch deduced a mathematical model which specifies exactly how to convert observed counts into linear (and ratio) measures" (p. 858). The Rasch units of measurement are termed *logits* and are determined through logistical transformation of a probabilistic function. A logit reflects the amount of information in one item (Tenenbaum, 1999). When data fit the model, Rasch enables respondents and items to be placed on the same interval scale (numerical continuum), allowing direct

comparison of the item difficulty and the person ability estimates through a common log-linear scale (ie: “logit” scale). Parameter estimates can be separated, leading to person ability measures that are not dependent on the distributional properties of the items used, and item difficulty measures that are not dependent on a specific sample. Unique standard errors are calculated for each person‟s measure and each item‟s difficulty estimates. In CTT, only a single standard error is calculated, which assumes that uncertainty in the estimates is constant across all people and items (Streiner, 2010). With the unique standard errors calculated through Rasch, it is possible to estimate the amount of uncertainty across each item and each ability level during re-sampling.

This means that the Rasch model provides a theoretical foundation for the use of sequential integer scores, in such a way that the total raw scores are sufficient statistics for the parameters of the model (Bond & Fox, 2007). As a result, the raw score for an item or person is the sufficient statistic for the item or person parameter, i.e., the person total score contains all information available within the specified context about the individual, and the item total score contains all information with respect to the item‟s relevant latent trait. In doing so, Rasch differs from CTT by recognising that total scores are actually counts of discrete observations rather than direct measurements of the underlying trait. The Rasch model thus provides a basis and justification for obtaining person locations on a continuum from total scores on assessments.

In the current study, Rasch analyses are carried out using the interactive software RUMM2030 (Andrich, Sheridan, & Luo, 2010). This program provides tests of fit to the model and estimates of item and person locations. Fit statistics report the degree to which the observations match the specifications necessary for

measurement (Wright & Linacre, 1989). Item fit statistics are used to detect if any items do not contribute to the measurement of the construct, i.e., there are violations of unidimensionality at particular levels of scale. Person fit statistics are used to detect if an individual answers an item in an unexpected manner, which then brings into question the validity of that person‟s measure (Andrich & Van Schoubroeck, 1989; Waugh, 1999).

While there are several other benefits of using the Rasch model for instrument validation in comparison with CTT, one of the most important is that it allows for the characteristics of the item to be examined independently of the sample of respondents. This provides an opportunity to address the CTT limitation of sample dependence in which the psychometric properties of the instrument are dependent on the sample in which it was tested (Streiner, 2010). IRT analyses the individual items within an instrument by using a set of properties that describe the relationship of the item with the underlying construct measured by the instrument. It provides a means to establish the relative difficulty of each item and if each item in the tool reliably measures abstract concepts such as attitudes or beliefs. A Rasch analysis therefore allows for the identification of measurement problems within the items themselves that may not be easily detected by traditional analysis (Luo et al., 2009; Thomas, 2010).

Rasch analysis of health data. The recent evaluation of four commonly used Falls Risk Assessment Tools (FRATs) in residential aged care facilities in Queensland by Barker, Nitz, Low Choy and Haines (2009) is given here as an example of how nursing could benefit from understanding Rasch analysis. In applying Rasch statistical techniques to their data, the authors found that all four FRATs currently being used in Queensland did not fit the Rasch measurement model

and concluded that the measurement properties of each tool was poor. Barker et al. (2009) therefore concluded that none of the tools was better than knowing if the resident had fallen in the previous 12 months. This is significant result. Australian nurses in acute care settings are required to complete a formalised FRAT for each patient admission. Anecdotally many nurses have stated that it would be more efficient to ask if the patient has had a previous fall and under what circumstances. Hence the article by Barker et al. (2009) provides support for what nurses have inherently known but have found difficult to articulate to those outside the profession: the time required to complete a FRAT and the information provided does not necessarily lead to a decrease in the patient‟s risk of a fall.

It is unlikely, however, that the significant findings of Barker et al. (2009) will be recognised by nurses as a valuable opportunity to re-shape current practice. Nursing is not familiar with Rasch analysis, and continues to favour the traditional statistical techniques derived from CTT. The evaluation of measurement tools using IRT by nursing researchers is therefore limited (Hagquist, et al., 2008). A CINAHL search of the phrase "item response theory" and "nursing" yielded nine references published from 1994 to 2010. Of these, three were research articles written by either gerontologists or psychiatrists with samples taken from nursing homes. Another provided a discussion on the use of Rasch analysis when Likert scales are used to collect data (van Alphen, et al., 1994). Three were unpublished doctoral dissertations. This left two research articles that covered nursing related topics. The most recent by Abe and Henly (2010) was reported in a nursing journal and investigated the use of a Japanese translation of the Negative Acts Questionnaire (NAQ-R) as a measure of bullying. The authors (who were both nurses) used statistical techniques from CTT in the reported study, but recommended that IRT be

used as an alternative method of analysing future data. The other published article detailed the work undertaken after the Pain Assessment Checklist for Seniors with Limited Ability to Communicate (PACSLAC) was translated into Dutch and refined for use by nursing personnel. The outcome was the PACSLAC-D, which was tested by van Nispen tot Pannerden et al. (2009) to determine if the number of items on the original tool (n = 60) could be reduced. Use of statistical techniques based in IRT enabled the authors to demonstrate that this was possible (n = 14). The article was written by staff from the Faculty of Health, Medicine and Life Sciences of Maastricht University in the Netherlands and it is unknown if the group included any nurses. The article was published in *The Journal of Pain*.

It is therefore reasonable to conclude that there are very few journal articles in the nursing literature that detail the use of IRT techniques by nurses when analysing data. According to Streiner (2010) this is because nurses simply do not realise that these types of statistical tests are available to them. IRT is not routinely covered in undergraduate and postgraduate nursing courses. Nurses that do become aware of IRT must learn a new statistical language and obtain access to an experienced mentor (either within or outside of the nursing profession) to guide interpretation of research results. In addition, the statistical packages based in IRT are not as readily available as those tailored to CTT and tend to be more expensive (Borsboom, 2006; Streiner, 2010; Thomas, 2010).

Context Assessment Index.

The CAI evolved from contextual data gathered while exploring the use of evidence-based practice in continence care within rehabilitation settings for older people across two health care jurisdictions in Ireland. Continence was chosen after consultation with clinicians who determined that although reasonable evidence

existed for this area of practice, it was not being reflected in the actual delivery of care. The initial phase of the project determined the contextual indicators that influenced continence promotion and management. These data were then used to develop and test a practice assessment tool for the target population so that appropriate strategies could be introduced that promoted evidence-based continence management. The measurement instrument was then tested for reliability and validity by analysing data collected in rehabilitation settings (Wright, et al., 2006). At the end of this phase, approximately 300 items that could be applied to a wider patient population were identified.

Using the principles of CTT, a five stage process was then used to develop and further test a refined measurement tool that eventually became known as the Context Assessment Index. During the first phase, the 300 items of the continence practice assessment tool were reviewed and a list of items for each of the characteristics of the variable context (culture, leadership and evaluation) which had previously been identified by McCormack et al. (2002) was created. Items determined to have a repetitive theme were then amalgamated or removed. This left 88 items for the second phase of development. This involved each item being tested for face and content validity using a combination of qualitative and quantitative methods. Some of these remaining items required rewording so that the earlier focus on continence care could be removed. It was also determined that respondents would be required to answer each of the items (forced choice) using a 4 point Likert Scale. The 88 items were then examined by a panel of expert nurses to test the clarity, specificity and comprehensiveness of the evolving measurement instrument. Analysis of the feedback resulted in changes to the terminology and structure of

some items. Further repetition was identified and another 5 items were removed (McCormack, et al., 2009).

During the third phase, the remaining 83 items were included in a pilot tool that was completed by 460 respondents. McCormack et al. (2009) decided that the number of factors extracted from the data would be determined by eigenvalues over 1, with the total number of factors set at 20. Preliminary factor analysis thus resulted in 20 factors that accounted for 64.27% of the variance within the data. These 20 factors were then reduced to 7 by ensuring that there was a factor loading of 0.4 for each item and that there were at least 2 items grouped under each factor. Another 32 items were removed using these criteria. Preliminary examination of the 7 remaining factors and the items assigned to them by the researchers identified that an additional item needed to be removed. Secondary factor analysis was then performed on the remaining 50 items with the number of factors reset to 7. The factor structure of the previous analysis was replicated with the 7 factors accounting for 52.19% of the data variance. Following review, an additional 2 items were removed and the number of factors was reduced to 6 as one factor did not included 2 or more items with a factor loading above 0.4. Examination by an expert panel followed to determine consensus on the factor names and the items that would be grouped within them. Thus a combination of factor analysis and expert consensus eventually resulted in 44 items being retained by McCormack et al. (2009). These items accounted for 48.08% of the data variance and were assigned to one of 5 subscales within the instrument: *collaborative practice*, *evidence-informed practice*, *respect for persons*, *practice boundaries* and *evaluation*.

During the fourth phase, these 44 items were tested using a purposive sample of 23 respondents for reliability over time and ease of use. The respondents were

asked to complete the pilot tool twice, the second time being 2 weeks after the first. McCormack et al. (2009) felt that this was long enough for respondents to have forgotten the answers provided the first time and short enough to ensure that the culture of the work organisation had not changed. Given that the response options consisted of a 4 point Likert scale, the likelihood that a respondent would select the same response for the same item both times the pilot tool was completed was deemed to be 25% for each item. Analysis revealed that the agreement between responses to all items across both completion dates was higher than chance alone. The 7 items that had agreement levels of 55% or less however were removed as McCormack et al. (2009) deemed that these items were negatively affecting the correlation of items within the previously identified sub-groups (factors). The final version of the CAI thus contained 37 items. The final phase undertaken by McCormack et al. (2009) involved a 20 minute telephone interview with previous participants to discuss the ease of use of the CAI within clinical environments. Although a variety of comments was provided, this process did not result in any further changes to the measurement tool. In summary, however, the authors noted that "for the CAI to be generalisable there is a need for further testing of its constructs and items" and that the CAI "has not been tested for cultural diversity" (McCormack, et al., 2009, p. 34).

In the primary CAI study, the focus was on using statistics and strategies based in CTT to assist in the development of the overall instrument with an emphasis on how the items worked together to measure the different aspects of a given context. As a result, the individual items themselves were not subject to the type of rigorous statistical scrutiny that is available when using Rasch analysis. The current study will therefore build on the work by McCormack et al. (2009) in two ways. Firstly, the collected data will be used to re-examine the results obtained by McCormack et

al. (2009) through traditional statistical methods. Secondly, the usefulness of the CAI across different populations will be explored using Rasch analysis to determine if the individual items measure what they are intended to measure when used outside of the original sample.

The method used to conduct a preliminary investigation of the CAI‟s validity in a Western Australian setting is explained in *Chapter Four,* and includes a description of the demographic characteristic of the participants who participated in the study.

Chapter Four

Method

This chapter describes the research methodology of the study and the process of obtaining access to the potential participants. This is followed by an outline of how the questionnaire was distributed and the data collected. The chapter then concludes with a description of the demographic characteristics of the sample.

Research method and selection of participants

For the purposes of this initial investigation, the sample was one of convenience at the hospital where the researcher was employed. Ideally, the researcher would calculate the required sample size for a full psychometric analysis of the CAI and attempt to obtain that number of respondents. To do this an estimate of the number of nurses actually working in the public health system, and how many of these were directly employed by public hospitals would be required. From these figures, the number of nurses in roles at the bedside on a general ward would need to be determined. A central repository of this information does not currently exist. While it is known that there were 33,703 nurses and midwives registered with the Nursing and Midwifery Board of Western Australia in the financial year 2007-2008, registration only indicates the intent to practise, rather than the number of nurses actually employed within the health system. Each hospital would therefore need to be approached individually and asked to provide this information. The resources each hospital would require to meet this request are unknown. This situation led to difficulties in establishing the total size of the specific nursing population from

which the sample was drawn for this study. Hence an alternative procedure was used and is now described.

Approximately 500 nurses work in the general ward environments of the tertiary metropolitan hospital from which the sample was drawn. Areas which do not provide a traditional ward environment, for example, the Emergency Department, Peri-operative Services, and Intensive Care Unit, were excluded from this study. The researcher was not employed in a position of influence over any of the nurses that were invited to participate. The recruitment of participants began after gaining approval from the Research and Ethics Committee of the School of Nursing and Midwifery and the Human Research Committee at Curtin University, and the Research and Ethics Committee of the participating hospital. The researcher met with each of the Nursing Directors and Clinical Nurse Managers for the medical and surgical wards to briefly explain the details of the study and seek permission to recruit participants from each clinical area. All nursing supervisors gave permission for the research instrument to be distributed to staff and for the placement of collection boxes in a room commonly used by participants within the clinical area.

A list of the nurses scheduled to work within the designated fortnight of the data collection period for each participating ward was obtained by electronic query of the rostering system. This list was compiled late in the week immediately preceding the data collection period to maximise the likelihood that the nurses who appeared on the list would actually be rostered to work in the following two weeks. First year Graduate Registered and Enrolled Nurses who appeared on the roster were excluded from the study because they regularly rotate between various clinical areas as part of their program and are not considered permanent staff during their graduate year. The participant information sheet, instructions and research instrument were

placed in an envelope addressed to each prospective participant nurse (refer to Appendix C). The participant information sheet outlined the purpose, aims, nature and benefits of the study; a statement assuring the voluntary nature of the study, anonymity, and confidentiality of all data; and the right to abstain or withdraw from participating in the study.

The list of nursing staff allocated to each day‟s roster for the participating wards was printed 24 hours in advance. The addressed envelopes in each ward group were then reviewed, matched to the roster, and collated for distribution during the morning, afternoon or night shift of the following day. Prior to any envelopes being distributed, a sealed and brightly labelled collection box was placed in the nursing handover room/office in each clinical area for receipt of the completed CAIs and demographic information sheets. The exact location of the collection box was negotiated with the nursing co-ordinator.

During the first few days, an envelope containing the CAI and other materials was allocated to nearly all of the rostered staff on each shift. Over time, the overall numbers that required daily distribution quickly reduced as the same nurses appeared on subsequent shifts. To ensure that all eligible nurses were receiving an invitation to participate, the researcher cross checked each rostered shift with the remaining envelopes. This enabled the researcher to identify some nurses who were rostered to work but had not shown up on the list that was generated the week before the data collection commenced. The researcher was then able to provide these nurses with a hand addressed envelope containing the CAI.

During the first week the researcher was present shortly before or after each shift commenced on each ward participating in the study. This provided the

opportunity for the researcher to hand deliver the envelope to a number of the nursing staff. In some instances, when nursing staff were already involved in patient care, the nursing co-ordinator was informed that the envelopes had been left on the table in the room where nursing handover usually occurred. In others, the nursing co-ordinator nominated where the envelopes could be left for the nurses to collect.

A visual reminder of the study (in the form of a flyer) was also placed within sight of the hand basin in the treatment preparation and cleans up rooms, as well as on the noticeboard of the nursing office within each clinical area. The flyer encouraged the nurses to take the opportunity to share their view of their working environment by participating in the study, and included the researcher‟s contact details. The flyer was updated at the commencement of the second week, with its colour and content modified to further encourage the completed CAIs to be placed in the collection box. This process was repeated in the third and final week.

As the data collection period moved into the second week, most of the envelopes had been distributed. By the middle of the second week, 343 (85%) of the permanent nursing staff scheduled to work on the designated fortnightly roster during the data collection period had received the envelope containing the CAI and demographic information sheet. Further opportunities to distribute additional envelopes were limited as some of those remaining on the initial list were either moved to cover shortfalls in other ward areas, or were on sick leave during this period.

The collection boxes were emptied at the end of the first week, in the middle and end of the second week, and then removed from the clinical area a week after the final shift for the designated roster period. Concentrated targeting of the clinical

areas resulted in a response rate of 61% which provided a total sample size of 210 responses (refer to Table 4.1). The clinical area in which the participant worked was known for all but two of the completed CAIs (which were returned to the researcher through the internal mail shortly after the collection boxes had been removed).

Table 4.1

Summary of Distributed and Returned Questionnaires by Clinical Area

De-Identified Wards	CAI Distributed	CAI Returned	Response Rate
1	23	15	65%
2	31	16	52%
3	27	15	56%
4	30	15	50%
5	28	18	64%
6	31	13	42%
7	35	19	54%
8	28	23	82%
9	44	29	66%
10	32	20	63%
11	34	25	74%
Unknown	-	2	-
TOTAL	**343**	**210**	**61%**

Characteristics of Sample

Participants completing the research instrument were asked to answer the CAI and provide demographic information to assist with the statistical examination of any variations identified. To reflect the study‟s primary purpose of validating the CAI outside of its country of origin, some questions sought information on the participants‟ personal and professional cultural backgrounds (refer to Appendix D).

Consideration of the demographic variation among the participants assists in understanding the general composition of the study‟s sample. For example, the cultural variation within the sample may influence the appropriateness of the individual item content for respondents belonging to different sub-groups,

particularly for those born or educated outside of the Oceania region (van de Vijer & Leung, 1997). For the purpose of this study, the main language spoken at home is considered an important demographic variable because it can be an indicator of whether individuals within migrant groups are experiencing assimilation, segregation or marginalisation. This in turn can influence the degree to which the dominant culture of the adopted country is integrated into an individual‟s pattern of thoughts and beliefs. Cultural identity can therefore influence the interaction of the individual with work colleagues and the interpretation of contextual variables in the work place (Berry, 1990; Berry, 2004; Prince, 2008).

Respondents were born in one of 29 different countries, and obtained their first nursing qualification in one of 16 countries. To ensure a minimum data set for exploratory analysis, the classifications by country of birth and nursing qualification location were collapsed to give 4 main geographical and/or cultural groups – Oceania (Australia and New Zealand), Europe/America (England, Scotland, Belgium, Croatia, Ireland, Sweden, The Netherlands, Romania, Yugoslavia, Serbia, North America), Asia (China, Japan, Singapore, Hong Kong, India, Philippines, Nepal, Thailand, Malaysia) and Africa (Zimbabwe, South Africa, Ghana, Kenya, West Africa, Mozambique).

Slightly more than half (53.5%) of the respondents were born in the Oceania region, followed by Europe (22.5%), Asia (15.5%) and Africa (8.5%). Two-thirds had obtained their first nursing qualification in Oceania (66.7%), with the rest from Europe (17.2%), Asia (9.9%) and Africa (6.3%) (refer to Table 4.2).

Table 4.2

Participants Birth Region and Region First Nursing Qualification Obtained

		Nsg Qual Country			
		Oceania	Asia	Europe	Africa
Birth Country	Oceania	103	0	1	0
	Asia	10	17	2	0
	Europe & Nth America	12	0	29	0
	Africa	4	0	0	12

While most (81.7%) indicated that English was the primary language spoken at home, another 21 languages were listed by the remaining respondents. These other languages were also combined to give 3 alternative groups: Asian (11.2%), European (3.6%) and African (3.6%). The only difference in assignment of groups between the countries of birth and qualification, and the primary language spoken at home, was that Afrikaans was grouped with the European languages to reflect its Dutch origin rather than those from Africa (refer to Table 4.3).

Table 4.3

Participants Birth Region and Main Language Spoken at Home

		Language			
		English	Asian	European	African
Birth Country	Oceania	108	1	0	0
	Asia	7	21	0	0
	Europe & Nth America	40	0	5	0
	Africa	8	0	2	7

An undergraduate Diploma or Degree (57.7%) or a Hospital Based Diploma (34.2%) accounted for nearly all (91.8%) of the primary nursing qualifications. Close to a quarter (24.6%) had an additional nursing qualification for which they were paid a formal allowance. While the participants who obtained an initial nursing qualification from the Oceania region were twice as likely to have a undergraduate

Diploma or Degree than a Hospital Based Diploma, those who had acquired an initial qualification in the Asian region were nearly four times as likely to have an undergraduate Diploma or Degree than a Hospital Based Diploma. Participants who acquired an initial qualification from the European region were just as likely to have an undergraduate Diploma or Degree or a Hospital Based Diploma, while almost all participants who acquired an initial qualification from the African region had a Hospital Based Diploma (refer to Table 4.4).

Table 4.4

Region First Nursing Qualification Obtained and Type of Qualification

		Nsg Qual Type				
		Hospital Based Diploma	Diploma or Degree	TAFE Diploma	College Diploma	Aust Defence Force
Nsg Qual Country	Oceania	36	79	14	0	1
	Asia	4	15	0	0	0
	Europe & Nth America	16	16	0	1	0
	Africa	11	1	0	0	0

For reasons similar to those explained previously, it is also possible that the length of time the respondents have worked in Australia may influence item interpretation. As hospitals are often known to have a discrete culture, it is possible that those respondents who have worked for a longer period in the hospital will have had more opportunity to accept certain behaviours than those who have worked there for a shorter period of time (Schultz & Kitson, 2010). Similarly, clinical areas also develop a distinct culture over time.

While slightly less than half (46.4%) of the participants had worked as a nurse in Australia for longer than 6 years, 41.8% had done so for 3 years or less. Just over half (53.4%) had spent 4 years or longer as a nurse in this particular hospital,

with two-thirds (67%) having worked in the current ward for less than 4 years (refer to Table 4.5).

Table 4.5

Number of Years as a Nurse in Australia, Study Hospital or Ward

	Years			
	< 1	1-3	4-6	> 6
Australia	14	62	23	92
Hospital	23	79	25	64
Ward	43	85	20	43

While 27.5% participants held the more senior position of Clinical Nurse (CN), the majority (53.9%) held a Registered Nurse (RN) position, with Enrolled Nurses (EN) accounting for 18.6% (refer to Table 4.6). The participants who held Clinical Nurse positions were much more likely to have obtained an initial nursing qualification in the Oceania or European regions (94.5%) than those whose obtained an initial nursing qualification in the Asian or African regions.

Table 4.6

Region First Nursing Qualification Obtained and Current Nursing Position

		Nsg Position		
		CN	RN	EN
Nsg Qual Country	Oceania	40	57	33
	Asia	2	15	0
	Europe & Nth America	12	18	2
	Africa	1	11	0

Most participants (40.1%) were between 20-34 years old, while the proportion between either 35-44 years (24.8%) or 45-54 years (25.7%) were similar. A small proportion (9.4%) were 55 years or older. While the proportions of participants across the four age groups were very similar for those born in the Oceania and European regions, and participants from the African region were fairly

evenly spread across the first three age groups, the majority (64.5%) of participants born in the Asian region belonged to the youngest cohort (refer to Table 4.7).

Table 4.7

Participants Birth Region and Age Group

		Age (years)			
		20-34	35-44	45-54	> 54
Birth Country	Oceania	40	23	34	12
	Asia	20	8	1	2
	Europe & Nth America	15	11	14	5
	Africa	6	6	4	1

The vast majority (94.1%) of the participants were female, which reflects the traditional composition of the nursing workforce. As women often combine formal employment with child rearing responsibilities, professions such as nursing are more likely to have a notable proportion of part time employees (Australian Bureau of Statistics, 2006). Recent data indicates that part time employees make up 44% of the Australian nursing workforce (Duffield, et al., 2007). The proportion of part time nurses in the study population was similar. Less than 60% (58.6%) of the sample participants worked more than 60.8 hours, which represents 0.8 of a Full Time Equivalent (FTE) per fortnight. Just under one-third (30.5%) worked between 60.8 and 37.5 hours (0.5 FTE) per fortnight, and the remaining 10.8% worked less than 37.5 hours per fortnight (refer to Table 4.8).

Table 4.8

Hours of Work per Fortnight and Participant Age

		Age (years)			
		20-34	35-44	45-54	> 55
Nsg Hrs per FN	> 60.8hrs	58	25	25	10
	> 37.5 but < 60.8hrs	15	21	22	5
	< 37.5hrs	8	4	6	5

In summary, there were some notable differences between the participants when the general composition of the sample was examined. While participants born in the Oceania, European/American and African regions were spread across the defined age groups, those born in the Asian region were more likely to be under 35 years of age. The vast majority of Clinical Nurse positions were held by those with initial nursing qualifications obtained from the Oceania and European regions. Most of the participants born in the Asian region and half of the participants born in the African region spoke a language other than English at home.

As these differences may have influenced the way in which the items on the CAI were interpreted by members of each of these 4 cultural groups, the most likely profile for a participant from each of the four regions was identified to assist in defining some of the statistical procedures in Rasch analysis:

- Participants born in either the Oceania or European/North American region were similar, and likely to be between 20 and 54 years old, have obtained either an undergraduate Diploma or Degree as an initial nursing qualification in the same region, be employed in a Clinical or Registered Nurse position, and speak English at home.

- Participants born in the Asian region were more likely to be less than 35 years old, have obtained an undergraduate Diploma or Degree as an initial nursing qualification in either Asia or Oceania, be employed in a Registered Nurse position and speak a language other than English at home.

- Participants born in the African region were likely to be between 20 and 54 years old, have obtained a Hospital Based Diploma as an initial nursing qualification in the

African region, be employed in a Registered Nurse position and just as likely to speak English at home as another language.

The first stage of data analysis in this study reflects the exploratory approach undertaken by McCormack et al. (2009) during the development and testing of the CAI. The findings are drawn from statistical procedures based in classical test theory and are reported in *Chapter Five*. The reliability of the CAI for the study sample and the results of exploratory factor analysis are provided. These findings are compared to that reported by McCormack et al. (2009). A tentative summation of the strengths and weaknesses identified by the data in each of the clinical contexts from which it was collected is included. The principles of the Rasch measurement model are then applied and include statistical procedures that examine the item and person parameters of the Context Assessment Index when used in the sample. As there is a distinct possibility of differential item functioning that reflects the diversity of sub-groups to which the respondents may belong, this is also explored using Rasch analysis, and the results are presented in *Chapter Five*.

Chapter Five

Findings

This chapter reports the findings from the statistical analysis procedures derived from classical test theory and item response theory. Evidence of the reliability of the Context Assessment Index in the study sample is provided. The results of the exploratory factor analysis are compared to that reported in the literature during the initial testing phase of the instrument‟s development. This is followed with a tentative summation of the strengths and weaknesses identified by the data in each of the clinical contexts from which it was collected. The study data are then used to demonstrate how Rasch analysis provides estimates of the item and person parameters for measurement tools such as the Context Assessment Index. Examination of a broad range of psychometric properties of the instrument is also possible. Test of fit statistics are presented and the interpretation of the results explained. Item characteristic curves are used to further describe the operation of the items. A discussion on the role of differential item functioning in understanding the robustness of individual items across groups of interest is provided. The chapter concludes with a summary of the complete findings derived from both forms of statistical analysis.

Data Screening

Prior to the analysis, the data were screened for accuracy of data entry and missing values using SPSS version 18 (Statistical Package for Social Sciences) frequencies. It was noted that 43 respondents had missing data for at least one of the 37 items on the CAI instrument, and that five of these respondents did not complete

any of the 19 items on the second page. It is not known if this was a conscious decision by these five respondents or if the second page was accidentally missed.

Factor Structure and Psychometric Properties

In classical test theory, Cronbach‟s alpha (α) coefficient and correlation are commonly calculated to quantify the reliability of the measurement tool (DeVellis, 2006; Tennant & Conaghan, 2007). This has lead to Cronbach‟s α being considered a sufficient statistic for the reporting of an instrument‟s internal consistency in nursing research (van Alphen, et al., 1994). Cronbach‟s α represents the proportion of variance in a set of scores that can be attributed to a common influence on the scores of the individual questions. The underlying premise is that if there is a correlation between questions, they have something in common that gives rise to that correlation. Unless the correlation is perfect, each question also has some variation that is unique. The closer the Cronbach‟s α is to 1, the more the questions have in common; thus, the greater the proportions of shared variation and the more strongly they reflect a common true score (DeVellis, 2006). For example, when applied to measurement instruments being developed for use in the physical sciences, researchers aim to achieve a Cronbach‟s α close to 1 to limit the potential risk of error in diagnosis and treatment. Alternatively, when a measurement instrument is required to assess psychological constructs in the social sciences, a Cronbach‟s α greater than 0.7 generally indicates an appropriate level of reliability (DeVellis, 2006; Field, 2005).

In the original study by McCormack et al. (2009) which led to the development of the CAI, the Cronbach‟s α for the instrument was 0.93. The Cronbach‟s α for the current study was 0.95. These results strongly suggest that the CAI offers a reliable and consistent measurement of the underlying construct in both

study populations, and provides preliminary support for the use of the CAI in Western Australia. In the original study the 37 items were grouped into five distinct factors (or subscales) within the overall questionnaire, which then required the calculation of the Cronbach‟s α for each of these factors. This step was therefore replicated in the current study. Four of the five factors (*collaborative practice, evidence-informed practice, respect for persons, and practice boundaries*) demonstrated a satisfactory level of internal consistency with Cronbach‟s α of 0.88, 0.86, 0.74 and 0.78 respectively. The fifth factor (*evaluation*) achieved a Cronbach‟s α of 0.62 (refer to Table 5.1). As the formula for calculating Cronbach‟s α includes the number of items squared, this result may reflect the smaller number of items (n=4) that have been grouped into this factor when compared with the others (items 9, 11, 7 & 6 respectively) (Cortina, 1993).

The mean factor scores that were calculated were based on a Likert scoring system which ranged from 1 (*strongly agree*) to 4 (*strongly disagree*). This ordering of the response categories needs to be noted when interpreting the data because measurement is usually expressed from low to high. According to McCormack et al. (2009), however, mean factor scores from 1.0 to 2.5 indicated a high degree of agreement among the respondents that the factor, e.g., respect for persons, existed in a beneficial way within the clinical environment. Alternatively, if the mean factor score was greater than 2.5, the factor was more likely to have a negative influence during times of change. This step was replicated using the data from the current study. The mean of the factor scores for each of the five subscales indicated that there was an overall level of agreement that each factor was present in the clinical setting. The standard deviation for each of the factor scores indicated that there was little variation in the strength of belief among the participants (refer to Table 5.1).

Table 5.1

Comparison of Cronbach's Alpha, Mean and Standard Deviation

Factor	Cronbach's Alpha		Mean (SD)	
	Current Sample	McCormack's Sample	Current Sample	McCormack's Sample
Collaborative Practice	0.88	0.91	2.04 (0.13)	2.28 (0.45)
3 A proactive approach to care is taken				
6 HCPs provide opportunities for patients to participate in care decisions				
10 HCPs in the MDT have equal authority in decision making				
14 Patients are encouraged to be active participants in their own care				
19 HCPs and patients work as partners, providing individual patient care				
22 Discussions are planned between HCPs and patients				
28 Patients have choice in assessing, planning, evaluating their care & treatment				
31 Clinical nurse leaders create environ conducive to development/sharing ideas				
33 Patients are encouraged to participate in feedback on care, culture & systems				
Evidence-informed practice	0.86	0.88	2.01 (0.21)	2.24 (0.48)
4 All aspects of care/treatment are based on evidence of best practice				
7 Education is a priority				
11 Audit and/or research findings are used to develop practice				
17 The management structure is democratic and inclusive				
23 The development of staff expertise is viewed as a priority by nurse leaders				
27 Evidence-based knowledge on care is available to staff				
29 HCPs have the opportunity to consult with specialists				
32 Guidelines and protocols based on evidence of best practice are available				
34 Resources are available to provide evidence-based care				
35 The organisation is non-hierarchical				
37 Structured programmes of education are available to all HCPs				
Respect for persons	0.74	0.81	1.90 (0.20)	1.92 (0.38)
2 Decisions on care and management are clearly documented by all staff				
5 The nurse leader acts as a role model of good practice				
8 There are good working relations between clinical and non-clinical staff				
15 There is high regard for patients privacy and dignity				
20 Care is based on a comprehensive assessment				
26 Staff welcome and accept cultural diversity				
36 HCPs share common goals and objectives about patient care				
Practice Boundaries	0.78	0.80	2.01 (0.10)	2.05 (0.44)
1 Personal and professional boundaries between HCPs are maintained				
13 Staff have explicit understanding of own attitudes/beliefs to provision of care				
16 HCPs and heath care support workers understand each others' role				
21 Challenges to practice are supported and encouraged by nurse leaders				
25 Organisational management has high regard for staff autonomy				
30 HCPs feel empowered to develop practice				
Evaluation	0.62	0.78	2.19 (0.20)	2.50 (0.52)
9 Staff receive feedback on the outcome of complaints				
12 A staff review process is in place that enables reflection on practice				
18 Appropriate information is accessible to patients				
24 Staff use reflective processes to evaluate and develop practice				

*Note.*HCP = Health care professionals. MDT = Multidisciplinary team.

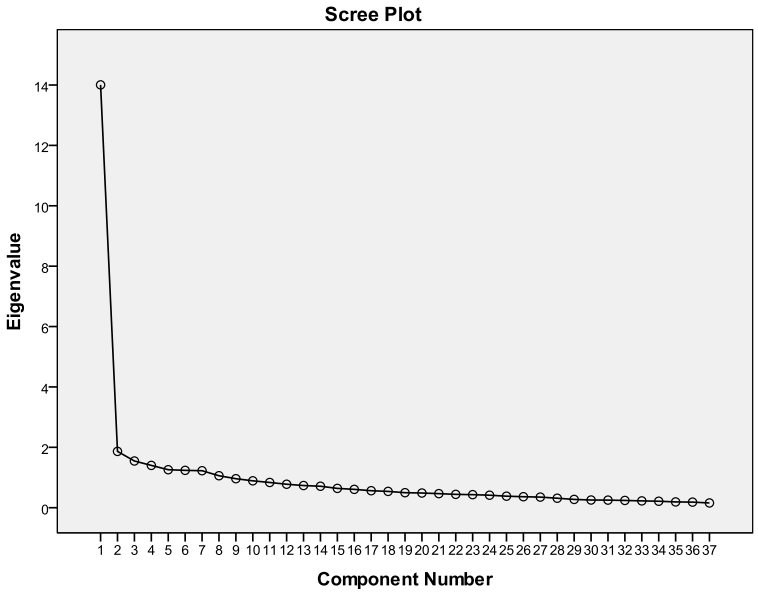
This result differs from the original study in which McCormack et al. (2009 p.30) found that Factor 5 (*evaluation*) was scored "with a sense of ambivalence towards the factor being present in the work place".

While a high Cronbach"s α that is close to 1 is often used as evidence that the items on a tool reliably measure a latent trait, it does not imply that the measure is unidimensional. Additional testing such as factor analysis is therefore required (Field, 2005). An eigenvalue conceptually represents the amount of variance accounted for by a factor within the data. Thus the eigenvalue associated with a latent variable indicates the substantiative importance of a factor (or factors) in explaining the phenomenon under study. Therefore, the main interest is in the factors with the largest eigenvalues. A technique to decide whether or not an eigenvalue is large enough to represent a meaningful factor is to plot a graph of each eigenvalue against the factor with which it is associated. The resultant graph is known as a scree plot. According to Field (2005), a scree plot based on a sample more than 200 can provide a fairly reliable criterion for factor selection. The current study"s data are therefore suitable for this type of analysis.

Usually the scree plot indicates that there are a few factors with high eigenvalues and many factors with relatively low eigenvalues. This results in a graph with a sharp descent followed by a tailing off. The decision point for selecting which factors are the most useful for explaining the variation within the data is thus at the inflexion of this curve. Generally it is recommended that all factors with eigenvalues over 1 be retained if the number of components is less than 30 and the resulting communalities after extraction are all greater than 0.7 (Field, 2005). This is not the

case with the data in the current study (refer to Figure 5.1). A total of 37 components are identified with 8 having eigenvalues greater than 1, while another 14 have eigenvalues greater than 0.7. The remaining 19 have eigenvalues less than 0.7.

Figure 5.1. The Scree Plot of Eigenvalues



There are several ways in which this result can be considered. The first retains the 8 factors with eigenvalues greater than 1, but this accounts for only 63.77% of the variation within the sample. Alternatively, all 14 factors with eigenvalues greater than 0.7 are retained as together they account for a more meaningful proportion (77.04%) of the variation (Field, 2005). The third and more practical approach, according to Field (2005), is to examine the scree plot itself. As the greatest difference in eigenvalues occurs between Factor 1 (*14.0*) and Factor 2 (*1.87*), with only a very gradual reduction in eigenvalues for the remaining 35 components, it appears that all the variation between the items is explained by only one underlying factor or latent trait. Rather than indicating a problem with the

instrument, this result strengthens the argument that the CAI is unidimensional. Interestingly, Kottner and Streiner (2010, p. 927) would argue that this conclusion could have been predicted from the relatively high Cronbach α for the entire scale as "it implies that the entire scale is tapping only one, homogenous construct." This possibility will be further explored using IRT, particularly as the analysis conducted by McCormack et al. (2009) identified 5 discrete factors (sub-scales) for the 37 CAI items. Any potential differences in the interpretation of the CAI items among the sub-groups can also be explored using Rasch analysis.

Rasch Analysis Overview

Using classical test theory (CTT), McCormack et al. (2009) found the CAI to have sound psychometric properties. This conclusion however is limited by the dependence of the results on the respondents to whom the instrument has been administered. In the current study, the Rasch measurement model is used to examine the psychometric properties of the CAI further because it is the only model that has the property of invariance, i.e., if the data can be shown to fit the model, then the instrument provides measures of both persons and items that remain invariant regardless of changes in the conditions in which that measurement takes place (Andrich, 1989; Wright, 1996). Thus every check of the data under the Rasch measurement model is really a check for different aspects of invariance of measurement. Once invariance has been established, it then becomes legitimate to add the scores on items so that the total score is a sufficient statistic to represent a person‟s ability, or in the case of the CAI, the person‟s level of opinion. The Rasch statistical techniques were therefore applied to data obtained from the current study to further test the construct validity and the reliability of the instrument. The polytomous Rasch model (PRM) was deemed appropriate for this analysis because of

the ordered response structure (4-point Likert response options) for all items in the CAI (Bond & Fox, 2007).

The Rasch Unidimensional Measurement Model (RUMM2030) computer software package (Andrich, et al., 2010) used in this analysis provides item and person tests of fit, as well as a graphical data indicating how items perform relative to the requirements of the model across a range of person estimates. RUMM2030 also allows for the investigation of differential item functioning (DIF), which recognises that invariant measurement requires that the total scores obtained on an instrument have the same meaning across all relevant sample groups (assuming it is necessary to make comparisons among groups or among items) and are thus measuring the same construct. Thus investigation of DIF examines whether the item appears to measure something different from what was intended for one or more subgroups of persons within the sample (Reise & Waller, 2009).

It is important to note that in the Rasch paradigm, there is no one way to establish the validity and reliability of the set of items that form a measurement instrument such as the CAI. Rather, the process relies on a series of checks of the data, which together provide a conclusion about the usefulness of the tool. The order of checking may vary, but here, the evidence for establishing the validity and reliability of the set of items is addressed in the following order: operation of item thresholds; test of fit of the items and persons to the model; differential item functioning; reliability, and targeting of the persons and items; the order of item locations; and, lastly, item dependencies and unidimensionality.

Item Thresholds and Response Categories

When an instrument is being used to measure an attitude, the relative difficulty of endorsing an item is often referred to as the intensity of the item. For example, if the item required a dichotomous response, such as true or false, to a knowledge question, the actual location or difficulty of each item along the continuum would be the point at which a person would have a 50% chance of being successful on that item. For polytomous data the number of decision thresholds at which the probability of endorsing an item in one or other of the adjacent categories is 50%, equals one less than the total number of response options. For example, in the Likert scale used in the CAI, there are three thresholds which identify the points at which the probability of a response of either *strongly agree* or *agree* (say, the first threshold), *agree* or *disagree* (second threshold), and *disagree* or *strongly disagree* (third threshold), are equally likely. These points are identified as the 1st, 2nd or 3rd response. The findings of the item threshold parameters for the CAI indicate that all item thresholds are ordered along the numerical continuum as required by the model, i.e., that threshold 1 occurred before threshold 2, which occurred before threshold 3 (refer to Table 5.2).

Table 5.2

CAI Item Threshold Parameters

Item	Description	Threshold		
		1	2	3
1	Personal & professional boundaries between HCPs maintained	-2.376	-1.006	3.382
2	Decisions on care & management documented by all staff	-3.026	-0.532	3.558
3	A proactive approach to care is taken	-2.994	-0.780	3.774
4	All aspects of care/treatment based on evidence of best practice	-3.421	-0.405	3.826
5	The nurse leader acts as a role model of good practice	-2.854	-0.752	3.606
6	HCPs provide opportunities for patients to participate in care decisions	-2.347	-0.754	3.101
7	Education is a priority	-2.346	-0.508	2.853
8	There are good working relations between clinical & non-clinical staff	-1.739	-1.350	3.089
9	Staff receive feedback on the outcome of complaints	-3.246	-0.825	4.072
10	HCPs in the MDT have equal authority in decision making	-3.251	-0.412	3.663
11	Audit and/or research findings are used to develop practice	-3.019	-0.917	3.935
12	A staff review process is in place that enables reflection on practice	-1.695	-1.275	2.970
13	Staff have explicit understanding of own attitudes/beliefs to provision of care	-3.149	-0.754	3.903
14	Patients are encouraged to be active participants in their own care	-2.271	-0.571	2.841
15	There is high regard for patients privacy & dignity	-4.173	0.145	4.028
16	HCPs & heath care support workers understand each others' role	-2.810	-0.903	3.713
17	The management structure is democratic & inclusive	-2.569	-0.539	3.108
18	Appropriate information is accessible to patients	-3.068	-0.114	3.183
19	HCPs & patients work as partners, providing individual patient care	-3.074	-0.497	3.571
20	Care is based on a comprehensive assessment	-4.044	-0.684	4.727
21	Challenges to practice are supported & encouraged by nurse leaders	-2.211	-1.114	3.325
22	Discussions are planned between HCPs & patients	-3.047	-0.650	3.697
23	The development of staff expertise is viewed as a priority by nurse leaders	-2.261	-0.693	2.954
24	Staff use reflective processes to evaluate & develop practice	-3.137	-0.533	3.671
25	Organisational management has high regard for staff autonomy	-2.870	-0.947	3.817
26	Staff welcome & accept cultural diversity	-2.674	-0.483	3.156
27	Evidence-based knowledge on care is available to staff	-2.724	-0.719	3.443
28	Patients have choice in assessing, planning & evaluating their care & treatment	-3.391	0.101	3.289
29	HCPs have the opportunity to consult with specialists	-3.394	-0.331	3.725
30	HCPs feel empowered to develop practice	-3.381	-0.617	3.998
31	Clinical nurse leaders create environ conducive to development/sharing of ideas	-2.291	-0.501	2.792
32	Guidelines and protocols based on evidence of best practice are available	-1.778	-1.170	2.948
33	Patients encouraged to participate in feedback on care, culture & systems	-3.405	-0.177	3.582
34	Resources are available to provide evidence-based care	-2.840	-0.682	3.522
35	The organisation is non-hierarchical	-2.537	0.047	2.489
36	HCPs share common goals & objectives about patient care	-2.567	-0.923	3.490
37	Structured programmes of education are available to all HCPs	-2.358	-0.780	3.138

Note.HCP = Health care professionals; MDT = Multidisciplinary team.

The ordering of the categories and thresholds may be seen in the category characteristic curve (CCC), which, for any given item, shows the most likely category response for a specific person location (the equivalent of a total score on all items). The intersections of the CCCs are the locations of the category thresholds. Figure 5.2 presents an example of the CCCs for Item 35 (*The organisation is non-hierarchical*) where the intersection of response thresholds 0 & 1, 1 & 2 and 2 & 3 occur in the expected order and are a similar distance apart.

Figure 5.2. *Category Characteristic Curve for Item 35*

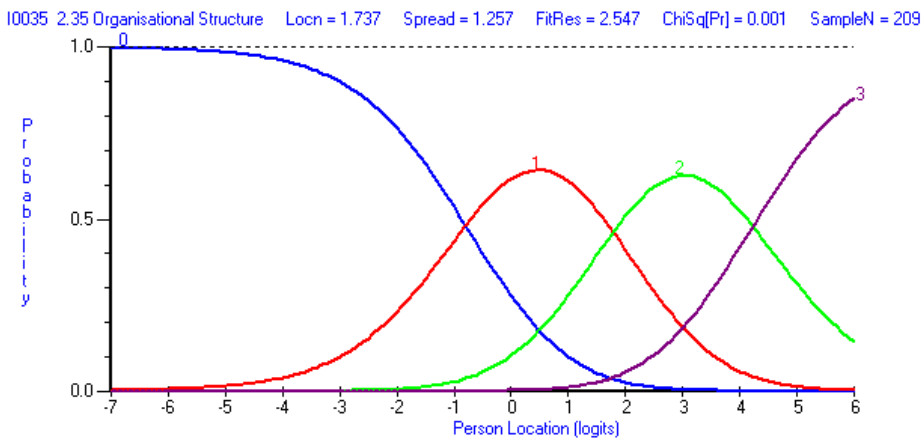
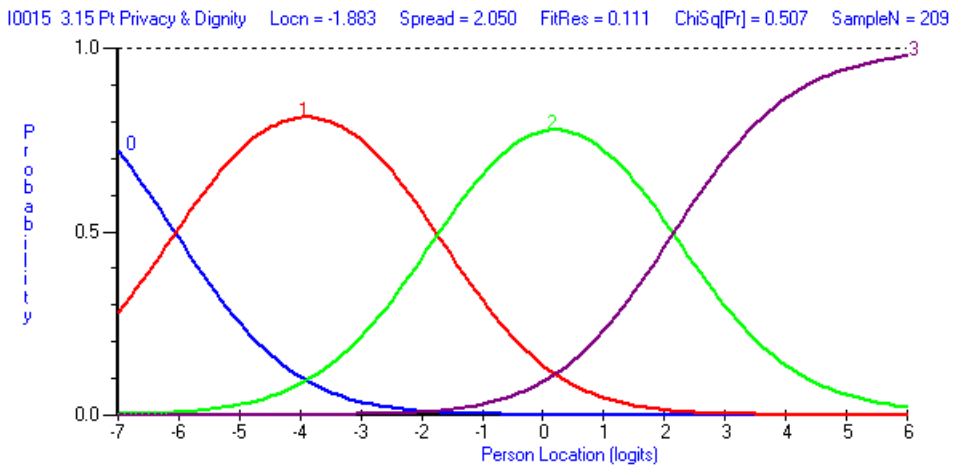


Figure 5.2 also shows that the item location for this item is shifted towards the higher end of the scale. This finding suggests that Item 35 is relatively intense (or hard) for respondents to agree with unless they hold a rather strong positive opinion about their working environment, i.e., only respondents whose mean location are located on the higher aspect of the person location continuum are likely to endorse this item. In addition, the chi square statistic (p = .001, which can be seen at the top right hand corner of the graph) suggests that the probability of this item fitting the

Rasch model is low ($p < .001$). These findings for Item 35 are explored further in the item fit statistics section.

In comparison, the CCC for Item 15 (*There is high regard for patients' privacy and dignity*) suggests that respondents at both the lower and higher person locations are likely to endorse this statement (refer to Figure 5.3). While the CCC thresholds for Item 15 also occur in the expected order, the intersection of the response category thresholds between 0 & 1and 1 & 2 are positioned to the left of the central person location (designated by 0). The CCC for Item 15 therefore suggests that it has a low level of intensity, and was easy for even respondents with a lower opinion about their working context to endorse. The chi square statistic of 0.507, however, indicates that the fit of this item to the model is good. In light of this finding, Item 15 will also be revisited in the section on item fit statistics.

Figure 5.3. *Category Characteristic Curve for Item 15*



Overall Test of Fit Statistics

The data were then examined using the test-of-fit summary statistics as presented in Table 5.3. If the data are in accord with the model then the mean of the item residuals should be close to 0 and the standard deviation close to 1. In the current study the RUMM2030 software calculated the CAI mean and standard deviation of the item fit residuals as -0.601 and 1.103 respectively. Thus, despite the small sample, the standard deviation of the data set is considered by the Rasch model to be close to the theoretical value (Andrich, et al., 2010).

Table 5.3

Summary of Item Test-of-Fit Statistics (All Data)

	Items	
	Location	Fit Residual
Mean	0.000	-0.601
SD	0.778	1.103
	Item Trait Interaction	
Total Item Chi Squ	108.377	
Total Chi Squ Prob	0.005708	

Individual item and person tests of fit

In addition to the mean fit statistics, there are a number of ways of examining the fit between the data and the model. These include three tests of fit of the items (item-trait interaction test of fit for each item, the log residual test of fit and graphs of the item characteristic curves), and one test of fit of the persons. No single fit statistic is sufficient to make a decision about the performance of an item or person. Rather, decisions are based on a combination of the above pieces of evidence. It is important to note that more than one method of fit analysis is always necessary when evaluating the fit of the data to the model.

The log residual test of fit provides information on the fit of the data to the model from the perspective of, first, the items and, second, the persons. The log residual test of fit statistic for the item tests the fit between the actual pattern of responses to an item across all persons and the pattern expected according to the model. The item-trait chi square indicates if the items are ordered from easy to medium to hard with good agreement by the persons about this item difficulty order. Each item‟s location is also noted because low discrimination can sometimes result from a low intensity item that most people endorse. A researcher may still wish to retain such an item, however, to maintain the instrument‟s validity, especially if it is to be used with a sample not as likely to endorse the items (Irene Styles, personal communication, December 9, 2010).The log residual test of fit for the persons does something similar in comparing patterns of responses obtained for each person across all items, with the expected or theoretical pattern. Like the item fit residuals, if the data are in accord with the model then the means of the person fit residuals should also be close to 0 and the standard deviation close to 1.

The log residual and item-trait test of fit statistics for each of the individual items are presented in Table 5.4. Data of particular note in the table are highlighted in boldface, with the implications for the CAI to be discussed in this section. Item 35 (*The organisation is non-hierarchical*) has a fit residual of 2.547 which is just outside of the acceptable range of -2.5 to 2.5. The chi square test of fit for this item shows that the probability that the responses for this item fit the model is low (p = .001). This statistical finding can be further explored using the third test of fit, the graphical inspection of the ICC.

Table 5.4
Summary of Item Fit Statistics

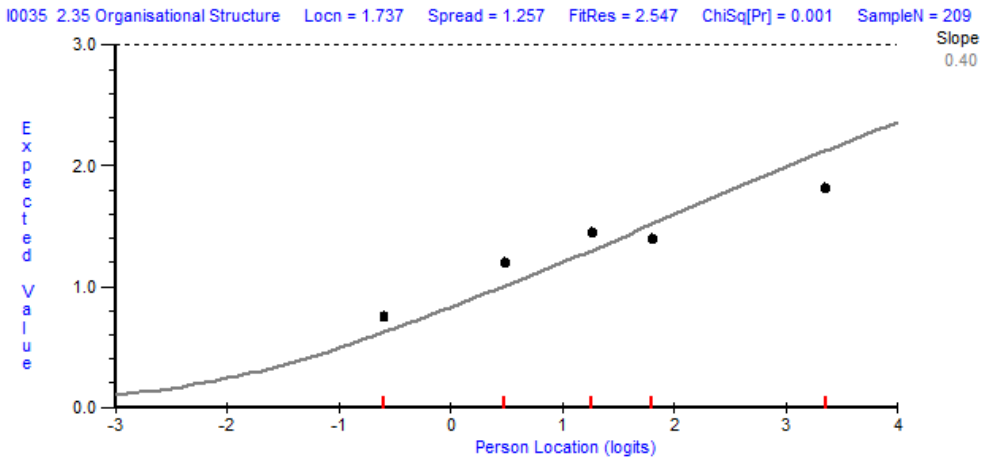
Item	Location	Fit Residual	Chi Square	Probability
1	-0.443	-0.494	11.655	.020
2	1.106	1.28	5.760	.218
3	-0.652	-0.934	1.559	.816
4	-0.394	-0.291	1.651	.800
5	-0.957	-1.332	4.676	.322
6	0.082	-1.577	3.036	.552
7	-0.425	0.229	9.015	.061
8	-0.495	-0.715	10.614	.031
9	1.481	-0.099	1.297	.862
10	0.837	0.689	14.477	**.006**
11	-0.162	-1.173	1.843	.765
12	-0.067	0.958	12.402	.015
13	-0.295	-0.386	3.213	.523
14	-0.338	-1.176	1.919	.751
15	-1.883	0.111	3.311	.507
16	-0.321	-1.835	5.636	.228
17	0.902	0.073	1.732	.785
18	0.961	1.835	10.015	.040
19	0.23	-1.525	8.151	.086
20	-1.517	0.437	2.929	.570
21	0.076	-2.263	0.527	.971
22	-0.171	-0.678	1.872	.759
23	-0.132	-0.398	1.568	.815
24	0.687	-0.843	4.054	.399
25	1.215	-1.388	5.341	.254
26	-1.008	-0.355	11.262	.024
27	-0.436	-1.21	0.841	.933
28	0.415	-0.563	2.879	.578
29	0.217	-0.914	3.029	.553
30	0.472	-1.409	4.041	.401
31	0.088	-2.398	3.901	.420
32	-0.356	-1.253	3.060	.548
33	0.569	-2.393	11.966	.018
34	-0.347	-0.922	3.410	.492
35	1.737	**2.547**	17.723	**.001**
36	-0.129	-1.526	5.069	.280
37	-0.546	-0.329	5.589	.232

Note. Fit residuals > 2.5 and p-values < 0.01 are in boldface.

The ICC is a central concept of the Rasch measurement model. It represents the probability of an item being endorsed for each possible location of a person on the continuum (Bond & Fox, 2007). Persons are ordered by the estimates of their locations on the latent continuum and classified into class intervals of between 40-50 people. This allows for the graphical inspection of the accordance of observations (mean obtained values for each group) with the model. The black dots which appear around or on the ICC plot represent the actual or observed proportions of persons within the class intervals for which the outcome is observed.

The overall pattern of responses to Item 35 (*The organisation is non-hierarchical*) demonstrate a relative lack of conformity of the data with the model because the dot points representing the location of each class intervals are situated either above or below the ICC (refer to Figure 5.4).

Figure 5.4. *Item Characteristic Curve for Item 35*

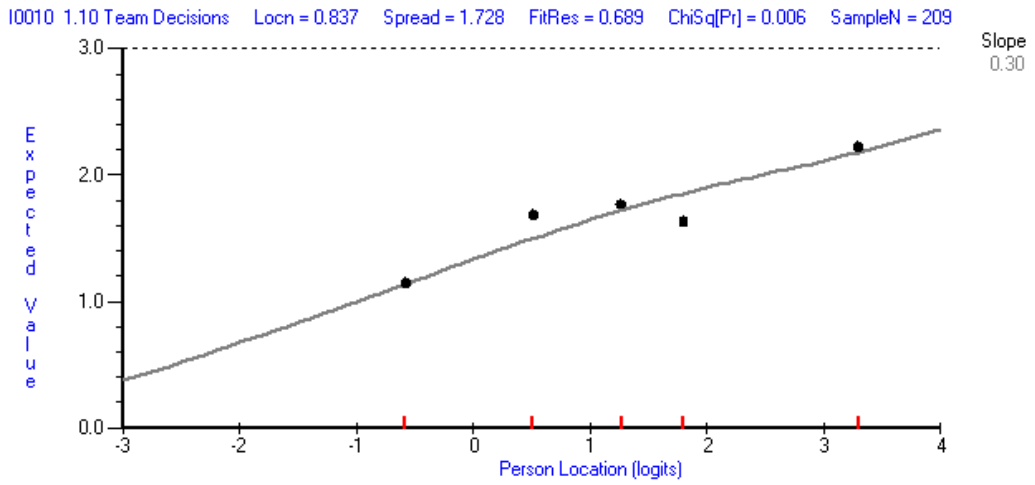


The ICC shows that the item is not discriminating as expected among groups of persons at different locations across the entire range of person locations. While the ICC itself generally increases with person locations, it is somewhat flat overall which

suggests that Item 35 discriminates less well than the other CAI items. For example, the respondents in the first 3 class intervals tend to over-endorse the statement, while the respondents in the last 2 class intervals tend to under-endorse the statement when compared to the expected curve represented by the ICC. The findings suggest that this item is not contributing to the person measures as well as other items.

Although Item 10 (*Healthcare practitioners in the multidisciplinary team have equal authority in decision making*) has an acceptable fit residual of 0.689, the low chi square test of fit indicates a low probability (p = .006) that the overall pattern of responses for this item fits the model. The ICC for Item 10 (refer to Figure 5.5) shows that although the expected values increased with increasing person locations, there was very little, if any, discrimination among the middle groups.

Figure 5.5. *Item Characteristic Curve for Item 10*



The overall responses to this item demonstrate a lack of conformity of the data with the model with the dot points for the second and fourth class intervals situated either above or below the ICC. This means that the respondents in the

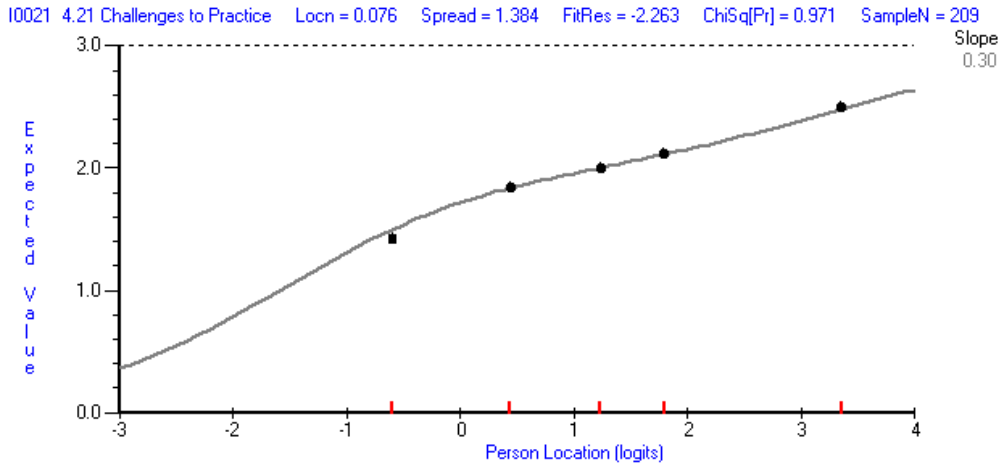
second class interval tended to over-endorse the statement, while the respondents in the fourth class interval tended to under-endorse the statement when compared to the expected responses represented by the ICC. This finding does not explain why the item is not discriminating properly for the respondents in the second and fourth class intervals only that it is happening. The conclusion, therefore, is that this item is not contributing to the person measures as well as other items because it does not discriminate sufficiently among many of the respondents in the sample.

The next step is to explore the overall impact of removing the least well fitting items (35 and 10), according to either one or two tests of fit. First, all responses for Item 35 are removed from the sample data as it is considered the item of least fit to the model. The overall tests of fit summary statistics are then re-calculated. The mean item fit residual increased from -0.601 to -0.649 with the standard deviation moving from 1.103 to 0.999. The Person Separation Index was still considered *Excellent* as it experienced only a marginal change from 0.944 to 0.942. The individual fit statistics for Item 10 were then re-visited. While the item location moved slightly from 0.837 to 0.893, the probability of Item 10 now fitting the model had notably improved from $p = .006$ to $p = .081$. The fit statistics for the other items remained acceptable.

The difference between the relative value of items such as 35 and 10 to the measurement of context by the CAI can also be illustrated by comparing their ICCs with other items that performed differently. For example, Item 21 (*Challenges to practice are supported and encouraged by nurse leaders and nurse managers*) had a Fit Residual of -2.263 suggesting that this item is somewhat over-discriminating. This may mean that something in addition to the property or latent trait of context is being assessed by this item. The chi square test of fit for this item was very

satisfactory (*p*= .971*)*. The ICC for Item 21 was checked and was observed to demonstrate an acceptable degree of conformity with the model. While the ICC"s curve wobbles a bit because the thresholds are not evenly spaced, the class interval dot points appear either very close to or on the ICC (refer to Figure 5.6).

Figure 5.6. *Item Characteristic Curve for Item 21*



This process of checking the log residuals, chi squares and ICC graphs was repeated for all items. At the end of this process, it was surmised that the inclusion of items 35 and 10 in the CAI may not add to the overall measurement of the latent trait (context) as these items demonstrate some lack of discrimination between person locations at different parts of the range of person locations. On the basis that the observed values for both items generally increased across the range of person locations and that only one item (Item 35) showed misfit across all three test of item fit (that is, less than 5% of all items), a decision was made to retain these items in future tests of the CAI so that there is an opportunity to re-visit how they operate during use in a larger test population. If this result were to be repeated over a larger

sample, an argument could be developed to remove them from revised editions of the measurement instrument.

Person Fit Statistics

In comparison to the item fit statistics, the overall mean and standard deviation of the person fit residuals for the CAI data were somewhat further from the ideal values of 0 and 1 at -0.765 and 2.203 respectively. The individual person fit residuals are useful in understanding the response profile of each participant.

The individual person fit statistics confirmed that one participant was notably different from the rest of the sample. The RUMM2030 program warned that the estimated person location of 7.806 for Participant 3 was an extrapolated value, and the log residual test of fit statistic was not calculated by the software because the pattern of responses for that person was considered to be extreme. The respondent had a total raw score of 110.81 out of a possible 111. The statistic indicates a too perfect response pattern thus rendering the data possibly unreliable. Examination of the responses indicated that this person chose the *strongly agree* response for all of the answered CAI items. This raises the possibility of a response set that quickly meets the request to complete the measurement instrument rather than providing considered responses. The responses from this participant were therefore removed from the data set for the purpose of further analysis.

After removing this outlying data, the test-of-fit summary statistics were re-calculated. The mean of the person location moved slightly closer to zero from 0.128 to 0.125, and the standard deviation reduced from 1.469 to 1.401. The power of test of fit statistic was still considered *Excellent* and the PSI decreased only slightly from 0.944 to 0.942.

Differential Item Functioning (DIF)

Although the overall fit of the CAI to the Rasch model is good, there is another aspect of the validity of measurement that needs to be considered. Fair and valid measurement requires that the results obtained from an instrument have the same meaning between and within all samples (Reise & Waller, 2009). For this to occur, items designed to measure a variable must function in the same way for all participants, and not reflect qualitative differences among different populations such as different cultures, sexes or ages. To be confident that this is the case when the CAI is used to measure the strengths of a particular clinical context in Western Australia, it is necessary to determine if all respondents, regardless of age, nursing experience or cultural background, are likely to interpret the item statements in the same way.

Differential item functioning (DIF) occurs when people from different sub-groups (commonly gender or ethnicity) but with the same level of the property of interest (ability/skill) have a different probability of giving a certain response to an item. This suggests that the item is functioning in an unexpected and different manner for a particular sub-group. To reflect this study‟s purpose of exploring the role of ethnicity in interpreting the items on the Context Assessment Index, DIF analysis was limited to the demographic variables that reflect the language currently spoken in the domestic setting by the participant, the geographical region where they were born, and where they obtained their first nursing qualification. The analysis was drawn from a modified data set that did not include the responses for Item 35 or from Participant 3. This decision reflected the lack of fit with the model that was identified in earlier tests.

The CAI items were all examined for DIF against each of the demographic variables of interest by using the RUMM2030 software to determine if each of the

remaining 36 items had the same meaning for sub-groups of the entire sample (N = 209). The probability that each item measured the same aspect of the underlying trait for all cultural sub-groups is examined. An item displays DIF when the item characteristic curves (ICCs) are distinctly different for different sub-groups of respondents. Statistically significant differences among the sub-groups are firstly identified using two-way ANOVA F statistics and probability (with Bonferroni adjustment), as these tests traditionally represent the level of variability among groups. The Bonferroni adjustment is necessary as multiple significance tests are being made simultaneously (and while a p value of .05 may be appropriate for an individual test, this may not be the case when examining a set of significance tests). Thus each test has a criterion of significance at the level of $p = .05$ which is then divided by the number of tests conducted (Field, 2005).

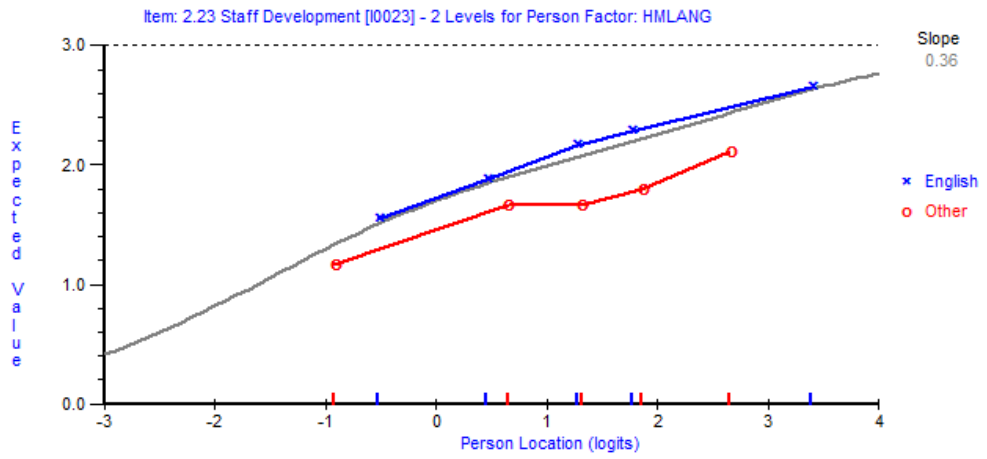
A display of each ICC with the set of observed means for each sub-group is then plotted across each of the class intervals. As described previously, these class intervals are ordered from low to high across the whole range of person estimates. The separation among the groups at each of these points is critical in the assessment of possible DIF. While lines that are parallel to the ICC and to each other indicate the possible existence of uniform DIF, lines that cross each other suggest a possible interaction effect and the presence of a non-uniform DIF structure.

The RUMM2030 program was used to randomly select equal numbers of participants that belonged to the different sub-groups for each of the cultural demographics of interest. As the resultant number of participants in each of the sub-groups was quite small ($n < 40$), 3 class intervals of approximately 10-15 respondents were selected. This process addressed the potential for problems in data

analysis that could arise when one sub-group was up to five times larger than the others.

For example, when the DIF analysis was performed for sub-groups of unequal size, Item 23 was identified as exhibiting possible DIF. One-way ANOVA statistics indicated a significant difference, $F(1) = 15.54$, $p = .000115$, Bonferroni = .000463 based on differences in the language spoken in the domestic setting (refer to Appendix E). The ICCs plotted for each of the sub-groups within the category of interest was then examined. Figure 5.7 represents the ICC for Item 23 based on the sub-group of home language. The blue line represents the persons who spoke English in the domestic setting (n=163) and the red line represents the persons who did not (n=36). The *x* (on the blue line) and *o* (on the red line) denotes the class interval mean for each sub-group. Notably the DIF appeared relatively uniform among the two sub-groups.

Figure 5.7. *Differential Item Functioning for Item 23 and Home Language*



Thus when based on sub-groups of unequal size, the language that was spoken in the domestic setting was found to influence the consistency of

interpretation of this item. The members of the sub-group who did not speak English in the domestic setting appeared to endorse the item much less than expected based on their total response to the items being used to measure context. In other words, persons who spoke a language other than English in their domestic setting, who overall have the exact same total score on the CAI, were observed to endorse this item too little when compared to the rest of the group. Thus it could have been concluded that rather than assessing an aspect of staff development practice within the clinical setting, this item was more likely measuring something else for persons who speak a language other than English in their domestic setting.

However, when using the RUMM2030 software to select equal sub-groups for Home Language, no DIF was identified for Item 23 ($F[1] = 8.218$, $p = .005608$, Bonferroni = .000463) or any of the other CAI items (refer to Appendix F). Although this finding cannot be considered robust due to the small numbers within each sub-group, the process of establishing DIF has been outlined for illustrative purposes. If a similar finding was obtained in a larger sample, the mean scores of the different sub-groups for each item could be compared with increased confidence. This process will be described later in this chapter when the CAI findings are presented for the sample.

Reliability

The Person Separation Index (PSI) indicates the extent to which the measurement instrument is able to discriminate among persons. The PSI represents the degree to which the ordering of the persons would be replicated if the same sample were given another set of parallel items measuring the same construct, i.e., Person A would still demonstrate a higher level of the measured trait than Person B and the difference in relative intensity would be similar. This requires that the set of items target a range of abilities with enough spread of ability across the sample to

demonstrate a graduation of intensity on the trait being measured. A PSI statistic close to 1.0 indicates that use of the instrument will result in some persons who score higher and some lower as the set of items can discriminate adequately among the persons in the sample. It also indicates that some consistency could be expected if the sample were re-tested using similar items (Bond & Fox, 2007). In the current study, the mean and standard deviation for the person locations of the CAI sample were 1.282 logits and 1.469, respectively.

The item reliability index indicates the degree to which the order of the item locations on the continuum would remain constant among different sample groups, i.e., item X would still be easier for most respondents to endorse when compared with item Y. The higher the item reliability index, the more likely it is that a range of items exist with some that are easier to endorse and some that are harder, and that some consistency exists if the same items were used across different groups (Bond & Fox, 2007).

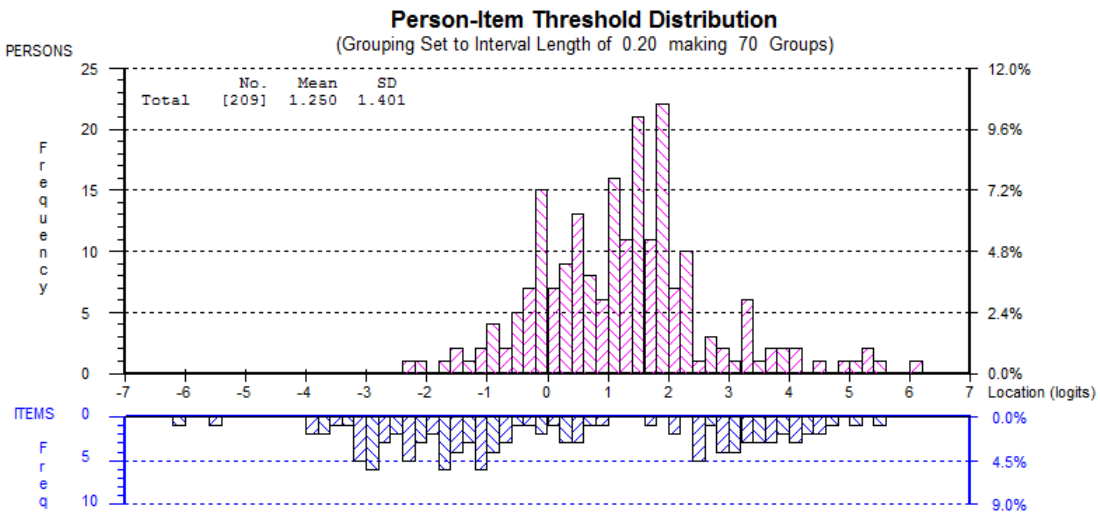
The PSI depends in part on the variance of the persons and the items. It is useful when considering if a range of item difficulty level exists within a measurement instrument, and it has an important role in understanding the fit statistics in the Rasch model. This is because it is the equivalent of the Cronbach"s α, i.e., it is the PSI that provides information about the reliability of the measure (Bond & Fox, 2007). Usually the higher the value, the less likely all items share the same difficulty level and the greater the likelihood that the ordering of the items will be similar on re-use in a different sample. If this value is close to 1.0, the model indicates that all persons tend to be spread across a continuum meaning that the items can discriminate well among persons. In the current study, the Power of Test of Fit for the CAI items with a PSI of 0.944 was considered *Excellent* by the RUMM2030

program. Thus the set of items in the CAI can be considered to provide a reliable measure for person locations across samples.

Targeting of the Persons and Items

The Rasch model allows for the determination of whether there are enough items spread along the continuum and enough spread of ability among the persons. The analysis of data thus includes an examination of the relative distribution of item threshold and person locations on the same continuum. Figure 5.8 provides a bar graph of the item and sample person locations.

Figure 5.8. *Person-Item Threshold Distribution Graph*



As stated before, the mean of the items is 0 logits. The person mean is 1.282 which indicates that it is relatively "easy" for this group of participants to agree with the items on the CAI. As the item statements are phrased in the positive, this finding suggests that the overall working environment of the sample is viewed somewhat favourably by the respondents.

The variance of the person locations in Figure 5.8 are illustrated in red (*SD* = 1.469) and are considerably greater than the variance of item locations illustrated in blue (*SD* = 0.778); however, the range of item locations is a little larger than that of the person locations. The important thing to note is the relative location of the persons and the items. The graph shows that nearly all of the respondents are targeted by the items, which suggests that the instrument provides an appropriate range of questions for a sample of persons of this kind. The gap between the item locations in the middle of the graph suggests that an opportunity exists to create additional items that are targeted to the majority of persons (Bond & Fox, 2007). The other point of interest is that two items are located to the far left on the numerical continuum indicating that they were very easy for all respondents to endorse. The process for identifying these two items is discussed in the next section.

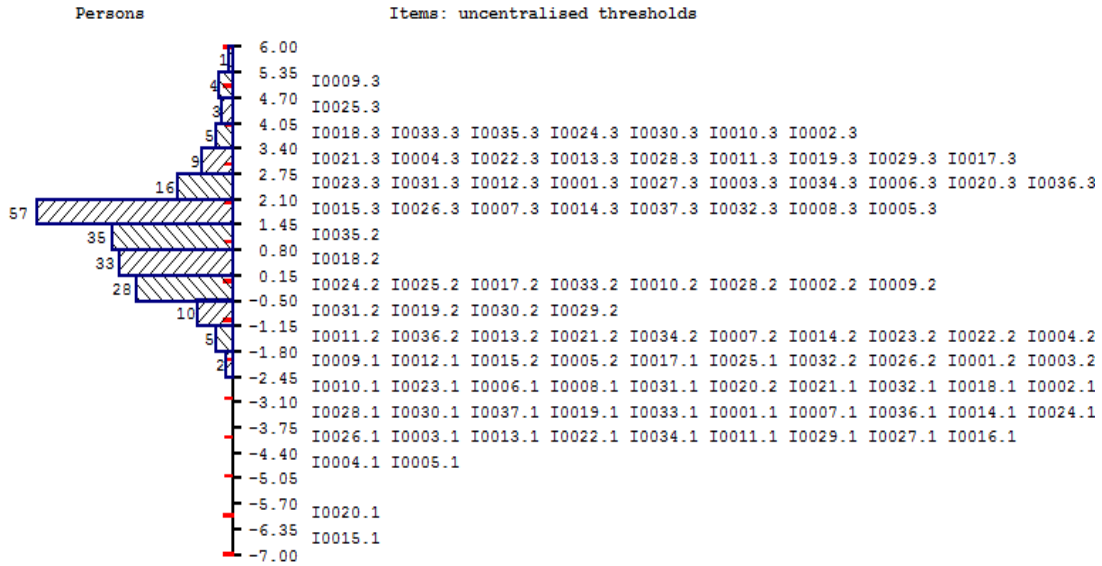
Item Threshold Map and Item Locations

The Item Map is an alternative way to examine the distribution of items and persons. The threshold locations for each item are indicated by the appropriate suffix appended to the item label. As each item in the instrument has four categories (*strongly agree*, *agree*, *disagree* and *strongly disagree*), there are three thresholds per item designated by „1", „2", and „3" respectively. For example, I0009.3 is the location of the third threshold for Item 9 (refer to Figure 5.12).

From the Item Threshold Map, the expected responses for persons plotted at different locations on the continuum can be surmised. While the Item Threshold Map displays the same information as the Person-Item Threshold Distribution, it has the advantage in displaying the location and label of item thresholds. It is therefore possible to identify that the first thresholds for Item 15 (*There is high regard for patient's privacy and dignity*) and Item 20 (*Care is based on a comprehensive*

assessment) were the two thresholds located to the far left on the Person-Item Threshold Distribution. This indicates that these items were the ones that persons with low total scores on the CAI were the most likely to endorse.

Figure 5.9. *Item Threshold Map*



This can be confirmed by examining the mean locations for all items. As mentioned in the background to the Rasch model, the locations of the items along the numerical continuum always have a mean of 0.0 logits. This is the arbitrary constraint imposed on the location parameters of the items in the RUMM2030 program because only the relative locations of the items, and not their absolute locations, can be estimated. The locations of the items are comparisons and independent of the person locations. This is a key feature of the Rasch model (Bond & Fox, 2007). In the current study, the standard deviation of the items around this mean is 0.778 logits. The locations of the individual items are provided in Table 5.5. This table presents the mean locations for each item in order of increasing intensity,

i.e., the location of each item is ordered from easiest (with the lowest location mean) for respondents to endorse to the hardest (with the highest location mean).

Table 5.5
Location of CAI items.

Item	Location	Item Statement	Item Sub-Scale
15	-1.883	There is high regard for patients privacy and dignity	Respect for Persons
20	-1.517	Care is based on a comprehensive assessment	Respect for Persons
26	-1.008	Staff welcome and accept cultural diversity	Respect for Persons
5	-0.957	The nurse leader acts as a role model of good practice	Respect for Persons
3	-0.652	A proactive approach to care is taken	Collaborative Practice
37	-0.546	Structured programmes of education are available to all HCPs	Evidenced informed Practice
8	-0.495	There are good working relations between clinical and non-clinical staff	Respect for Persons
1	-0.443	Personal and professional boundaries between HCPs are maintained	Practice Boundaries
27	-0.436	Evidence-based knowledge on care is available to staff	Evidence informed Practice
7	-0.425	Education is a priority	Evidence informed Practice
4	-0.394	All aspects of care/treatment are based on evidence of best practice	Evidence informed Practice
32	-0.356	Guidelines and protocols based on evidence of best practice are available	Evidence informed Practice
34	-0.347	Resources are available to provide evidence-based care	Evidence informed Practice
14	-0.338	Patients are encouraged to be active participants in their own care	Collaborative Practice
16	-0.321	HCPs and heath care support workers understand each others' role	Practice Boundaries
13	-0.295	Staff have explicit understanding of own attitudes/beliefs to provision of care	Practice Boundaries
22	-0.171	Discussions are planned between HCPs and patients	Collaborative Practice
11	-0.162	Audit and/or research findings are used to develop practice	Evidence informed Practice
23	-0.132	The development of staff expertise is viewed as a priority by nurse leaders	Evidence informed Practice
36	-0.129	HCPs share common goals and objectives about patient care	Respect for Persons
12	-0.067	A staff review process is in place that enables reflection on practice	Evaluation
21	0.076	Challenges to practice are supported and encouraged by nurse leaders	Practice Boundaries
6	0.082	HCPs provide opportunities for patients to participate in care decisions	Collaborative Practice
31	0.088	Clinical nurse leaders create environ conducive to development/sharing ideas	Collaborative Practice
29	0.217	HCPs have the opportunity to consult with specialists	Evidence informed Practice
19	0.230	HCPs and patients work as partners, providing individual patient care	Collaborative Practice
28	0.415	Patients have choice in assessing, planning and evaluating their care and Rx	Collaborative Practice
30	0.472	HCPs feel empowered to develop practice	Practice Boundaries
33	0.569	Patients are encouraged to participate in feedback on care, culture and systems	Collaborative Practice
24	0.687	Staff use reflective processes to evaluate and develop practice	Evaluation
10	0.837	HCPs in the MDT have equal authority in decision making	Collaborative Practice
17	0.902	The management structure is democratic and inclusive	Evidence informed Practice
18	0.961	Appropriate information is accessible to patients	Evaluation
2	1.106	Decisions on care and management are clearly documented by all staff	Respect for Persons
25	1.215	Organisational management has high regard for staff autonomy	Practice Boundaries
9	1.481	Staff receive feedback on the outcome of complaints	Evaluation
35	1.737	The organisation is non-hierarchical	Evidence informed Practice

Note. HCPs = Health Care Professionals; Rx = Treatment; MDT = Multidisciplinary Team.

Persons with low overall locations on the continuum are likely to endorse the items with the lowest locations, such as items 15, 20, 26 and 5. Notably, all of these

items were grouped by McCormack et.al (2009) into the *Respect for Persons* sub-scale. It is likely that only persons at the higher locations on the continuum will routinely endorse the more intense items such as 2, 25, 9 and 35 which are at the other end of the scale. Unlike the easiest items, each of these items belonged to a different sub-scale in the original study.

It could be argued that, in principle, the first few items on an instrument that tests attainment or ability should be the easiest for a respondent to endorse and the last items the hardest. While it has not been established if this is also the case with attitudinal scales such as the CAI, it is worth noting that the data from the current study would suggest that many of the items in sub-scale *Respect for Persons* and *Evidenced informed Practice* were easier for the participants to endorse than those items in sub-scales *Collaborative Practice*, *Practice Boundaries* and *Evaluation*.

The calculation of residual correlations indicates that only two sets of items are dependent on each other. The first set are items 28 and 33 (r = .352) which pertained to patient involvement in care processes (*Patients have choice in assessing, planning and evaluating their care and treatment* and *Patients are encouraged to participate in feedback on care, culture and systems* respectively). The second are items 32 and 34 (r = .319) which pertained to the provision of evidence-based material (*Guidelines and protocols based on evidence of best practice are available* and *Resources are available to provide evidence-based care* respectively). No systematic dependencies were identified.

Unidimensionality

As has been explained previously, unidimensionality, an underlying principle of the Rasch measurement model, is the requirement that an instrument should

include only items that measure the same construct. However, constructs can be regarded as being at different levels of scale. For instance, in a mathematics test, some items might assess algebra, other geometry and yet others, measurement. At one level of scale, the whole test is assessing mathematical achievement, but at lower levels of scale, it is comprised of sub-scales measuring its three components. Hence the aim in Rasch analysis is to evaluate whether or not a particular set of items is unidimensional at a particular level of scale. What that level of scale is will depend on the purpose of the instrument.

The Rasch analyses of the study data indicate that all 37 items are measuring a single variable, that is, the CAI is a unidimensional instrument. Thus it can be surmised that when grouped together the items on the CAI measure respondents‟ attitudes towards the context of the clinical work environment at a particular hospital. According to McCormack et al. (2009), however, the overall measurement scale also consists of five subscales, with each measuring a defined element of context. Rasch analysis can be used to establish whether this is the case or not.

A factor analysis of the current study‟s data suggests that the items actually contribute to the measurement of only one construct rather than the five factors to which the items were assigned during the development of the CAI. This was explored further using the RUMM2030 software. Firstly, only responses with complete data sets (N = 166) were selected so that Cronbach‟s α could be calculated. This resulted in a Person Separation Index (which is the RUMM equivalent of Cronbach‟s α) of 0.95. Then each of the CAI items was grouped into the factors assigned by McCormack et al. (2009), with each factor forming an item. The PSI was then re-calculated. Theoretically, if there is a notable change in the PSI, it indicates the presence of sub-scales in the CAI. However, if the outcome is similar, then it

supports the earlier premise that the items measure one common variable, which in this instance is postulated to be the context of the clinical working environment.

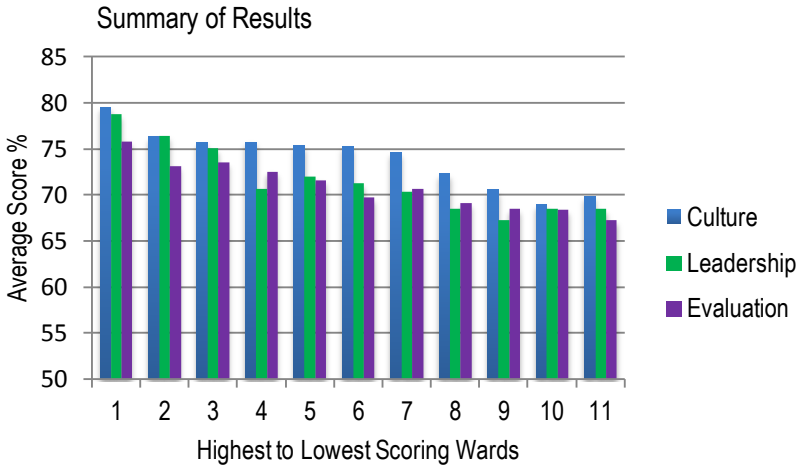
The power of the test of fit statistic after this process is still considered *Excellent* and the PSI decreases only slightly from 0.948 to 0.941. This result is supported by a minimal change in Cronbach"s α from 0.95 to 0.91. Thus the RUMM2030 analysis confirms the CTT result obtained earlier in this study – that the items comprising the CAI are measuring a single variable at this level of scale, or in CTT terminology, a single factor. If these participants are representative of Western Australian public hospital nurses, then larger scale studies may also find that the CAI measures one rather than five factors from the CTT perspective, or that the CAI represents one general level of scale rather than five lower level scales from the Rasch perspective.

CAI Findings for Sample

Although not available in the literature, the guide for use in interpreting the CAI was provided by Professor Brendan McCormack (personal communication, March 19, 2009). The guide offers a step by step process to determine what the results of the CAI data indicates when applied to the clinical setting in which it was collected (refer to Appendix G). Firstly, each of the 37 CAI items is grouped into the elements of context identified in the PARiHS framework; thus, 16 items are assigned to *culture*, 7 to *leadership* and 14 to *evaluation*. The responses given by each participant for each item are then allocated a numerical value according to the chosen Likert option: *strongly agree* = 4, *agree* = 3, *disagree* = 2 and *strongly disagree* = 1. The total scores for each context element are then multiplied by the number provided in the guide. It is unknown how this number was derived as no additional information is provided.

The participant responses from each clinical area were grouped together and the above process was followed. The combined results for the participants from each clinical area were then expressed as a percentage score for each of the three elements of the PARiHS framework. All wards scored above 65% for each of the contextual element (refer to Figure 5.10).

Figure 5.10. *Average Factor Scores by Ward and Context Elements*



The items grouped under the element of *culture* were given a more positive endorsement from the participants in each of the clinical areas than for the items grouped under the elements of *leadership* and *evaluation*. Thus, while the individual ward results for this element ranged from 69 - 80%, it appears that irrespective of the overall result for each clinical area, the individual *culture* within a ward environment is perceived by the respondents to be the strongest aspect of context at the study hospital.

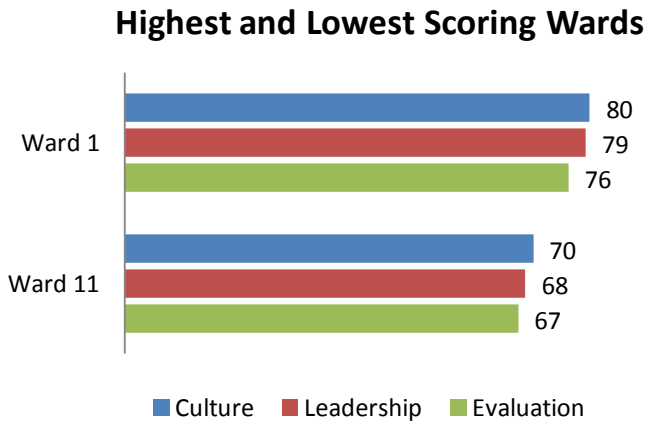
The strength of the perceived level of *leadership* demonstrated a similar level of variation (67 – 79%) among the clinical areas. Wards 1, 2 and 3 had the most

experienced Clinical Nurse Managers (all had been in the position for over ten years) and these respondents rated the *leadership* items more positively than the other areas. It was also interesting to note that at the time of data collection, Wards 10 and 11 were both being managed by Acting Clinical Nurse Managers with far less managerial experience (less than 12 months). Thus these preliminary results support research by Duffield, et al. (2007) demonstrating that consistent nursing leadership contributes to a positive working environment.

The items comprising the e*valuation* subscale obtained the lowest level of agreement among the participants in 7 of the 11 clinical areas. In the remaining 4 wards, *evaluation* was more positively endorsed than *leadership*. The reasons for this are unclear, and while it would have been interesting to explore the differences in these results with staff involved at the ward level, this work is outside the scope of the current study.

When the results for the three identified elements of context were combined for each of the clinical areas the greatest difference in the total scores was observed between Ward 1 and Ward 11 (refer to Figure 5.11).

Figure 5.11. *Comparison of CAI Scores*



There are notable variations in the composition and function of these two wards, not least of which is the patient admission process. Ward 11 is an acute surgical ward with a high level of patient turnover. Most patients are admitted via the Emergency Department after sustaining an injury and are operated on within 48 hours of admission. Ward 1 is a medical ward that receives most patients as a direct transfer from other areas of the hospital either following the immediate post-operative period or having experienced an acute episode of a chronic condition. Many of Ward 1 patients are direct admissions from Ward 11.

Apart from the difference in the type of service provided, there was also a disparity in the composition of the respondents from these two areas. The greatest variations were in the type of undergraduate qualification, length of time respondents had been part of the clinical team, and the age of the respondents themselves (refer to Table 5.6).

Table 5.6
Composition of Respondents in Wards 1 and 11

	Ward	
	1	11
Hospital based nursing diploma	80%	43%
Team member for > 4 years	60%	26%
Age > 35 years	86%	66%

While the relative influence of the ward and patient type is beyond the scope of the current study, the likelihood that differences in the demographic variables have influenced the respondents‟ perceptions of working context can be explored. It is possible to do this with some confidence as Rasch analysis of this sample determined that DIF does not exist.

The logits for the person locations of the sample data were imported into SPSS. As explained in previous chapters, logits are a more accurate description of the strength of the latent trait for each participant than the total score and help to define the relationship with other respondents in the sample group. The logits were then used as raw scores in SPSS to determine if any significant differences in response patterns existed that could be linked to differences in any of the demographic characteristics of the sample. For this sample, one-way ANOVA statistics indicated that there is a significant difference in the response patterns between those participants who had worked in Australia for less than six years and for those that had done so for six years or greater (refer to Table 5.7).

Table 5.7
ANOVA for Person Location by Time Nursing in Australia

	Sum of Squares	df	Mean Square	F	*p*
Between Groups	5.158	1	5.158	3.936	.049
Within Groups	255.539	195	1.310		
Total	260.697	196			

Thus, in the current study, the number of years each respondent has practiced as a nurse in Australia may account for most of the difference in factor scores between Ward 1 and Ward 11.

Summary of Results

Overall the results of this study indicate that the CAI functions as it was intended, that is, as a reliable and valid instrument to measure of respondents‟ attitudes towards their work context. Using classical testing techniques, the Cronbach‟s α of .95 found in this study is slightly higher than the Cronbach‟s α of .93 reported by McCormack et al. (2009). These findings attest to the reliability of the CAI. Rasch analysis confirms this result with a PSI of .944 and indicates that the

set of items are internally consistent, i.e., there is internal validity (with the possible exception of one item).

The possibility that the items can be grouped together to represent different elements of a common theme was explored. Internal consistency was examined by evaluating the scree plot and the eigenvalues from factor analysis. Only one overarching factor was identified by this process, suggesting that the items on the CAI work together to help define the underlying construct of context and cannot be legitimately separated into different sub-groups. Rather than indicating a problem with the instrument, this strengthened the likelihood that the criterion for unidimensionality has been met.

Using Rasch analysis, the current study further explored the concept of unidimensionality. RUMM2030 software provides comparative information about a measure‟s reliability by calculating the PSI and Cronbach‟s α for a complete data set (with no missing values) when the items are grouped together and when divided into factors such as those identified by McCormack et al. (2009). The findings of the PSI and Cronbach‟s α for both these data sets were nearly identical, suggesting that no identifiable subscales exist within the CAI. Therefore, while this study has confirmed that the CAI can be considered to be a unidimensional instrument, the separation of the 37 items by McCormack et al. (2009) into 5 distinct sub-scales is not supported.

In the current study, the unequal number of participants in the demographic sub-groups required the random reduction of data for the analysis of differential item functioning. While this provided an opportunity to illustrate the process involved, the overall size of the sample did not provide sufficient numbers to either establish or refute the existence of DIF.

Rasch analysis provided additional information about the interaction between the persons in the sample and the items on the CAI. The person location mean indicated that this group of participants found it relatively "easy" to agree with the CAI items. The CAI also provided an appropriate range of item statements for the sample group because Rasch analysis indicated that nearly all the participants were well targeted by the 37 items.

Further, by using the Rasch model of IRT it became possible to determine each of the CAI item"s contribution to the measurement of context within clinical environments. In the current study, the fit residual for Item 35 (*The organisation is non-hierarchical*) was considered to be extreme, and the item characteristic curve indicated that this item was not contributing as expected to the measure of the person location. Finally a high chi square statistic (and low *p* value) confirmed that the probability that the individual responses for this item fitted the Rasch measurement model was low.

Chapter Six

Discussion and Conclusion

"A theoretical discovery has but the merits of its existence, it will awaken hope and that is all. But, let it be cultivated, let it grow, and you will see what it will become."

(Louis Paster, 1854)

The present study is the first attempt to explore the construct validity of the Context Assessment Index in an Australian setting. It is also the first known attempt to examine the Context Assessment Index using a combination of statistical techniques drawn from classical test theory and item response theory. The two theoretical perspectives complement each other. The main objective of classical test theory is to explore how the items operate together at the test level, while the approach of item response theory is to investigate what each item contributes to the overall measure. Rarely used in nursing, Rasch analysis provides information about each item that is not available when measurement instruments are created using only classical test theory techniques. Importantly, the use of Rasch analysis provides a greater level of certainty that the final tool meets the property of invariance and remains unchanged regardless of the conditions in which measurement takes place. The inclusion of Rasch analysis is therefore one of the strengths of this study.

Reliability and Validity

Overall the results of this study indicate that the CAI functions as it was intended, as a single internally consistent measure of the attitude of the health care workforce towards the contextual aspects of their work environment. There is internal validity with the possible exception of one item, and the instrument has high

reliability. Thus when used together, the items can provide valid, reliable measures, of opinions about work context across a sample of people with different ethnic backgrounds and depth of nursing experience.

Differential Item Functioning

While exploring the tenets of general measurement principles, a particular focus of the current study has been on the potential influence of culture on the interpretation and meaning of the items included in the CAI. This consideration is especially important when attempting to transpose a measurement instrument from one country to another. There is always the possibility that the items on the measurement instrument are interpreted differently by individuals within different sample groups, as intrinsic characteristics such as culture, age and gender can produce different responses. This is where examination of how the items perform across different groups can be invaluable. RUMM2030 software provides a means of identifying differential item functioning to determine if the responses for any of the 37 items on the CAI are influenced by cultural differences that exist among the participants, and result in the item measuring a different variable for a specific sub-group.

In this study, no item was identified as exhibiting differential item functioning. While these findings must be considered preliminary as the analysis was based a relatively small base population, it appears that the items on the CAI may have the same meaning and thereby measuring the same variable for respondents working in different clinical settings. This is an encouraging result and justifies further testing of the CAI in a larger sample.

Unidimensionality

In this sample, all the items worked together to help define the underlying construct of context. The current study could not justify separating the items into different sub-groups. No systematic dependencies that are usually seen when discrete subgroups exist within the items were identified. Analysis of the data indicated that it does not matter which way the items are grouped, all contribute to the measurement of a single variable. Thus this study has found that the items on the CAI comprise one scale, and that the measurement instrument is unidimensional. As a result, the separation of the 37 items into 5 distinct sub-scales is not supported.

Unidimensionality is an important measurement concept, and raises the question of why factor analysis is commonly used to identify sub-scales during the development and testing of new measurement instruments. As factor analysis is based on sample dependent correlations, there is no certainty that any relationship that is identified between items in one sample will be replicated in another (Kottner & Streiner, 2010). This was clearly the case when attempting to confirm the discrete factors identified by McCormack et al. (2009) using the eigenvalues derived from the study sample. This finding supports the proponents of Rasch analysis such as Fox and Bond (2007) who would argue that the value of factor analysis in understanding how the items work together to form a linear measure is limited.

The current study considered the possibility that the division of the items by McCormack et al. (2009) into the five sub-groups could be useful when attempting to identify the relative strengths and weaknesses of context in a particular clinical setting. For example, the summary of person locations produced during Rasch analysis suggested that more items belonging to the sub-scales of *Respect for Persons* and *Evidenced Informed Practice* were positively endorsed by the

participants (that is, these items were located at the less intense end of the continuum) than those belonging to the sub-scales *Collaborative Practice*, *Practice Boundaries* and *Evaluation* (which were located at the more intense end of the continuum). It was therefore surmised that if a particular clinical setting obtained a similar result (as it should do, if the scale can be shown to be invariant across different contexts), the clinical leaders may decide to concentrate on strategies that improve teamwork within and between different health professionals, while also providing opportunities for team debriefing and reflective practice.

The problem with this line of reasoning became clear when attempting to use the sample data as initially intended, that is, to identify the strengths and weaknesses of context within a given setting. Following personal communication, the suggested process of data analysis provided by McCormack (2008) was not based on the 5 sub-scales discussed in the published literature. Instead, prior to data analysis, the 37 CAI items were intended to be grouped together under the headings of culture, leadership or evaluation. The reason being that these were the elements of context identified in the PARiHS framework as influencing the ability of a particular clinical setting to successfully adopt new evidence-based guidelines. Thus the procedure recommended by McCormack (2008) for data analysis reflects the theoretical origin of the CAI rather than the factors that were later identified during testing and refinement of the instrument.

Results for Sample

Rasch analysis allowed for the testing of what each of the items on the CAI contributed to the measurement of context within clinical environments. The gap that was identified between the item locations in the middle of the Person-Item Threshold Distribution graph suggested that there is an opportunity to create additional items if

there is a need to discriminate among persons located at this point on the continuum. As aggregate data are not available to determine if the sample was representative of the nursing staff both in Western Australia and other Australian states, larger scale studies would be needed to confirm that the CAI currently contains a range of items that are appropriate for the wider groups of interest.

In this study, the amount of information that is provided by Item 35 (*The organisation is non-hierarchical*) is questionable, as it appears to add little to the overall understanding of the latent trait for these participants. Given the relative homogeneity of the group from which the sample was obtained, this is not an unexpected result. The historical lines of management within nursing are based on the traditional hierarchies found in military and religious orders, and continue to form the foundation upon which the nursing workforce infrastructure is based. This is probably useful because hospitals have evolved into complex, dynamic organisations which are subjected to multiple demands and expectations. To meet these challenges and function efficiently, decisions and information often need to originate at the top and flow in a downward direction. Thus, if similar results are obtained in future studies based on larger samples, the removal of Item 35 from the CAI would need to be considered.

The traditional management hierarchy within tertiary hospitals can mean that it is possible for policies and procedures to be created, revised or removed without active debate by the majority of clinicians at the bedside. It is also easy to assume that once formally distributed that the specific detail of a policy or procedural change will be immediately included in the delivery of care. However, as discussed in Chapters 1 and 2, just being aware that a change in practice is recommended is not enough to ensure that it is accepted and then implemented as standard practice.

Despite working within a traditional nursing hierarchy, today‟s nurses are educated in university environments which encourage graduates to question and critically evaluate the information that surrounds them. The implementation of new or altered practice therefore requires a considered approach that takes into account the context into which those changes are to be introduced and a concerted effort to tailor strategies that are congruent with the needs of that clinical area. This is because the differences between clinical areas can either work with or against the standard strategies that are often used in a whole of hospital approach. It is therefore prudent that those involved in the implementation and monitoring of evidence-based practice take the time to pre-emptively collect area specific data and assess the contextual strengths and weaknesses of each clinical environment.

Analysis of the sample data using the guide provided by McCormack et al. (2008) supported anecdotal evidence of differences between clinical areas. For example, the study participants who obtained a higher score for the cultural and leadership items on the CAI worked in clinical areas that had previously been identified as early adopters of change by project leaders within the study hospital. A change management approach that utilises the strengths of ward culture may therefore be appropriate for those wards that scored higher on this element. This could include change management initiatives such as the “productive” ward program, which empowers the nursing team to assess the efficiency of workplace norms and immediately implement agreed changes with the goal of releasing time to care (Wilson, 2009). Coincidentally, towards the end of 2010, this strategy was implemented as a trial program in Ward 6 and Ward 11. It was interesting to note in discussion with the project facilitator five months after the Productive Ward project was introduced that Ward 11 has experienced many more challenges than Ward 6

and has not achieved the expected results in key areas. It is likely that the differences in culture that were identified by the CAI in this study could help to explain this outcome. Thus the ways in which this information can be used to shape change management practices at the ward level can be the focus of future studies now that the validity of the CAI as a measurement instrument has been established.

Thus, preliminary analysis of the study data has enabled the relative influence of leadership, culture and evaluation processes within different clinical areas to be quantified. Importantly the nature of this information allows for the observed variation in the achievement of objectives in specific change management projects to be linked to the differences in context that this study has identified. The ability of the CAI to measure those underlying differences in clinical context has reinforced the strong influence that ward culture can have on the success of change management projects while highlighting the importance of experienced leadership in providing stability and direction to nurses experiencing practice change at the patient‟s bedside.

Challenges for Implementation of the CAI in Healthcare Settings

Unfortunately, adequate warning of a change in practice and adequate time to prepare for it is becoming rare in health. Bed management practices are undergoing a revolution with many care processes being condensed into shorter time frames. Very short notice is often given of whole ward movements and reconfigurations. While the impact on morale and the working relationships between and within staff groups are sometimes recognised, the subsequent impacts on the clinical context and delivery of care are rarely measured.

For example, when the study hospital experienced several months of intense change last year, anecdotal evidence indicated that nursing staff were under

considerable pressure to increase the flow of patients throughout the hospital. As bed management processes were constantly modified to meet the expectations of external stakeholders, many more patients with numerous co-morbidities had multiple ward and bed transfers within a short period of time. Data capture and reporting during this time focussed on hospital-wide performance indicators, and not the cumulative effects on the quality of patient care.

Routine audit data collected as part of the SQuIRe Pressure Ulcer Prevention project for this period, however, identified a notable decrease in the number of patients for whom a pressure ulcer risk assessment was documented and the number of high risk patients placed on an alternating pressure relieving mattresses as a preventative strategy. At the same time, the number of severe hospital acquired pressure ulcers that were reported by clinical staff more than doubled. When discussing this observation at a regular SQuIRe meeting, the other clinical improvement projects also reported a rise in patient related incidents or near misses. Thus while the bed management changes resulted in time to patient admission targets being met, data capture for other aspects of patient care were not as positive.

Unfortunately the disparate nature of the collection and reporting of this type of data, not just in the SQuIRe program but throughout other areas in the hospital, made it difficult to definitively link the „one size fits all" approach of the bed management changes to the observed decline in the quality of care that was being delivered in the shorter time frames. It is therefore surmised that the measurement of contextual factors before, during and after key milestones in projects of this nature, could have provided a more robust understanding of potential problems in care delivery and the opportunity to implement tailored strategies that strengthened the

ability of each clinical environment to successfully meet the expectations of all stakeholders.

Study Limitations

It should be acknowledged that the study was subject to a number of limitations. Despite a response rate of 61%, the sample size (N = 210) was relatively small for Rasch analysis which as a general rule requires 10 responses per item (Streiner, 2010). This would have equated to a sample size of 370, which was greater than the total population available. For this reason, the study was always deemed to be exploratory in nature with any conclusions forming the basis of future investigations.

The voluntary nature of the participants and the convenience sampling method may have limited the representativeness of the sample. There were no constraints imposed on the completion of the CAI. It is therefore unknown if each nurse completed the CAI in isolation from the other nurses who worked in the same clinical area. It is possible that some nurses compared answers or discussed the questions with other nurses either during or after completing the CAI. Some participants may have even taken the research instrument home rather than complete it in the ward environment. It is also possible that only nurses with a stronger opinion of the work context decided to participate in the study. It was not feasible to try to control for these possibilities, and the impact on the data is unknown but unlikely to be significant.

A designated time to complete the CAI was not provided and the responses of some participants may have reflected this. Rasch analysis determined that this was the case for at least one participant who was subsequently removed from data

analysis. It is also unknown why 43 respondents did not complete at least one CAI item, although it is suspected that 5 of the respondents were not aware that the items continued on a second page. The inclusion at the bottom of the first page of an instruction to continue answering the questionnaire on the second page may have prevented this from occurring.

The CAI was designed for use in the entire population of a designated working context. Prior to introducing a change in practice, all staff within a clinical area would be encouraged to complete the CAI using a range of management and education strategies. This study only targeted nursing staff. The voluntary nature of this study meant that nurses could not be actively encouraged to complete the CAI, with the return rate for some areas being much higher than others. The preliminary results for wards in which less than 75% of the total nursing population are represented should therefore be viewed with caution.

Recommendations

This study provides some of the first information about the construct validity of the Context Assessment Index when tested outside of the original study population in Ireland. Pending larger scale studies, the current findings confirm that most of the items included on the CAI are useful in assessing the context in which nurses‟ work. Although the value of two items is not yet confirmed (23 & 35) , a larger study with a sample of at least 370 would allow for expansion of this current work, and the possibility of designing a shorter version more suitable for use by busy clinicians.

Ideally the importance of the successful introduction of new and updated best practice to both the consumer and those responsible for the fiscal outcomes would be recognised with an allocated budget in every health care facility. Instead, the current

dichotomy between how health care professionals are taught to deliver care, and the increasing pressure from external stakeholders for the organisation to focus on such things as bed management is creating conflict at the bedside. The effective delivery of nursing care in this type of environment is a growing challenge for the nursing profession. Simply providing money for limited periods of time to address narrowly defined aspects of care is not an efficient way to allocate resources, particularly if the current context is impeding the ability of the clinician to provide best practice.

While this study has demonstrated the potential value in the use of the CAI to assess contextual strengths and weaknesses prior to the introduction of a change in practice, there is often limited time in which to do so. Thus the main issue standing in the way of the use of the CAI is the current lack of commitment to not only identifying the underlying aspects of context that exist in a given environment, but an investment in both time and resources to address the issues that are identified. Recognition by senior hospital management of the importance in allocating time for the collection and analysis of CAI data, and identification of appropriate strategies would be necessary. In order to make the most of what the CAI has to offer, nursing leaders within tertiary environments therefore need to argue for the time and resources to implement this type of assessment process. Using the results to guide change management strategies would go some way to ensuring that the delivery of patient care reflects the evolving science and evidence base of nursing practice.

Conclusion

This was a small scale study of an instrument that was recently developed to measure the strengths of contextual factors within health care environments. The results provide direction for future study in this area, and may be useful in

understanding the observed differences in the uptake of evidence-based practice guidelines between clinical areas at the study hospital.

It is important to note that unlike some other instruments available in the literature, the CAI is derived from a theoretical model and reflects the contextual challenges of transferring knowledge into practice. The CAI was specifically designed to gather data that would aid in the determination of the strengths and weaknesses of contextual factors within in a particular clinical setting.

As this paper has argued, the collection of meaningful data on the relative influence of contextual factors requires the use of a well defined calibrated instrument. The current study is part of an emerging body of research that uses a variety of analytical techniques to explore the construct validity of new measurement tools. In this instance, the Rasch measurement model has been used to compliment the traditional statistical approach, as it provides a mathematical description of how measurement principles can be applied when investigating variables such as attitude. The other key advantage to this approach is the recognition that items that have been designed and tested in one setting may be interpreted differently when used in environments outside the original sample.

There is also a growing recognition in the literature that the likelihood of successfully implementing a sustainable practice change may reflect the inherent strengths and weaknesses of a given context. It is now understood that robust evidence by itself is not enough to ensure that new or updated clinical behaviours are adopted. The strong and constant influence of the interrelationship between team members and other aspects of the work environment on the way in which patient care is delivered is being acknowledged. Understanding these strength and weaknesses of

these relationships will help to tailor strategies that are appropriate for each setting, and to explain why the adoption of a change in practice is not uniform across all groups within and outside of specific organisations. Assessing contextual influences therefore has the potential to speed up and spread practice improvement by providing insight into the challenges inherent in a particular change and the strategies required for its successful implementation in any given setting.

In times of increasing patient complexity coupled with decreasing lengths of stay within tertiary hospitals, the judicious adoption and use of measurement instruments by the nursing profession is essential. Nursing research therefore has an active role to play in the evaluation of emerging measurement instruments to prevent the widespread adoption of those with limited clinical value. The combined use and understanding of statistical techniques based in classical test theory and item response theory can assist the nursing profession to achieve this objective. It is hoped that the discussion and application of this type of approach in this study will encourage the nursing profession to consider the use of these types of statistical techniques to evaluate emerging and existing measurement tools that are used to either capture the nature of nursing work or the factors that influence the provision of nursing care.

References

Abe, K., & Henly, S. J. (2010). Bullying (ijime) among Japanese hospital nurses: modeling responses to the revised Negative Acts Questionnaire. *Nursing research, 59*(2), 110-118.

Andrich, D. (1982). An Index of Person Separation in Latent Trait Theory, the Traditional KR-20 Index, and the Guttman Scale Response Pattern. *Education Research and Perspectives, 9*(1), 95-104.

Andrich, D. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. In J. Keats, R. Taft, R. Heath & S. Lovibond (Eds.), *Mathematical and Theoretical Systems* (pp. 7-16). North Holland: Elsevier Science Publishers.

Andrich, D., Sheridan, B., & Luo, G. (2010). Rasch Unidimensional Measurement Model (RUMM) 2030 [Computer software]. Perth, Western Australia: RUMM Labs.

Andrich, D., & Van Schoubroeck, L. (1989). The general health questionnaire: a psychometric analysis using latent trait theory. *Psychological Medicine, 19*, 469-485.

Australian Bureau of Statistics. (2006). Selected Health Occupations: Australia [Electronic Version]. Available at: http://www.abs.gov.au/ausstats/abs@nsf/mf/4819.0 (accessed 5 September 2010).

Baker, F. (2001). *The Basics of Item Response Theory*. [Electronic Version] Available at: http://eric.ed.gov/PDFS/ED458219.pdf (accessed 12 July 2010), ERIC Clearinghouse on Assessment and Evaluation: University ot Maryland, College Park.

Barker, A. L., Nitz, J. C., Low Choy, N. L., & Haines, T. (2009). Measuring fall risk and predicting who will fall: clinimetric properties of four fall risk assessment tools for residential aged care. *Journal of Gerontology, 64A*(8), 916-924. doi: 10.1093/gerona/glp041

Berry, J. W. (1990). Psychology of acculturation: understanding individuals moving between cultures. In R. W. Brislin (Ed.), *Applied Cross-Cultural Psychology* (pp. 232-253). Newbury Park: Sage Publications.

Berry, J. W. (2004). Psychology of group relations: cultural and social dimensions. *Aviation, space and environmental medicine, 75*(7), C52-57.

Berwick, D. M. (2003). Disseminating innovations in health care. *Journal of the American Medical Association, 289*(15), 1969-1975. doi: 10.1001/jama.289.15.1969

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: fundamental measurement in the human sciences* (2nd ed.). Mahwah: Lawrence Erlbaum Associates.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika, 71*(3), 425-440. doi: 10.1007/s11336-006-1447-6

Buchan, J., Parkin, T., & Sochalski, J. (2003). International nurse mobility: Trends and policy implications [Electronic Version]. Available at: www.who.int/entity/workforcealliance/knowledge/resources/nursesmobility (accessed 18 July 2010). Geneva: World Health Organisation.

Comfort, E. H. (2008). Reducing pressure ulcer incidence through Braden Scale risk assessment and support surface use. *Advances in Skin and Wound Care, 21*(7), 330-334. doi: 10.1097/01.ASW.0000323519.08306.ea

Commonwealth Government Department of Immigration and Citizenship. (2010). Report on Migration Program 2008-09 [Electronic Version]. Available at: http://www.immi.gov.au/media/statistics/pdf/report-on-media-program-2008-09.pdf (accessed 12 July 2010).

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

Cummings, G. G., Estabrooks, C. A., Midodzi, W. K., Wallin, L., & Hayduk, L. (2007). Influence of organisational characteristics and context on research utilization. *Nursing Research, 54*(S4), S24-S39. doi: 10.1097/01.NNR.0000280629.63654.95

Currie, L., & Loftus-Hills, A. (2002). The nursing view of clinical governance. *Nursing Standard, 16*(27), 40-44.

Currie, L., & Watterson, L. (2007). Challenges in delivering safe patient care: a commentary on a quality improvement initiative. *Journal of Nursing Management, 15*(2), 162-168. doi: 10.1111/j.1365-2834.2007.00627.x

Currie, V., Harvey, G., West, E., & Keeney, S. (2005). Relationship between quality of care, staffing levels, skill mix and nurse autonomy: literature review. *Journal of Advanced Nursing, 51*(1), 73-82. doi: 10.1111/j.1365-2648.2005.03462.x

De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education, 44*(1), 109-117. doi: 10.1111/j.1365-2923.2009.03425.x

DeVellis, R. F. (2006). Classical Test Theory. *Medical Care, 44*(11), S50-59. doi: 10.1097/01.mlr.0000245426.10853.30

Dodek, P., Cahill, N. E., & Heyland, D. K. (2010). The relationship between organisational culture and implementation of clinical practice guidelines: a narrative view. *Journal of Parenteral and Enteral Nutrition, 34*(6), 669-674. doi: 10.1177/0148607110361905

Dopson, S., FitzGerald, L., Ferlie, E., Gabbay, J., & Locock, L. (2002). No magic targets! Changing clinical practice to become more evidenced based. *Health Care Management Review, 27*(3), 35-47.

Dougherty, E. R. (2009). Translational science: epistemology and the investigative process. *Current Genomics, 10*, 102-109. doi: 10.2174/138920209787847005

Doyle, L., Brady, A.-M., & Byrne, G. (2009). An overview of mixed methods research. *Journal of Research in Nursing, 14*, 175-185. doi: 10.1177/1744987108093962

Duffield, C., Gardner, G., & Catling-Paul, C. (2008). Nursing work and the use of nursing time. *Journal of Clinical Nursing, 17*, 3269-3274. doi: 10.1111/j.1365-2702.2008.02637.x

Duffield, C., Roche, M., O'Brien-Pallas, L., Diers, D., Aisbett, C., & King, M. (2007). Glueing it together: nurses, their work environment and patient safety [Electronic version]. Available at: http://www.health.nsw.gov.au/pubs/2007/pdf/nwr_report_final.html (accessed 23 December 2010).

Dunbar, J. A., Reddy, P., Beresford, B., Ramsey, W. P., & Lord, R. S. A. (2007). In the wake of hospital inquiries: impact on staff and safety. *Medical Journal of Australia, 186*(2), 80-83.

Eccles, M., Grimshaw, J., Walker, A., Johnston, M., & Pitts, N. (2005). Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *Journal of Clinical Epidemiology, 58*(2), 107-112. doi: 10.1016/j.clinepi.2004.09.002

Eisenbeiss, S. A., van Knippenberg, D., & Boerner, S. (2008). Transformational leadership and team innovation: integrating team climate principles. *Journal of Applied Psychology, 93*(6), 1438-1446. doi: 10.1037/a0012716

Estabrooks, C. A., Squires, J. E., Cummings, G. G., Birdsell, J. M., & Norton, P. G. (2009). Development and assessment of the Alberta Context Tool. *BMC Health Services Research, 9*, 234-245.

Estabrooks, C. A., Thompson, D. S., Lovely, J. J. E., & Hofmeyer, A. (2006). A guide to knowledge translation theory. *Journal of Continuing Education in the Health Professionals, 26*(1), 25-36. doi: 10.1002/chp.48

Field, A. (2005). *Discovering Statistics Using SPSS* (Second ed.). London: SAGE Publications.

Finlayson, M. L., Peterson, E. W., Fujimoto, K. A., & Plow, M., A. (2009). Rasch validation of the falls prevention strategies survey. *Archives of Physical Medicine Rehabililitation, 90*, 2039-2046. doi: 10.1016/j.apmr.2009.07.013

FitzGerald, L., Ferlie, E., & Hawkins, C. (2003). Innovation in healthcare: how does credible evidence influence professionals? *Health and Social Care in the Community, 11*(3), 219-228.

Foy, R., Ovretveit, J., Shekelle, P. G., Pronovost, P. J., Taylor, S. L., Dy, S., et al. (2011). The role of theory in research to develop and evaluate the implementation of patient safety practices. *BMJ Quality and Safety, 20*, 453-459. doi: 10.1136/bmjqs.2010.047993

Funk, S. G., Champagne, M. T., Wiese, R. A., & Torniquist, E. M. (1991). BARRIERS: The Barriers to Research Utilisation Scale. *Applied Nursing Research, 4*, 39-45.

Gallant, C., Morin, D., St-Germain, D., & Dallaire, D. (2010). Prevention and treatment of pressure ulcers in a university hospital centre: a correlational study examining nurses' knowledge and best practice. *International Journal of Nursing Practice, 16*(2), 183-187. doi: 10.1111/j.1440-172X.2010.01828.x

Gleason, S. J., Sochalski, J., & Aiken, L. (1991). Review of magnet hospital research. *Journal of Nursing Administration, 29*(1), 9-19.

Government of Western Australia Department of Health. (December 2009). *SQuIRe2 CPI Guide Pressure Ulcer Prevention*.

Greenhalgh, T., Robert, G., MacFarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organisations: systematic review and recommendations. *The Milbank Quarterly, 82*(4), 581-629. doi: 10.1111/j.0887-378X.2004.00325.x

Grol, R., & Grimshaw, J. (2003). From best evidence to best practice: effective implementation of change in patients' care. *The Lancet, 362*(9391), 1225-1230. doi: 10.10.1016/S0140-6736(03)14546-1

Hagquist, C., Bruce, M., & Gustavsson, J. P. (2008). Using the Rasch model in nursing research: an introduction and illustrative example. *International Journal of Nursing Studies, 46*, 380-393. doi: 10.1016/j.ijnurstu.2008.10.007

Hardy, J. (2010). [SQuIRe Pressure Ulcer Prevention Project]. Unpublished raw data.

Harvey, G., Kitson, A., McCormack, B., Rycroft-Malone, J., Seers, K., & Titchen, A. (2008). PARiHS Overview [Electronic Version]. Available at: http://www.parihs.org/data/files/kt_share_fair_presentation.pdf (accessed 14 Feb 2011)

Helfrich, C. D., Damschroder, L. J., Hagedorn, H. J., Daggett, G. S., Sahay, A., Ritchie, M., et al. (2010). A critical synthesis of literature on the promoting

action on research implementation in health services (PARIHS) framework. *Implementation Science, 5*(82), 1-20. doi: 10.1186/1748-5908-5-82

Ho Yu, C. (2009). Item Response Theory: a multimedia tutorial developed by TEAR [Electronic Version]. Available at: www.creative-wisdom.com/multimedia/IRTTHA.htm (accessed 19 March 2010).

Hofstede, G. (1997). *Culture and Organisations: Software of the Mind*. Sydney: McGraw-Hill.

Hughes, C. F., & Mackay, P. (2006). Sea change: public reporting and the safety and quality of the Australian health care system. *Medical Journal of Australia, 184*(10), S44-47.

Humphries, N., Brugha, R., & McGee, H. (2008). Overseas nurse recruitment: Ireland as an illustration of the dynamic nature of nurse migration. *Health Policy, 87*(2), 264-272. doi: DOI: 10.1016/j.healthpol.2007.12.014

Jones, K. R. (2010). Rating the level, quality, and strength of the research evidence. *Journal of Nursing Care Quality, 25*(4), 304-312.

Kajermo, K. N., Bostrom, A.-M., Thompson, D. S., Hutchinson, A. M., Estabrooks, C. A., & Wallin, L. (2010). The BARRIERS scale- the barriers to research utilisation scale: a systematic review. *Implementation science, 5*(32), 1-22. doi: 10.1186/1748-5908-5-32

Kitson, A., Harvey, G., & McCormack, B. (1998). Enabling the implementation of evidence based practice: a conceptual framework. *Quality & Safety in Health Care, 7*, 149-158.

Kitson, A., Rycroft-Malone, J., Harvey, G., McCormack, B., Seers, K., & Titchen, A. (2008). Evaluating the successful implementation of evidence into practice using the PARiHS framework: theoretical and practical challenges. *Implementation science, 3*(1), 1-12. doi: 10.1186/1748-5908-3-1

Kitson, A. L. (2009). The need for systems change: reflections on knowledge translation and organizational change. *Journal of Advanced Nursing, 65*(1), 217-228.

Kottner, J., & Streiner, D. L. (2010). Internal consistency and Cronbach's alpha: a comment on Beekman et al (2010). *Internal Journal of Nursing Studies, 47*, 926-928. doi: 10.1016/j.ijnurstu.2009.12.018

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1-55.

Luo, X., Cappelleri, J. C., Cella, D., Charbonneau, C., Kim, S. T., Chen, I., et al. (2009). Using the Rasch Model to validate and enhance the interpretation of the Functional Assessment of Cancer Therapy-Kidney Symptom Index-Disease-Related Symptoms Scale. *International Society for*

Pharmacoeconomics and Outcomes Research (ISPOR), 12(4), 580-586. doi: 10.1111/j.1524-4733.2008.00473.x

Mallory, G. A. (2010). Professional nursing societies and evidence-based practice: strategies to cross the quality chasm. *Nursing Outlook, 58*(6), 279-286. doi: 10.1016/j.outlook.2010.06.005

Mantzoukas, S. (2007). A review of evidenced-based practice, nursing research and reflection: levelling the hierachy. *Journal of Clinical Nursing, 17*, 214-223. doi: 10.1111/j.1365-2702.2006.01912.x

Mantzoukas, S. (2009). The research evidence published in high impact nursing journals between 2000 and 2006: a quantitative content analysis. *Int J Nurs Stud, 46*(4), 479-489. doi: 10.1016/j.ijnurstu.2008.12.016

Marchionni, C., & Ritchie, J. (2008). Organizational factors that support the implementation of a nursing best practice guideline. *Journal of Nursing Management, 16*, 266-274. doi: 10.1111/j.1365-2834.2007.00775.x

Marvin, M. M. (1927). Research in nursing: the place of research and experimentation in improving the nursing care of the patient. *The American Journal of Nursing, 27*(5), 331-335 [*reprinted in American Journal of Nursing, October 2010*]. doi: 10.1097/01.NAJ.0000389680.64897.75

McCormack, B., Kitson, A., Harvey, G., Rycroft-Malone, J., Titchen, A., & Seers, K. (2002). Getting evidence into practice: the meaning of context. *Journal of Advanced Nursing, 38*(1), 94-104.

McCormack, B., & McCarthy, G. (2008). Development of the Context Assessment Index (CAI): University of Ulster & University of College Cork.

McCormack, B., McCarthy, G., Wright, J., Slater, P., & Coffey, A. (2009). Development and testing of the Context Assessment Index (CAI). *Worldviews on Evidence-Based Nursing, 6*(1), 27-35. doi: 10.1111/j.1741-6787.2008.00130.x

McKibbon, K. A., Lokker, C., Wilczynski, N. L., Ciliska, D., Dobbins, M., Davis, D. A., et al. (2010). A cross-sectional study of the number and frequency of terms used to refer to knowledge translation in a body of health literature in 2006: a tower of Babel? *Implementation Science, 5*(16), 1-11. doi: 10.1186/1748-5908-5-16

Meeuwesen, L., van den Brink-Muinen, A., & Hofstede, G. (2009). Can dimensions of national culture predict cross-national differences in medical communication? *Patient Education and Counseling, 75*(1), 58-66. doi: DOI: 10.1016/j.pec.2008.09.015

Parker, V., & McMillan, M. (2007). Challenges facing internationalisation of nursing practice, nurse education and nursing workforce in Australia. *Contemporary Nurse : a Journal for the Australian Nursing Profession, 24*(2), 128-136. doi: 10.5555/conu.2007.24.2.128

Ploeg, J., Davies, B., Edwards, N., Gifford, W., & Elliot Miller, P. (2009). Factors influencing best-practice guideline implementation: lessons learned from administrators, nursing staff and project leaders. *Worldviews on Evidence-based Nursing, 4*(4), 210-219.

Plsek, P., & Greenhalgh, T. (2001). Complexity science: the challenge of complexity in health care. *British Medical Journal, 323*, 625-628.

Plsek, P., & Wilson, T. (2001). Complexity science: complexity, leadership, and management in healthcare organisations. *British Medical Journal, 323*, 746-749.

Prieto, L., Alonso, J., & Lamarca, R. (2003). Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health & Quality of Life Outcomes, 1*(27), 1-13. doi: 10.1186/1477-7525-1-27

Prince, M. (2008). Measurement validity in cross cultural comparative research. *Epidemiologia e Psichiatria Sociale, 17*(3), 211-220.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability IV* (pp. 321-334). Berkeley CA: University of California Press.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology, 5*, 27-48. doi: 10.1146/annurev.clinpsy.032408.153553

Rycroft-Malone, J., Kitson, A. L., Harvey, G., McCormack, B., Seers, K., Titchen, A., et al. (2002). Ingredients for change: revisiting a conceptual framework. *Quality and Safety in Health Care, 11*, 174-180.

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't: it's about integrating individual clinical expertise and the best external evidence. *British Medical Journal, 312* (7023), 71-72.

Schultz, T. J., & Kitson, A. L. (2010). Measuring the context of care in an Australian acute care hospital: a nurse survey. *Implementation science, 5*(60), 1-11. doi: 10.1186/1748-5908-5-60

Scott, I. A., Poole, P. J., & Jayathissa, S. (2008). Improving quality and safety of hospital care: a reappraisal and an agenda for clinically relevant reform. *Internal Medicine Journal, 38*(1), 44-55. doi: 10.1111/j.1445-5994.2007.01456.x

Stevens, D. (2008). How do we know? *Quality and Safety in Health Care, 17*, 154-155.

Straus, S. E., & McAlister, F. A. (2000). Evidence-based medicine: a commentary on common criticisms. *CMAJ, 163*(7), 837-841.

Streiner, D. L. (2010). Measure for measure: new developments in measurement and item response theory. *The Canadian Journal of Psychiatry, 55*(3), 180-186.

Tansella, M., & Thornicroft, G. (2009). Implementation science: understanding the translation of evidence into practice. *The British Journal of Psychiatry, 195*, 283-285. doi: 10.1192/bjp.bp.109.065565

Tenenbaum, G. (1999). The implementation of Thurstone's and Guttman's measurement ideas in Rasch analysis. *International Journal of Sports Psychology, 30*, 3-16.

Tennant, A., & Conaghan, P. (2007). The Rasch Measurement Model in Rheumatology: what is it and why use it? When should it be applied and what should one look for in a Rasch paper? *Arthritis & Rheumatism (Arthritis Care & Research), 57*(8), 1358-1362. doi: 10.1002/art.23108

Thomas, M. L. (2010). The value of Item Response Theory in clinical assessment: a review. *Assessment*. doi: 10.1177/1073191110374797

Thompson, D. N., Wolf, G. A., & Spear, S. J. (2003). Driving improvement in patient care. *Journal of Nursing Administration, 33*(11), 585-595.

Thurstone, L. L. (1927). A law of comparative judgement. Article Series B, no.107 (pp. 273-286): The University of Chicago.

Titler, M. G. (2010). Translation science and context. *Research and Theory for Nursing Practice: an international journal, 24*(1), 35-55. doi: 10.1891/1541-6577.24.1.35

Titler, M. G., Everett, L. Q., & Adams, S. (2007). Implications for implementation science. *Nursing Research, 56*(S4), S53-S59.

Tracey, C. (2006). The glass ceiling in Irish healthcare: a nursing perspective. *Journal of Health Organization Management, 20*(6), 502-511.

van Achterberg, T., Schoonhoven, L., & Grol, R. (2008). Nursing implementation science: how evidence-based nursing requires evidence-based implementation. *Journal of Nursing Scholarship, 40*(4), 302-310.

van Alphen, A., Halfens, R., Hasman, A., & Imbos, T. (1994). Likert or Rasch? Nothing is more applicable than good theory. *Journal of Advanced Nursing, 20*, 196-201.

van de Vijer, F., & Leung, K. (1997). *Methods and Data Analysis for Cross-Cultural Research*. California: Sage Publications.

van Herck, P., Sermeus, W., Jylha, V., Michiels, D., & van den Heede, K. (2009). Using hospital administrative data to evaluate the knowledge-to-action gap in pressure ulcer preventative care. *Journal of Evaluation in Clinical Practice, 15*, 375-382. doi: 10.1111/j.1365-2753.2008.01019.x

van Nispen tot Pannerden, S. C., Candel, M. J. J. M., Zwakhalen, S. M. G., Hamers, J. P. H., Curfs, L. M. G., & Berger, M. P. F. (2009). An Item Response Theory-Based Assessment of the Pain Assessment Checklist for Seniors With Limited Ability to Communicate (PACSLAC). *The Journal of Pain, 10*(8), 844-853. doi: DOI: 10.1016/j.jpain.2009.02.007

Wallin, L. (2009). Knowledge translation and implementation research in nursing: a discussion paper. *International Journal of Nursing Studies, 46*(4), 576-587.

Waugh, R. F. (1999). Approaches to studying for higher education: A Rasch measurement model analysis. *British Journal of Educational Psychology, 69*, 63-79.

Wilson, G. (2009). Implementation of releasing time to care – the productive ward. *Journal of Nursing Management, 17*(5), 647-654. doi: 10.1111/j.1365-2834.2009.01026.x

Wilson, R. M., & Van Der Weyden, M. B. (2005). The safety of Australian healthcare: 10 years after QAHCS. *Medical Journal of Australia, 182*(6), 260-261.

Wright, B. D. (1996). Comparing Rasch measurement and factor analysis. *Structural Equation Modeling, 3*(1), 3-24.

Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal; measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation, 70*(12), 857-860.

Wright, J., McCormack, B., Coffey, A., & McCarthy, G. (2006). Developing a tool to assess person-centred continence care. *Nursing Older People, 18*(6), 23-28.

Appendices

Appendix A: Permission to implement the CAI in an Australian setting

Hardy, Joanne

From: Brendan McCormack [bg.mccormack@ulster.ac.uk]
Sent: Thursday, 19 March 2009 6:46 PM
To: Hardy, Joanne
Subject: RE: Context Assessment Index (CAI)
Attachments: CAI Final Report April 2008.pdf; CAI instrument pack FINAL.pdf

HI Joanne,
Thanks for this. I attach the final report of the research that resulted in the CAIK and also the instrument pack in case you don't have this. Hopefully this contains the information you need.
There is an article coming out in the journal *Worldviews on Evidence Based Nursing* (it is currently in early release, so should be generally available in next few weeks – visit the WVEBN website.
Re q13: this item is about how the attitudes and beliefs of staff impact on the way they engage in care practices, something that is widely recognised in the caring literature and something that emerged strongly in the focus groups. Hope that makes sense
Best Regards
Brendan

Professor Brendan McCormack
Professor of Nursing Research/Postgraduate Tutor
Institute of Nursing Research/School of Nursing
Room 17C19
University of Ulster
Shore Road
Newtownabbey
Co Antrim BT37 0QB
Tel: +44(0)2890368187
Fax: +44(0)2890368202
email: bg.mccormack@ulster.ac.uk

From: Hardy, Joanne [mailto:Joanne.Hardy@health.wa.gov.au]
Sent: 19 March 2009 00:51
To: bg.mccormack@ulster.ac.uk
Subject: Context Assessment Index (CAI)

Dear Brendan,

I am currently enrolled in a Master of Nursing by Thesis at Curtin University, and as part of this study I would like to seek your permission to implement the CAI at the tertiary hospital in which I work.

I am interested in exploring the usefulness of the tool in relation to the experiences I have had in encouraging nurses to apply evidenced based practice for the assessment and prevention of pressure ulcers.

I would like to know if it is possible to obtain any aggregated demographic details of the sample respondents (eg: age, sex, country of birth, length of time nursing) that guided refinement of the tool.

I am also interested in knowing the intent of the thirteenth item "Staff have explicit understanding of their own attitudes and beliefs toward the provision of care" as this is not clear to either myself or my supervisor.

I appreciate your time and consideration of this matter.

University of ULSTER

UCC University College Cork

CONTEXT ASSESSMENT INDEX (C.A.I)

The Context Assessment Index (CAI)

For each of the following statements, please put a cross in one box only.
A – Strongly agree; A – Agree; D – Disagree; SD – Strongly disagree

HCP= Healthcare professionals

		SA	A	D	SD
01	Personal and professional boundaries between HCPs are maintained	☐	☐	☐	☐
02	Decisions on care and management are clearly documented by all staff	☐	☐	☐	☐
03	A proactive approach to care is taken	☐	☐	☐	☐
04	All aspects of care/treatment are based on evidence of best practice	☐	☐	☐	☐
05	The nurse leader acts as a role model of good practice	☐	☐	☐	☐
06	HCPs provide opportunities for patients to participate in decisions about their own care	☐	☐	☐	☐
07	Education is a priority	☐	☐	☐	☐
08	There are good working relations between clinical and non-clinical staff	☐	☐	☐	☐
09	Staff receive feedback on the outcomes of complaints	☐	☐	☐	☐
10	HCPs in the MDT have equal authority in decision making	☐	☐	☐	☐
11	Audit and/or research findings are used to develop practice	☐	☐	☐	☐
12	A staff performance review process is in place which enables reflection on practice, goal setting and is regularly reviewed	☐	☐	☐	☐
13	Staff have explicit understanding of their own attitudes and beliefs towards the provision of care	☐	☐	☐	☐
14	Patients are encouraged to be active participants in their own care	☐	☐	☐	☐
15	There is high regard for patients privacy and dignity	☐	☐	☐	☐
16	HCPs and healthcare support workers understand each others role	☐	☐	☐	☐
17	The management structure is democratic and inclusive	☐	☐	☐	☐
18	Appropriate information (large written print, tapes, etc) is accessible to patients	☐	☐	☐	☐
19	HCPs and patients work as partners providing individual patient care	☐	☐	☐	☐

20	Care is based on comprehensive assessment	☐	☐	☐	☐
21	Challenges to practice are supported and encouraged by nurse leaders and nurse managers	☐	☐	☐	☐
22	Discussions are planned between HCPs and patients	☐	☐	☐	☐
23	The development of staff expertise is viewed as a priority by nurse leaders	☐	☐	☐	☐
24	Staff use reflective processes (e.g. action learning, clinical supervision or reflective diaries) to evaluate and develop practice	☐	☐	☐	☐
25	Organisational management has high regard for staff autonomy	☐	☐	☐	☐
26	Staff welcome and accept cultural diversity	☐	☐	☐	☐
27	Evidenced-based knowledge on care is available to staff	☐	☐	☐	☐
28	Patients have choice in assessing, planning and evaluating their care and treatment	☐	☐	☐	☐
29	HCPs have the opportunity to consult with specialists	☐	☐	☐	☐
30	HCPs feel empowered to develop practice	☐	☐	☐	☐
31	Clinical nurse leaders create an environment conducive to the development and sharing of ideas	☐	☐	☐	☐
32	Guidelines and protocols based on evidence of best practice (patient experience, clinical experience, research) are available	☐	☐	☐	☐
33	Patients are encouraged to participate in feedback on care, culture and systems	☐	☐	☐	☐
34	Resources are available to provide evidence-based care	☐	☐	☐	☐
35	The organisation is non-hierarchical	☐	☐	☐	☐
36	HCPs share common goals and objectives about patient care	☐	☐	☐	☐
37	Structured programmes of education are available to all HCPs	☐	☐	☐	☐

PARTICIPANT INFORMATION SHEET October 2009

Assessing the usefulness of a recently developed questionnaire in determining how nurses view the clinical area in which they work.

My name is Joanne Hardy and I am currently completing a research study for my Master of Philosophy in Nursing at Curtin University of Technology.

You are being invited to participate in this study because you are a Registered or Enrolled Nurse working within a Medical or Surgical ward at Fremantle or Kaleeya Hospitals.

Purpose of Research
The aim of this study is to assess the usefulness of a recently developed questionnaire in determining how nurses view the clinical area in which they work. This questionnaire is called the Context Assessment Index (CAI) and usually takes a maximum of 20 minutes to complete.

Participation
Completion of the questionnaire will indicate that you agree to participate in this study. Your involvement in the study is entirely voluntary, and there is no penalty for not participating.

Confidentiality
The completed Context Assessment Index will be completely de-identified and no personal information will be provided to your employer. I will be the only person that will have access to the information that you provide. It will not be possible to link any personal information to the completed tool, and you will remain anonymous. The background information will only be used to compare the general characteristics of the total study population with other nursing groups that have used the same tool. The complete questionnaires will be kept in a locked cabinet for 5 years, before being destroyed.

Further Information
This research has been reviewed and approved by the Curtin University of Technology Human Research Ethics Committee (approval number xxxxx) as well as the South Metropolitan Area Health Service Nursing Research Committee. If you have any concerns about this study you can contact either the secretary of the Curtin University of Technology Human Research Ethics Committee on 9266 2784 or the secretary of the South Metropolitan Area Health Service Nursing Research Committee on 9431 2929.

If you would like further information about this study, please feel free to contact me on 9431 3325 or by email: joanne.hardy@health.wa.gov.au. Alternatively if you wish to discuss the study with someone who is not directly involved in this study you can contact my supervisor Paul Snider by phone on 9266 2071 or email: P.Snider@email.curtin.edu.au.

Instructions for completing the Context Assessment Index

1. Please read each of the statements.

2. Answer each question as you feel it relates to your experience. (There are no wrong answers).

3. Do not dwell too long on each statement and go with your first response to each one.

4. Select the response that best describes your view or opinion by colouring in the corresponding circle like this:

5. If you decide to change your answer, place a cross through the first answer like this:

6. Answer all the questions.

7. Place the completed form in the collection box located in the Nursing Handover room.

Thank you

	Context Assessment Index (McCormack et al 2009)	Strongly Agree	Agree	Disagree	Strongly Disagree
1.	Personal and professional boundaries between healthcare professionals are maintained	O	O	O	O
2.	Decisions on care and management are clearly documented by all staff	O	O	O	O
3.	A proactive approach to care is taken	O	O	O	O
4.	All aspects of care are based on evidence of best practice	O	O	O	O
5.	The nurse leader acts as a model of good practice	O	O	O	O
6.	Healthcare professionals provide opportunities for patients to participate in decisions about their own care	O	O	O	O
7.	Education is a priority	O	O	O	O
8.	There are good working relations between clinical and non-clinical staff	O	O	O	O
9.	Staff receive feedback on outcomes of complaints	O	O	O	O
10.	Healthcare professionals have equal authority in decision making	O	O	O	O
11.	Audit and/or research findings are used to develop practice	O	O	O	O
12.	A staff review process is in place which enables reflection on practice and goal setting and is regularly reviewed	O	O	O	O
13.	Staff have explicit understanding of their own attitudes and beliefs towards the provision of care	O	O	O	O
14.	Patients are encouraged to be active participants in their own care	O	O	O	O
15.	There is high regard for patient privacy and dignity	O	O	O	O
16.	Healthcare professionals and healthcare support workers understand their roles	O	O	O	O
17.	The management structure is democratic and inclusive	O	O	O	O
18.	Appropriate information (large written print, tapes etc) is accessible for patients	O	O	O	O
19.	Healthcare professionals and patients work as partners providing individual patient care	O	O	O	O

Context Assessment Index (McCormack et al 2009)	Strongly Agree	Agree	Disagree	Strongly Disagree
20. Care is based on comprehensive assessment	O	O	O	O
21. Challenges to practice are supported and encouraged by nurse leaders and nurse managers	O	O	O	O
22. Discussions are planned between healthcare professionals and patients	O	O	O	O
23. The development of staff expertise is viewed as a priority by nurse leaders	O	O	O	O
24. Staff use reflective processes (eg: clinical supervision, reflective diaries) to evaluate and develop practice	O	O	O	O
25. Organisational management has high regard for staff autonomy	O	O	O	O
26. Staff welcome and accept cultural diversity	O	O	O	O
27. Evidenced-based knowledge on aspects of care is available to staff	O	O	O	O
28. Patients have choice in assessment, planning and evaluating their care and treatment	O	O	O	O
29. Healthcare professionals have the opportunity to consult with specialists	O	O	O	O
30. Healthcare professionals feel empowered to develop practice	O	O	O	O
31. Clinical nurse leaders create an environment that is conducive to the development and sharing of ideas	O	O	O	O
32. Guidelines and protocols based on evidence of best practice (patient experience, clinical experience, research) are available	O	O	O	O
33. Patients are encouraged to participate in feedback on care, culture and systems	O	O	O	O
34. Resources are available to provide care	O	O	O	O
35. The organisation is non-hierarchical	O	O	O	O
36. Healthcare professionals share common goals and objectives about patient care	O	O	O	O
37. Structured programmes of education are available to all Healthcare professionals	O	O	O	O

Context Assessment Index
BACKGROUND INFORMATION

It would be appreciated if you could answer the following questions.

1. In which country did you complete your first nursing qualification?

_____

2. What type of nursing qualification did you complete?
   - O Hospital based diploma
   - O Undergraduate university diploma or degree
   - O TAFE diploma

3. How long have your worked as a nurse in Australia?
   - O less than 1 year
   - O between 1 - 3 years
   - O between 3 - 6 years
   - O longer than 6 years

4. How long have you worked as a nurse at this hospital?
   - O less than 1 year
   - O between 1 - 3 years
   - O between 3 - 6 years
   - O longer than 6 years

5. How long have you worked as a nurse in your current ward at this hospital?
   - O less than 1 year
   - O between 1 - 3 years
   - O between 3 - 6 years
   - O longer than 6 years

Context Assessment Index

BACKGROUND INFORMATION

6. What is your current nursing position?
   - O Clinical Nurse
   - O Registered Nurse
   - O Enrolled Nurse

7. Are you paid a qualification allowance?
   - O Yes
   - O No

8. How many hours do you work in a fortnight?
   - O less than 37.5 hours
   - O more than 37.5 hours, but less than 60.8 hours
   - O more than 60.8 hours

9. What is your country of birth?

   _____

10. What is the main language spoken at home?

    _____

11. How old are you?
    - O between 20 - 34 years of age
    - O between 35 - 44 years of age
    - O between 45 - 54 years of age
    - O older than 55 years of age

12. What is your gender?
    - O Male
    - O Female

Thank you for answering these questions.

Your time and contribution to this research project is appreciated.

Appendix E: DIF summary table based on home language (unequal groups)

Person Factor ANOVA for all 36 items (one item with significance level below Bonferroni probability adjustment of 0.000463 is highlighted)

Item	Class Interval				Home Language				Class Interval by Home Language				Total DIF			
	MS	F	DF	p	MS	F	DF	p	MS	F	DF	p	MS	F	DF	p
I0001	2.5	2.819	4	0.026466	0.01	0.013	1	0.910601	0.72	0.81	4	0.520381	2.89	0.65	5	0.661618
I0002	1.44	1.252	4	0.290409	0.04	0.038	1	0.8447	1.02	0.885	4	0.473936	4.12	0.716	5	0.612321
I0003	0.53	0.601	4	0.662509	1.73	1.97	1	0.162087	0.99	1.122	4	0.347483	5.68	1.292	5	0.269244
I0004	0.61	0.629	4	0.642296	0.51	0.528	1	0.468483	0.81	0.842	4	0.500173	3.75	0.779	5	0.565908
I0005	1.42	1.715	4	0.148474	1.02	1.229	1	0.268992	0.72	0.865	4	0.486096	3.88	0.938	5	0.457734
I0006	0.81	1.01	4	0.403686	0.71	0.887	1	0.347397	0.4	0.499	4	0.736255	2.31	0.577	5	0.717622
I0007	2.22	2.257	4	0.064514	2.31	2.349	1	0.127025	0.36	0.363	4	0.834934	3.73	0.76	5	0.579727
I0008	2.76	3.275	4	0.012665	3.7	4.397	1	0.037332	0.94	1.119	4	0.348692	7.48	1.775	5	0.119871
I0009	0.79	0.782	4	0.538053	0.34	0.331	1	0.565479	0.33	0.329	4	0.858133	1.67	0.33	5	0.894631
I0010	3.62	3.646	4	0.006953	4.05	4.078	1	0.044886	0.76	0.768	4	0.547142	7.1	1.43	5	0.215286
I0011	0.7	0.806	4	0.522735	0.32	0.372	1	0.542625	0.58	0.667	4	0.615901	2.63	0.608	5	0.694076
I0012	3.03	2.846	4	0.025363	3.78	3.551	1	0.061055	0.81	0.764	4	0.549756	7.04	1.322	5	0.256656
I0013	0.56	0.59	4	0.670319	0.05	0.057	1	0.812066	0.15	0.156	4	0.959924	0.64	0.136	5	0.983705
I0014	0.63	0.717	4	0.581379	0.08	0.086	1	0.769073	0.65	0.742	4	0.564214	2.69	0.611	5	0.691423
I0015	1.77	1.81	4	0.128509	0.78	0.794	1	0.374148	0.08	0.087	4	0.986509	1.11	0.228	5	0.949983
I0016	0.91	1.147	4	0.335863	0.05	0.068	1	0.794659	0.77	0.968	4	0.426037	3.12	0.788	5	0.559271
I0017	0.52	0.483	4	0.74793	0.18	0.166	1	0.684275	0.54	0.507	4	0.730993	2.34	0.438	5	0.82131
I0018	3.65	3.259	4	0.013022	1.28	1.145	1	0.286068	0.93	0.833	4	0.505778	5.01	0.895	5	0.485411
I0019	2.12	3.024	4	0.019046	4.1	5.844	1	0.016593	0.31	0.439	4	0.78006	5.34	1.52	5	0.18538
I0020	0.42	0.392	4	0.814307	0.38	0.354	1	0.55248	0.55	0.514	4	0.725182	2.57	0.482	5	0.789119
I0021	0.15	0.206	4	0.934624	0.04	0.049	1	0.82538	0.74	1.021	4	0.397911	2.99	0.826	5	0.53235
I0022	0.52	0.573	4	0.682252	2.32	2.553	1	0.111802	1.07	1.183	4	0.319744	6.61	1.457	5	0.206002
I0023	0.56	0.637	4	0.636437	13.76	15.54	1	0.000115	0.41	0.46	4	0.765033	15.39	3.476	5	0.005012
I0024	0.68	0.772	4	0.544774	0.24	0.276	1	0.599707	0.67	0.753	4	0.557434	2.91	0.657	5	0.656228
I0025	1.3	1.563	4	0.18619	0.78	0.933	1	0.335487	0.61	0.73	4	0.572583	3.21	0.77	5	0.572176
I0026	3.05	3.497	4	0.008859	0.05	0.057	1	0.811633	0.77	0.883	4	0.475399	3.13	0.718	5	0.610999
I0027	0.16	0.188	4	0.944266	0.38	0.439	1	0.508331	0.3	0.35	4	0.843825	1.59	0.368	5	0.870169
I0028	0.58	0.645	4	0.631104	2.14	2.362	1	0.126025	1.42	1.565	4	0.185424	7.81	1.725	5	0.130999
I0029	1.59	1.824	4	0.125996	0	0.003	1	0.954473	2.09	2.397	4	0.051929	8.36	1.918	5	0.09329
I0030	0.34	0.418	4	0.795668	0.04	0.047	1	0.829442	0.32	0.388	4	0.816748	1.32	0.32	5	0.900472
I0031	0.97	1.335	4	0.258569	2.94	4.057	1	0.045462	1.13	1.566	4	0.18515	7.47	2.065	5	0.071851
I0032	0.58	0.717	4	0.581306	0.09	0.111	1	0.739256	1.09	1.357	4	0.250735	4.45	1.108	5	0.35789
I0033	2.67	3.798	4	0.005443	1.15	1.629	1	0.203457	1.58	2.241	4	0.066327	7.46	2.119	5	0.065126
I0034	0.24	0.293	4	0.882265	8	9.598	1	0.002257	1.23	1.479	4	0.210149	12.93	3.103	5	0.010293
I0036	0.85	1.059	4	0.378143	3.88	4.843	1	0.029004	0.33	0.418	4	0.795804	5.22	1.303	5	0.264614
I0037	1.54	1.655	4	0.162399	5.46	5.877	1	0.016305	0.73	0.79	4	0.532791	8.4	1.808	5	0.113336

Appendix F: DIF summary table based on home language (equal groups)

Person Factor ANOVA for all 36 items using equal groups (no items with significance level below Bonferroni probability adjustment of 0.000463 exist)

Item	Class Interval				Home Language				Class Interval by Home Language				Total DIF			
	MS	F	DF	p	MS	F	DF	p	MS	F	DF	p	MS	F	DF	p
I0001	0.08	0.094	2	0.909988	0.34	0.389	1	0.534807	2.49	2.808	2	0.067549	5.32	2.002	3	0.122255
I0002	1.77	1.41	2	0.25133	1.28	1.024	1	0.315333	2.13	1.696	2	0.191376	5.54	1.472	3	0.230298
I0003	1.36	1.356	2	0.264673	5.14	5.131	1	0.026792	0.9	0.897	2	0.412738	6.94	2.308	3	0.084518
I0004	0.1	0.122	2	0.885756	0.35	0.411	1	0.523592	0.21	0.252	2	0.777853	0.78	0.305	3	0.821521
I0005	0.05	0.046	2	0.954722	0.58	0.578	1	0.44989	1.35	1.36	2	0.263795	3.28	1.099	3	0.355812
I0006	0.25	0.295	2	0.74553	1.03	1.228	1	0.271942	0.01	0.01	2	0.990156	1.05	0.416	3	0.742214
I0007	2.09	2.249	2	0.113541	3.71	3.998	1	0.049665	2	2.157	2	0.12376	7.71	2.771	3	0.0484
I0008	1.46	1.486	2	0.23387	2.03	2.068	1	0.155248	0.69	0.7	2	0.50021	3.41	1.156	3	0.33338
I0009	0.18	0.17	2	0.843946	1.7	1.575	1	0.214066	1.09	1.009	2	0.370136	3.87	1.198	3	0.31772
I0010	0.05	0.05	2	0.951705	2.98	2.774	1	0.100829	0.33	0.31	2	0.734925	3.65	1.131	3	0.343468
I0011	1.5	1.893	2	0.158943	0.01	0.016	1	0.898549	0.57	0.716	2	0.492695	1.14	0.483	3	0.695522
I0012	1.71	1.196	2	0.309018	0.4	0.278	1	0.600093	0.4	0.281	2	0.755895	1.2	0.28	3	0.839735
I0013	5.95	6.191	2	0.003509	0	0.001	1	0.980705	2.33	2.426	2	0.096563	4.67	1.618	3	0.194116
I0014	0.02	0.018	2	0.982411	0.31	0.297	1	0.587714	1.33	1.276	2	0.285947	2.96	0.95	3	0.421829
I0015	1.12	0.955	2	0.390153	0.24	0.206	1	0.651283	0.21	0.18	2	0.835802	0.66	0.189	3	0.903787
I0016	0.78	0.867	2	0.42502	0.08	0.091	1	0.763642	0.99	1.106	2	0.337079	2.07	0.767	3	0.516359
I0017	0.23	0.236	2	0.790105	0.31	0.321	1	0.572879	0.35	0.354	2	0.702977	1	0.343	3	0.794021
I0018	1.23	1.467	2	0.238046	1.63	1.954	1	0.166917	0.39	0.461	2	0.632797	2.4	0.959	3	0.417732
I0019	0.48	0.598	2	0.5527	4.7	5.898	1	0.017976	0.1	0.129	2	0.879309	4.9	2.052	3	0.115423
I0020	1.75	1.584	2	0.213055	0.07	0.064	1	0.801445	0.52	0.473	2	0.625573	1.12	0.336	3	0.799136
I0021	1.73	2.693	2	0.075465	0.02	0.037	1	0.848821	0.95	1.476	2	0.236294	1.92	0.996	3	0.400453
I0022	0.87	0.944	2	0.394374	3.62	3.909	1	0.052334	0.68	0.729	2	0.486395	4.97	1.789	3	0.158159
I0023	1.64	2.052	2	0.136905	6.59	8.218	1	0.005608	1.34	1.676	2	0.195187	9.27	3.857	3	0.013358
I0024	0.55	0.611	2	0.545966	1.13	1.251	1	0.267818	0.59	0.651	2	0.524896	2.3	0.851	3	0.471373
I0025	0.48	0.546	2	0.582091	0.07	0.08	1	0.778433	2.04	2.326	2	0.106266	4.15	1.577	3	0.204003
I0026	0.75	1.004	2	0.372103	0.51	0.688	1	0.409894	0.94	1.27	2	0.287756	2.4	1.076	3	0.365559
I0027	0.38	0.46	2	0.633627	0.44	0.538	1	0.465878	0.27	0.335	2	0.716794	0.99	0.403	3	0.751667
I0028	0.86	0.849	2	0.432589	0	0.001	1	0.970032	0.22	0.22	2	0.803205	0.45	0.147	3	0.931176
I0029	1.89	2.529	2	0.087652	0.07	0.099	1	0.754315	0.48	0.646	2	0.527502	1.04	0.464	3	0.708674
I0030	0.23	0.289	2	0.750023	0.34	0.423	1	0.518063	0.69	0.865	2	0.425943	1.72	0.718	3	0.545157
I0031	0.25	0.266	2	0.766953	1.55	1.664	1	0.20199	1.08	1.161	2	0.320048	3.72	1.328	3	0.273396
I0032	2.59	4.423	2	0.015944	0.17	0.289	1	0.59283	0.84	1.43	2	0.24705	1.84	1.049	3	0.377031
I0033	1.39	1.768	2	0.179233	0.57	0.723	1	0.398496	2.16	2.741	2	0.072451	4.88	2.068	3	0.113714
I0034	1.76	2.661	2	0.077715	2.25	3.395	1	-0.070114	0.31	-0.465	2	0.999999	1.63	0.822	3	0.486777
I0036	1.37	2.369	2	0.1019	0.45	0.781	1	0.380168	1.21	2.08	2	0.133412	2.87	1.647	3	0.187488
I0037	1.17	0.979	2	0.3812	2.99	2.51	1	0.118058	0.83	0.697	2	0.502062	4.65	1.301	3	0.281817

Appendix G: Analysis Guide for CAI data

GUIDENCE ON INTREPRETATING THE CAI

This guide provides step by step process for interpreting the results of the CAI questionnaire. A calculator would help in this process. Also enclosed are two options you may want to consider in developing an action plan following the process of interpreting the CAI

Step 1

1. Put the score for each question in the boxes provided. Score 4 for 'strongly agree' 3 for 'agree' 2 'disagree' and 1 'strongly disagree'. Add up the total score for each element. Then times the score with the number provided in each box to gain the percentage score.

Example;

Statement numbers	3	5	7	14	18	21	31	34	**Total Score**
Scores	1	3	4	2	3	2	2	4	21

Element: Culture

Statement numbers	1	3	07	9	12	15	16	18	21	23	24	28	31	33	34	36	**Total Scores X 1.5625**
Scores																	= %

Element: Leadership

Statement numbers	2	6	10	17	22	27	29	**Total Scores X 3.57**
Scores								=.....%

Element: Evaluation

Statement numbers	4	5	8	11	13	14	19	20	25	26	30	32	35	37	**Total Scores X 1.78**
Scores															=.....%

Step 2: Plotting the results

In Step 2plot your scores for each element along the appropriate continuums (lowest score to the highest score you can achieve for each element)

Plotting scores along the continuum from weak to strong will indicate the existing context of the clinical area that is being assessed. Refer to the description of each element provided in the user guide for further information.

Culture

0% ←——————————|——————————→ **100%**

Leadership

0% ←——————————|——————————→ **100%**

Evaluation

0% ←——————————|——————————→ **100%**

Step 3: Overall individual score

The next step is to determine the overall score in order to identify the characteristics of the existing context that enhance or hinder person-centred care and the receptiveness of the clinical area to change.

Add together the scores for each element above and divide by 3

Total % Score =

OVERALL CONTEXT

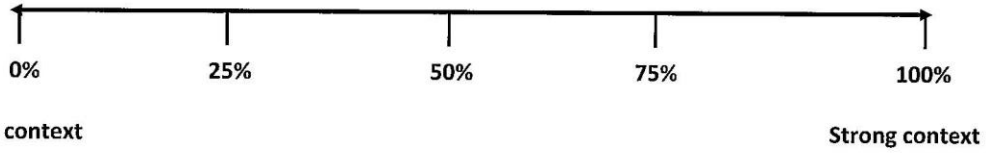
0% **25%** **50%** **75%** **100%**

Weak context **Strong context**

Step 4: Total team score

Add together all the scores from each person and divide by the number of practitioners who completed the CAI to gain the overall team context. You can do this for all three elements if you wish

OVERALL CONTEXT



IDENTIFYING AREAS FOR DEVELOPMENT

Option 1: Identifying key areas of development

This option enables the team to identify areas they scored lowest on and thus are areas for development.

Step 1: Place all the scores in the table over the page. On the left is an option to put a number or the initial of the person completing the CAI

Example

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
J	4	3	2	4	3	3	3	3	3	3	3	1	4	2	1	4	3	3	3	3	2	2	3	3	3	3	3	4	1	2	3	3	3	3	3	3	3
S	3	3	2	3	3	4	4	2	2	1	4	2	4	2	2	3	3	3	3	3	1	1	4	4	4	4	4	3	4	2	4	4	4	2	3	3	4
T	2	4	1	3	4	4	3	3	4	2	3	1	3	3	2	3	3	4	4	4	2	1	4	2	2	4	4	3	2	3	3	3	3	3	3	4	4
J	3	2	2	2	3	4	4	2	3	1	4	1	3	1	2	2	4	4	4	4	2	1	3	4	3	3	3	3	1	4	3	3	2	3	3	4	3

Step 2
Note the questions that the team has consistently scored low. For this team the questions were;

Questions 3, 12, 14, 15, 21, 22, 29,

03 A proactive approach to care is taken
12 A staff performance review process is in place which enables reflection on practice, goal setting and is regularly reviewed
14 Patients are encouraged to be active participants in their own care
15 There is high regard for patient's privacy and dignity
21 Challenges to practice are supported and encouraged by nurse leaders and nurse managers
29 HCPs have the opportunity to consult with specialists

Step 3
This has illustrated the contextual areas the team agrees are weak. The team can now use the information to develop an action plan that focuses on addressing areas of weakness.

Identifying key areas of development

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37

Now list questions that the team has scored lowest

Option 2: Reflective questions These questions aim to enable you/the team to reflect on a way forward in developing an action plan from learning about the context of care. You can use these questions to built on option 1 or move directly from the scoring to this option.

Reflective questions	Elements	Weak characteristics	Strong characteristics	Reflective questions

What is the team's consensus on the important beliefs and values they hold about person centred care?
 	
Are their opportunities for the team to reflect on their practice?
 	
What is the team's understanding of their work based culture?
 	
What do you/team see as the barriers to change in the area?
 	
How does the team gain the views and thoughts of patients about practice in the area?
 	
How does the team ensure the views of staff and patients on patient care and wider organisation are integrated into practice development plans?
 	
Have other tools been utilised (e.g. observation of practice, patient stories), in your areas as a means to understand present practice?
	Culture	• Lack of clarity around boundaries
		• Unclear values and beliefs
		• Low regard for individuals
		• Task driven organisation
		• Lack of consistency
		• Not receptive to change
			• Boundaries clearly defined (physical, social, cultural and structural)
			• Able to define culture(s) in terms of prevailing values/beliefs
			• Values individual staff and clients
			• Promotes learning organisation
			• Consistency of individuals role/experience to value;
			  ○ Relationship with others
			  ○ Team working
			  ○ Power and authority
			  ○ Rewards/recognition
			• Receptiveness to change
				How does the team/you disseminate their achievements with other areas?
				
				Does the team/you reflect on the key factors which have enabled a strong culture to develop and focus on developing these?
				
				Are the values and beliefs the team hold about their practice documented and share with others?
				
				Is there a practice development plan for the area?
				
				What processes have been used to develop a strong culture?

Is the team familiar with the theory of transformational leadership? *Does the team understand the*	**Leadership**	• Traditional, command and control leadership • Lack of role clarity • Lack of teamwork • Didactic approaches to	• Transformational leadership • Role clarity • Effective teamwork • Enabling/empowering approach	*Does the team understand the term transformational leadership?* *Have the team explored in*

impact of different leadership styles on practice? *Do team members engage in clinical supervision as a team or individually?* *Have the clinical leaders ever undertaken 360 degree feedback with support of their supervisor/manager?* *Do the registered nurses have clarity about their roles?*		teaching/learning/managing • Autocratic decision making processes • Lack of appropriateness and transparency • Lack of power and authority	to teaching/learning/managing • Enabling/empowering approach to learning/teaching/managing • Appropriate and transparent decision making processes • Power and authority understood	*action learning/clinical supervision leadership strengths and weakness to gain individual/team insight into the impact of their leadership on practice* *Is the practice of the area shared with other clinical areas?* *Is there clear succession planning of clinical leaders taking place in the area?*
Is feedback gained about practice from patients such as, patient stories, questionnaires or a patient forum etc? *How is data used to inform practice development e.g. audits of practice, length of stay etc?.* *Are staff and patient forums used for the feedback of information from audits, complaints etc?*	**Evaluation**	• Absence of any form of feedback and information • Narrow use of performance information sources • Evaluations rely on single rather than multiple methods • Poor organisational structure like what? How do you judge?	• Feedback on individual, team and systems • Use of multiple sources of information on performance • Use of multiple methods, clinical, performance and experience. • Effective organisational structure	*Have findings from own area been compared with other areas and evidence of best practice* *Have successes been documented and disseminated?* *Are patient's forums used to ensure the voice of the patient is heard?*