DOCTOR OF PHILOSOPHY

Dementia Knowledge - Psychometric Evaluation in Healthcare Staff and Students

Gamble, Clair Marie

Award date:
2022

Awarding institution:
University of Dundee

Dementia Knowledge - psychometric evaluation in healthcare staff and students

Clair Marie Gamble

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

University of Dundee

February 2022

Table of Contents

List of tables

List of figures

List of key/common abbreviations

AD	Alzheimer's Disease
CFA	Confirmatory Factor Analysis
CML	Conditional Maximum Likelihood
CTT	Classical Test Theory
DKAS	Dementia Knowledge Assessment Scale
EFA	Exploratory Factor Analysis
HCP	Health Care Professionals
IRT	Item Response Theory
KIDE	Knowledge in Dementia Scale
KoD	Knowledge of Dementia
MOOC	Massive Open Online Course
VaD	Vascular Dementia
WHO	World Health Organisation

Acknowledgements

To every person who has believed in me: words cannot express my gratitude.

PhD supervisors:

Professor Tim J. Croudace (principal supervisor)

Professor Judith Sixsmith

Professor Wendy Moncur

Declaration

The candidate confirms that she is the author of this thesis and that, unless otherwise stated, all references cited have been consulted by the candidate. The work of which this thesis is a record has been done by the candidate and has not been previously accepted for a higher degree. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Name: Clair M. Gamble

Signature: [redacted]

Date: 06/09/21

Abstract

Dementia is a progressive life-limiting condition. Global prevalence is increasing in line with current rates of population ageing. A number of disciplines have argued or found it useful to measure dementia knowledge levels. Challenges exist in achieving effective measurement across a range of healthcare professionals, but this is important to inform health-related education and professional development interventions. Similarly, effective measurement is required in informal care givers as well as lay populations, to improve evaluation of national public health and awareness campaigns.

The field of dementia knowledge measurement has seen rapid growth over the past decade, with a proliferation in the number of available instruments. To identify and appraise existing measurement instruments, a systematic review was undertaken. 14 instruments were identified; critical appraisal showed that the overall psychometric quality was poor, with no instrument emerging as 'gold standard'. A key finding from this review was that measurement of dementia knowledge is currently anchored by classical test theory (CTT) methods, with a notable absence of any item response theory (IRT) methods used in instrument development or evaluation. To address this gap, and to highlight the potential usefulness of IRT methods in evaluating the measurement properties of instruments, this study sought to apply IRT methods in dementia knowledge instruments, to determine what information could be generated about the measurement properties of items and item-sets using IRT methods, as opposed to CTT methods alone.

Two datasets were used in this thesis: the first comprised 521 sets of responses from healthcare professionals to the 16-item Knowledge in Dementia Scale (KIDE). The second comprised 404 sets of responses to the KIDE and the 25-item Dementia Knowledge Assessment Scale (DKAS).

This PhD thesis advances the field and contributes to knowledge by synthesising the current measurement literature and demonstrating the value and potential of strong measurement modelling under the Rasch paradigm, highlighting areas of further development but also areas of weakness within current items and instrumentation.

CHAPTER 1

BACKGROUND AND OVERVIEW

This chapter provides an introduction to dementia and the importance of raising awareness of the condition, as well as an overview of the current literature on knowledge of dementia, and current practices in dementia knowledge measurement. Following discussion on the background for this thesis, justification for the PhD project and the structure of this thesis will be outlined.

Although by no means exhaustive, the descriptions and context of dementia are outlined here to demonstrate the complexity of the condition, and how understanding, or 'knowledge', of dementia might (and indeed, must) exist on many levels. For example, from basic knowledge of symptoms and risk factors for public awareness, to knowledge of evidence-based strategies for management of dementia in healthcare workers, through to high-level knowledge of each dementia-related condition for practitioners involved in diagnosis, treatment, and clinical research.

1.1 Dementia

Dementia is the 5th leading cause of death globally (World Health Organisation (WHO), 2020) and the second highest cause of death in Scotland, only 0.2% behind ischaemic heart disease (National Records of Scotland (NRS), 2021). As such, the condition remains a global and national public health priority.

1.1.1 Causes and symptoms of dementia

Dementia is a syndrome caused by pathophysiological processes in the brain, which lead to progressive impairment of cognitive functioning. There are many illnesses, injuries, and conditions which contribute to, or result in dementia; for example, the most common form of dementia, *Alzheimer's disease* (AD), is caused by a build-up of abnormal proteins around the brain cells, resulting in damage to internal cell structures. Another common form, *vascular dementia* (VaD), is caused by hypoxic

ischaemia (cell death) due to a deprivation of oxygen to the brain. VaD can be caused by a narrowing or blockage of blood vessels in the brain, or by a stroke or multiple small strokes (Alzheimer's Disease International (ADI), 2019a). Other known causes of dementia include *dementia with Lewy bodies*, where small abnormal structures (Lewy bodies) form in the brain; and *frontotemporal dementias*, where clumps of abnormal proteins form inside nerve cells in the frontal and/or temporal lobes of the brain. A*lcohol-related brain damage* is caused by excessive and prolonged alcohol use and is responsible for approximately 10% of *young-onset dementia* diagnoses, where symptoms occur before the age of 65 (Dementia UK, 2021). This list is not comprehensive, as there are currently thought to be more than 100 types of dementia, with each presenting as a unique but often overlapping set of symptoms with varying trajectories of progressive functional impairment (ADI, 2020).

Although there are agreed medical classifications for the different types of dementia, the boundaries between each are often blurred, leading to difficulties in specific identification of type. For example, it is not uncommon for someone to have both AD and VaD, with these types accounting for up to 90% of all cases of dementia (ADI, 2019a). Diagnosis, management, and care can be impeded by confusion between dementia symptoms and symptoms of common or comorbid illness such as depression (Alzheimer's Society, 2021a). It is generally easier to distinguish between different forms of dementia in the early stages of the condition, though unfortunately, symptoms during this stage can be overlooked completely due to the often-gradual onset of dementia, and in occasional cases where social and emotional deterioration precede any noticeable cognitive impairment.

The majority of dementias cause gradual decline in a progressive manner, and as such dementia is widely categorised into three stages: early (mild), middle (moderate) and late (severe) stage (Alzheimer's Society, 2021b). Symptoms include gradual deterioration of memory, communication, and thinking abilities, but also changes in behaviour, mood, and difficulty in performing everyday tasks such as dressing or preparing meals (WHO, 2020). A more comprehensive overview of symptoms can be found in *Table 1.1*. Four dementias have been represented in this table to illustrate the variability of symptoms in the early stages, before symptom presentation merges in the middle and late stages, rendering identification of type

more problematic. The **text in bold** represents the primary symptom by which dementia-type is likely to be identified. Knowledge and understanding of these symptoms is likely to be a requirement for healthcare staff involved in the management, treatment, and diagnosis of dementia-related conditions. For the sources of information used to collate this table, see references: Alzheimer's Research UK (2018), Alzheimer's Society (2021b) & WHO (2020).

At present, dementia is classed as a terminal illness; there is no cure and conclusive evidence on modifiable risk factors such as vascular disease, hypertension, depression, and lifestyle factors is limited (ADI, 2019a). As a result, a growing body of research exists in the area of brain health and dementia prevention, with innovative initiatives in the UK leading this field forward; an example of this narrative can be found in the Scottish Dementia Research Consortium (SDRC) Impact report (SDRC, 2019).

Table 1.1. Symptoms across the stages of four types of dementia (AD, VaD, DLB & FTD).

	Alzheimer's Disease (AD)	Vascular dementia (VaD)	Dementia with Lewy bodies (DLB)	Frontotemporal dementia (FTD)
Early (mild) stage	• **Loss of memory of recent events** • Difficulty retaining new information • Unable to find words • Difficulty in making decisions • Poor judgement • Depression, distress and anxiety	• **Difficulty in planning and following logical steps** • Poor concentration • Slower thought processes • Episodes of heightened emotion • Apathy • Mood swings	• **Fluctuating alertness and attention span** • Detailed visual hallucinations • Auditory hallucinations • Problems with movement – slow, stiff gait	• **Uncharacteristic behaviour such as selfishness and/or rudeness** • Repetitive or ritualised behaviour such as tapping or returning repeatedly to a specific location • Loss of inhibition
		All dementias		
Middle (moderate) stage	• **Changes in behaviour** • Needing reminders and/or assistance to perform activities of daily living such as washing, dressing and eating • Increasing forgetfulness, including the names of loved ones (especially in AD) • Disorientation to place and time – increasingly likely to get lost • Misperceptions – suspicion, sometimes delusions and/or hallucinations (more common in DLB) • Anger and aggression • Social inappropriateness • Particular risk of falls (more pronounced in DLB)			

Late (severe) stage	• **Pronounced loss of memory** • **Loss of ability to perform activities of daily living such as washing, dressing, and eating** • Increasing weight loss • Difficulty in eating and swallowing • Incontinence • Episodes of aggression, particularly during personal care • Progressive loss of speech • Eventually, complete dependence

1.1.2 The scale and impact of dementia

In terms of global burden, dementia conditions are significant. Currently there are an estimated 50 million people worldwide living with dementia. It has been predicted that this estimate will more than triple by 2050 (WHO, 2020). In the UK, 850 000 people live with the condition (Alzheimer Society, 2021a), with more than 90,000 of these people living in Scotland (Alzheimer Scotland, 2021).

The economic impact of dementia is profound, with the World Health Organisation (WHO) stating that 1.4% (and rising) of gross domestic product (GDP) in high income countries is spent on these diseases (WHO, 2020). The most recent estimates of the cost of dementia in the UK range from £26 billion (Lewis *et al*., 2014) to £37.4 billion per year (Alzheimer's Society, 2021c), with almost £12 billion of this being paid by people with dementia and their families through the provision of informal care (Prince *et al*., 2016).

A major contributor to this scale of economic impact is the costs associated with secondary care, as a high percentage of people living with dementia have comorbid conditions which can lead to frequent hospital admissions (ADI, 2019a) and therefore the associated high costs. It is estimated that at least 25% of hospital beds in the UK are occupied by people with dementia; in addition, these patients often stay on the wards up to five times longer than other patients over the age of 65 (Alzheimer Society, 2016). However, current numbers are likely to be underreported as recent estimates report that only 53-73% of people with dementia in the UK have been formally diagnosed (Alzheimer's Research UK, 2021).

Aside from the economic impact, dementia is known to have profound impact on quality of life with its features influencing the emotional, social, and psychological welfare of both the person living with the disease as well as their loved ones (WHO, 2020), some of whom will be involved in roles as carers. As such, dementia has been identified as a global public health priority by the World Health Organisation and has been a national priority in National Dementia Strategies throughout the UK for over a decade (Department of Health, 2015; Scottish Government, 2017).

1.1.3 The need to raise awareness of dementia

In response to the growing urgency to prioritise dementia, the need for awareness raising initiatives across populations is critical to:

- the timely seeking of advice where potential symptoms are present
- the creation and uptake of person-centred services
- provision of the most effective evidence-based diagnostic and post-diagnostic support
- the generation of research funds to identify any disease-modifying treatments

(ADI, 2019b)

Aside from those who live with dementia or provide any form of care or treatment, the same issues of knowledge as a form of health-literacy are relevant to the general public, as people living with dementia are becoming increasingly prevalent in our ageing population and initiatives are in place to advocate for community inclusiveness and engagement in decision making (Scottish Government, 2017).

The need to raise awareness of dementia has been a consistent feature of dementia strategies globally and remains a key mission for global and national organisations such as the World Health Organisation, Alzheimer's Disease International, Alzheimer's Society, and Alzheimer Scotland. Common misconceptions remain deep-rooted and may in turn fuel the stigma which remains pervasive in society. An example from a recent global survey from Alzheimer's Disease International is that two out of three people believe that dementia is a normal part of ageing, and that between 25% - 67% (estimates are wide due to regional variations) of people would conceal their diagnosis of dementia when meeting people (ADI, 2019a).

Since this awareness or 'knowledge' of dementia is a primary focus of this thesis, the following section provides operational definitions of 'knowledge' and 'knowledge of dementia', for clarity.

1.2 Operational definitions of 'knowledge' and 'knowledge of dementia'

A widely recognised definition of *knowledge* is: 'the theoretical or practical understanding of a subject; facts, information and skills acquired through experience

or education' (Oxford University Press, 2021). As such, for this PhD research, '*knowledge of dementia'* (henceforth 'KoD') has been operationalised as: 'the theoretical or practical understanding of dementia, acquired through experience or education', for this project.

Varying levels of knowledge exist across population groups, for example healthcare professionals would be expected to have completed some formal training, whereas members of the public may have awareness based only on personal experience and/or media coverage. In theory, the more knowledge one has about the condition, the more likely they may be to seek help, encourage others to seek help, or recognise and understand clinical symptoms and behaviours (ADI, 2019a).

1.3 Knowledge and awareness of dementia across groups

The following sections will discuss the current context of dementia knowledge and the potential implications of poor levels of knowledge across population groups.

1.3.1 Generation of dementia knowledge datasets

A sound empirical understanding of where we stand as a population regarding KoD holds importance in the design and development of knowledge and awareness campaigns, as well as policy-making. The generation of evidence (such as datasets) relating to dementia knowledge across populations has proliferated in the past decade, likely as a result of dementia's status as a global public health priority. In the UK, national social surveys regularly contain sections on dementia knowledge and attitudes, with many having a focus on risk factors and recognition of symptoms (NatCen, 2015; Devine, 2016; NatCen, 2017).

Research studies have published results in relation to instrument development and attempts to improve dementia knowledge nationally and globally. For example, a massive open online course (MOOC) entitled 'Understanding Dementia' (Eccleston *et al*., 2019) sought to improve dementia knowledge across populations of adult learners who may have limited prior education. An outcome of this intervention was the generation of a large dataset (n=4984) containing responses to a validated

dementia knowledge instrument. Instrument development studies traditionally generate datasets of participant responses during the pilot phase. The available studies and instruments in relation to dementia knowledge are discussed later, as part of the literature review in Chapter 3.

A wide scope of item content with regard to dementia knowledge has been assessed across various population groups. Biomedical, clinical and psychosocial aspects of care of people with dementia are commonly broken down into domains, which might include: pathology; diagnosis and assessment; signs and symptoms; risk factors; prevalence; treatment; care management (Annear *et al*., 2017, Cahill *et al*., 2015).

Knowledge levels have been found to vary throughout different populations and deficits have been identified across several domains (Cahill *et al*., 2015, ADI, 2019a, Eccleston *et al*., 2019). In recent years, assessment of the British publics' KoD has taken place in the form of large-scale social surveys, including the British Social Attitudes (BSA) to Dementia Survey (Marcinkiewicz *et al*., 2015) which assessed knowledge and attitudes in 2,167 British households. The findings of this survey indicated that the majority of people in Britain know of someone with dementia, and have an awareness of the condition, but that a significant gap in knowledge exists regarding dementia risk factors. Further, the findings suggest that negative attitudes and stigma around the condition remain prevalent in our society (Marcinkiewicz *et al*., 2015).

Other social surveys include The Northern Ireland Life and Times (NILT) survey, which contained questions on knowledge and attitudes on dementia in 2010 and 2014, for approximately 1,200 participants. Also in 2014, the Scottish Social Attitudes (SSA) survey included questions about dementia during 1,501 face-to-face interviews, with the intention of using the data to inform future policy. Finally, The Healthy Ireland Survey in 2014/5 included questions about dementia in 7,539 individual interviews. The findings of the latter three surveys have been compared together in a very useful discussion paper by Devine (2016). Combined, these findings echo the BSA survey findings in that the majority of people in Scotland, Northern Ireland and the Republic of Ireland report knowing a person living with dementia, and also the majority of respondents had some knowledge of the condition, but overall ability to identify risk factors was poor. Key messages to arise from this

discussion paper include recommendations for a public health campaign that highlights the facts on dementia, and a further campaign directed at young people with the intention of reducing dementia prevalence in the future (Devine, 2016).

However, despite the volume of available research results, accurate comparison of dementia knowledge levels across groups remains problematic due to the lack of use of validated measurement instruments. Again, this issue has been discussed in detail later in the thesis, in Chapter 3.

1.3.2 Synthesis of dementia knowledge data

To the author's knowledge, and at the time of writing, there is currently only one systematic review of dementia knowledge studies, and one to synthesise instrument development articles.

A systematic review (Cahill *et al*., 2015) of the public's knowledge of AD and dementia offered a synthesis of the results of 40 heterogeneous studies. In 19 of the 40 studies reviewed, poor or very poor levels of public knowledge were found, but the study concluded that overall, the public had a fair to moderate knowledge of dementia, with the exception of ethnic minority populations, who were identified to have poor knowledge; Cahill *et al*. (2015) suggested that this effect may have been due to English proficiency rather than ethnicity as such, there may also have been cultural barriers in these groups of participants. It should however be noted that more than three quarters of the studies in this review developed their own data collection instruments; the review authors did not assess psychometric quality or state any psychometric properties for these instruments, which may limit the interpretation of results.

A systematic review of dementia knowledge instruments (Spector *et al*., 2012) examined the reported psychometric properties and appropriateness of administrations of five instruments across populations including healthcare professionals, informal caregivers, and lay populations. Spector *et al*.'s (2012) search strategy was robust and data on reported psychometric strength were extracted not only from the instrument development studies but also from subsequent administrations of the measures. The review authors concluded that the existing

instruments were predominantly weak in psychometric strength, and that practical application was limited due to outdated item content and restrictive target populations.

1.3.3 Implications of poor knowledge of dementia in formal care providers

It is known that limitations in knowledge and awareness, particularly among healthcare professionals, can be detrimental to the provision of person-centred care (Mid Staffordshire NHS Foundation Trust Public Inquiry, 2013; ADI, 2019a; WHO, 2020). Amongst clinicians, poor knowledge of dementia can lead to misinterpretation of symptoms and delayed diagnosis (Annear *et al*., 2017); this does not support current political objectives that aim to increase early diagnosis rates and provide timely and effective care (DoH, 2015; Scottish Government, 2017). Further, low levels of knowledge can result in inappropriate care and therapeutic treatments (ADI, 2019a) particularly within secondary care settings, which have, at times, unfortunately fallen short in the provision of safe, dignified and effective care (Mid Staffordshire NHS Foundation Trust Public Inquiry, 2013; Dewing and Dijk, 2016; Scottish Government, 2017).

As such, it is inferred that higher levels of knowledge of dementia among healthcare professionals may increase the quality of care provided, but also facilitate social inclusion and psychosocial support (Annear *et al*., 2017, ADI, 2019a). This in turn works toward addressing the continuing stigma and negative attitudes commonly associated with dementia (Spector *et al.*, 2012). In addition, it has been suggested that higher levels of knowledge among staff may reduce work related stress (Sullivan & Mullan, 2016) because a deeper understanding of the causes of behavioural and psychological symptoms of dementia can lead to more effective person-centred care, and therefore greater sensations of reward and job satisfaction.

1.3.4 Implications of poor knowledge of dementia in the community

Early detection of dementia is crucial to the facilitation of patient-led care planning: through timely diagnosis, patients and families may have more time to make

decisions and plan their care in advance, hereby affording them choice (ADI, 2019b). Knowledge of dementia symptoms and, crucially, a widespread understanding that dementia is *not* a normal part of the ageing process are critical to the seeking of healthcare advice in those who have concerns about potential symptoms. Further, as stated above, the general public's knowledge of risk factors for dementia is somewhat limited (Cahill *et al*., 2015), therefore people living in the community may not be aware that some risk factors are in fact modifiable. Given these modifiable risk factors are very similar to those for better known conditions including coronary artery disease, stroke, and several cancers (Brown *et al*., 2018), it could be considered that communities are equipped with enough health literacy to reduce the primary risk factors for dementia. However, even in populations of healthcare professionals, knowledge of the risk factors for dementia has been found to be poorer than knowledge of other aspects such as symptoms, diagnosis, and treatment (60% vs >80%, respectively, using a validated KoD instrument) (Alacreu *et al*., 2019), potentially demonstrating the significance of the problem.

1.4 Measurement of dementia knowledge

In recent years, there has been a proliferation in the number of measurement instruments developed to assess knowledge of dementia in populations including healthcare professionals, patients in health and social services, and the general public. This proliferation was justification for the systematic review conducted for this PhD thesis, which was an update of a previous review of dementia knowledge instruments by Spector *et al*. (2012). The appearance of several new instruments may be due to the increased recognition of the need for high quality measures that are realistically complex, that is, able to assess person-centred aspects of care and treatment as opposed to only a narrow focus on biomedical aspects of dementia and related conditions.

The previous discussion has established the incentive to measure dementia knowledge across groups of health professionals, caregivers, and members of the public; to facilitate benchmarking and reliable evaluation of change. To reliably measure a latent (unobservable) construct such as knowledge, instruments must be designed and constructed in a logical, structured, and systematic manner. Guidance

on test development is available and widely cited, for example in texts by Streiner and Norman (2015) and DeVellis (2016). The use of instruments or tests that do not meet technical requirements or demonstrate psychometric quality may bias results or lead to inconsistencies in findings (Sullivan and Mullan, 2016) therefore contributing to methodologically weak studies.

A comprehensive discussion on the fundamentals of test development and the array of models for psychometric evaluation of tests and items is provided in *Chapter 2: Theoretical foundations of measurement*, however, to aid in the justification of this PhD research, an introductory outline is also provided here.

Psychometrics or psychometric theory is defined as "the theory of psychological tests and measurements" (McDonald, 2013, pg.1); traditionally used in the fields of psychological and educational measurement, but recently having gained traction in health-related research (Gomes *et al*., 2018). Psychometric theory posits that where direct measurement of a concept is not possible, as is the case for social attitudes, mood states, pain severity, or in this instance *knowledge*, all appropriate aspects of the concept(s) should be compiled to form a measurement scale, or 'test' (Polit and Yang, 2018).

In educational and health-related measurement and test development, it is generally accepted that there are two model-based frameworks of measurement theory: *classical test theory* (CTT) and *item response theory* (IRT). Under both frameworks, methods of item (question) generation are common, often beginning with a comprehensive pool of questions, or 'items' related to the construct of interest. These items may be formed using factual evidence or content from other relevant, validated tests; however, all items (existing and new) should receive consensus on their appropriateness, relevance, and interpretation from experts and target population groups, where appropriate (Terwee *et al*., 2007).

1.4.1 Classical Test Theory

The CTT model is a predominantly *test-based* framework; that is, an individual's ability or level of dementia knowledge is determined by the *overall number of items* they correctly respond to in a test. Under the CTT framework, the initial item pool is

refined based on psychometric evaluation to determine whether the test indeed measures the phenomenon of interest it intends to (establishing validity) and that it measures this construct in a consistent manner (establishing reliability). Psychometric testing should also determine whether the test is sensitive to changes between populations and over time (Streiner and Norman, 2015). This process of psychometric testing is commonly referred to as *instrument validation* and involves administering the test to samples of the target population and carrying out a series of defined tests (DeVellis, 2016).

CTT is underpinned by relatively soft assumptions about tests and the items they contain; these assumptions facilitate the applicability of CTT methods widely, and with relative ease (Streiner and Norman, 2015). There are a number of limitations with such *test-based* models: information generated by aggregated scores is entirely sample dependent, and also dependent on the number of items within a test. Further, the assumption of item equivalence does not facilitate equal distribution of measurement precision across the underlying construct (Embretson and Reise, 2013).

1.4.2 Item Response Theory

The framework of IRT covers probabilistic *item-based* models that guide researchers in the examination of relationships between an individual's ability (or level of dementia knowledge) and their responses to individual items, rather than entire tests. Test development using IRT methods is not concerned with how many items are scored correctly by an individual, but with how the relationship between item endorsement and respondent's ability, and the estimation of item parameters, can be used to facilitate greater levels of measurement precision across groups of items (Embretson and Reise, 2013). IRT models are underpinned by stronger assumptions than the CTT model, rendering it less widely applicable yet able to generate robust and replicable outcomes.

1.5 Justification for the study

Discussion in this chapter has outlined the global, national, and individual burden caused by dementia. The necessity of KoD measurement across populations has been

discussed: to inform public health initiatives and educational and clinical competencies, as well as for benchmarking purposes and the identification of shifts or trends in knowledge. To reliably measure unobservable traits; tests or questionnaires must be developed using established guidelines. Of the two main schools of theory in measurement research, test development in relation to KoD is currently anchored by CTT methods. Reports of IRT modelling are predominantly absent from this field, despite growing popularity of these methods in test development in educational and health-related research (Embretson & Reise, 2013; Desjardins & Bulut, 2018). IRT modelling of sets of questions that form dementia knowledge tests is an achievable goal, and one that could compliment and build upon the existing CTT narrative, but also provide informative evidence on the usefulness of KoD tests in specified population groups.

Currently, advanced psychometric methods in dementia knowledge assessment are very limited, with a general lack of robust examination of dimensionality, measurement range, scoring patterns, and item-level analysis. Models that incorporate these techniques have yet to be published in the literature in relation to dementia knowledge. This PhD study therefore contributes to the existing literature and advances scientific knowledge about the scaling properties of KoD tests and items. This exploration of IRT modelling in currently available instruments also offers insight into how improvements in instrumentation may be advanced on a practical level.

1.6. Research aim, objectives, and research question

The aim of this PhD research was to critically explore how dementia knowledge is measured across population groups with a view to advancing the understanding of how measurement instruments work in populations of healthcare staff and students.

The research objectives for this thesis are as follows:

1. Explore the dementia knowledge testing landscape through identification and psychometric appraisal of currently available measurement instruments (systematic review).

2. Acquire and/or generate appropriate dementia knowledge datasets of responses to currently available instruments for further examination using IRT-based modelling.

3. Application and calibration of the Rasch model to evaluate measurement range, measurement precision, Fairness to test takers (presence of item bias), and areas for improvement in methods to help understand KoD instruments.

The primary research question is as follows**:** Can item response theory applications be used to make dementia knowledge tests and testing more informative?

1.7 Thesis structure

The following provides an outline of the structure of this thesis:

Chapter 2 - Theoretical frameworks of measurement

This chapter provides a detailed description of the two theoretical frameworks that underpin measurement of latent concepts, specifically measurement of knowledge. Key developments in measurement theory are discussed, along with the usefulness of specific methods in test development and evaluation procedures. The discussion in this chapter is purposefully broader than that of the specific IRT methods employed in this thesis, the purpose being to provide a robust context of the sophistication and possibilities afforded by measurement theory in general.

Chapter 3 - Systematic review and psychometric appraisal of KoD instruments

Following the background contexts of dementia, knowledge, and aspects of measurement theory that underpin instrument development, this chapter details a critical review of currently available KoD instruments. Comprehensive literature searches sourced fourteen dementia knowledge instruments; these were critically appraised through examination of their reported psychometric properties from instrument development studies and any subsequent studies where further psychometric evaluation took place.

Chapter 4 - Measurement instruments and datasets

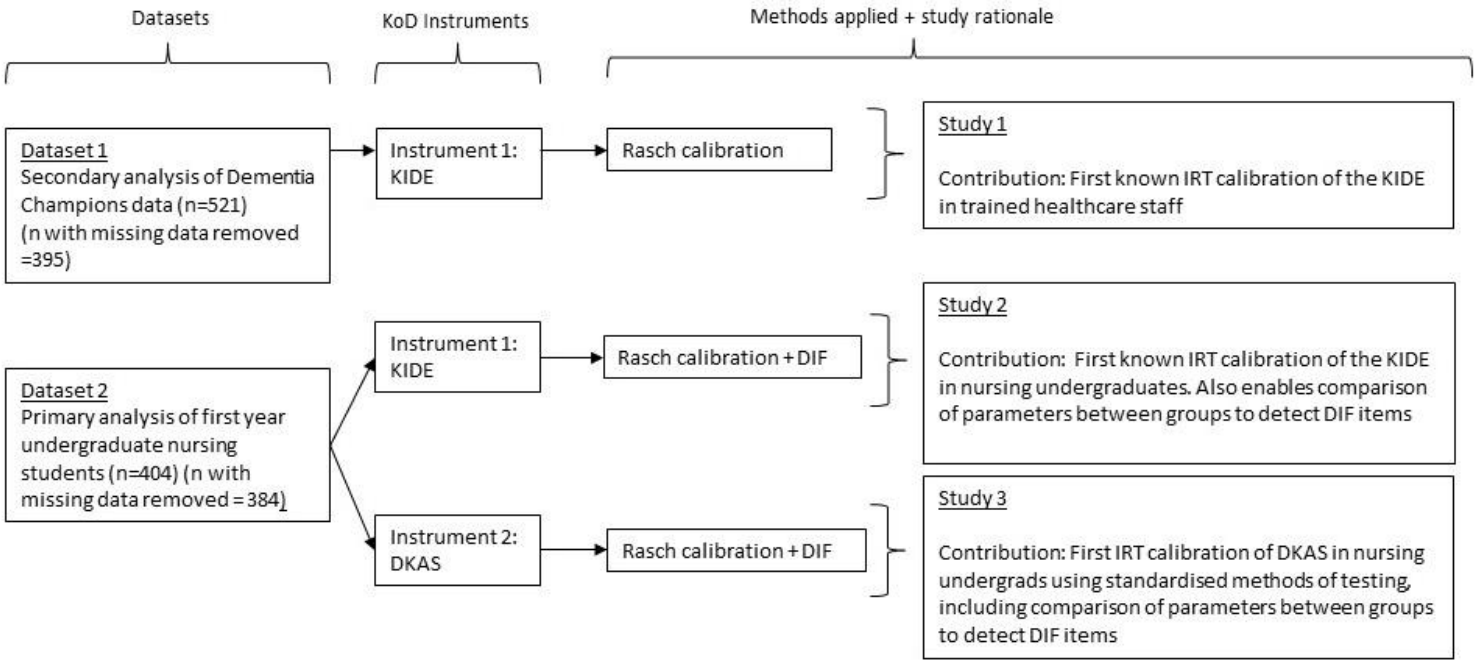
This chapter provides a detailed introduction to, and rationale for, the measurement instruments and datasets that are used throughout the empirical sections of this thesis. Using questionnaire responses from undergraduate nursing students (n=404) and healthcare professionals (n=521), this thesis offers an IRT-based evaluation of two established dementia knowledge instruments. The KoD instruments are: 1. The Knowledge in Dementia Scale **(**KIDE**)** (Elvish *et al*., 2014), and 2. The Dementia Knowledge Assessment Scale **(**DKAS**)** (Annear *et al*., 2015). This thesis details the first application of IRT models in the evaluation of dementia knowledge scales, to the author's knowledge.

To help guide the structure of the empirical work, the methods have been mapped onto a schematic that appears within each empirical chapter with the relevant pathways highlighted. A clean copy of the methods schematic can be viewed below, in *Figure 1.1*.

Chapter 5 – Rasch calibration of the KIDE in healthcare professionals

In this chapter, the Rasch model was used to examine the psychometric properties of the **KIDE** scale in registered healthcare professionals (HCP). This was a secondary data analysis study using data obtained through a collaborative partnership with the Dementia Champions team at the University of the West of Scotland (**Dataset one**). Probabilistic item and person parameters were estimated independently using a Rasch calibration, and the data were evaluated for model and item fit. Effective measurement range of the KIDE items in HCPs was very limited; the Rasch model showed sufficient fit to eleven out of sixteen items. Further, significant mistargeting was identified, with significant ceiling effects evident in the Dementia Champions participants.

Figure 1.1 Schematic detailing the empirical chapters of this thesis



Chapter 6 – Rasch calibration of the KIDE in undergraduate student nurses

This chapter details the application and calibration of the Rasch model to the **KIDE** in **Dataset two**. This was a primary analysis using data collected from first year undergraduate nursing students; these data were collected as an attempt to address the problems around inappropriate targeting of the KIDE that were identified in Dataset one. Probabilistic item and person parameters were estimated independently using a Rasch calibration, and the data were evaluated for model and item fit, including examination for differential item functioning. The effective measurement range of the KIDE items was limited, though less-so than in Dataset one with HCPs. The Rasch model showed sufficient fit to all 16 items and a wider range of ability was captured in the undergraduate nurse sample, however item redundancy and an element of mistargeting lowered the measurement precision of the KIDE.

Chapter 7 – Rasch calibration of the DKAS in undergraduate student nurses

Chapter 7 reports the application and calibration of the Rasch model to a set of undergraduate student nurse responses to the 25-item **DKAS** (**Dataset two**). The objective was to determine whether Rasch calibration could provide additional information about the performance of the DKAS items in this sample, and whether the DKAS dataset demonstrated adequate fit to the Rasch model. The DKAS demonstrated a wide measurement range, however, problems were detected in relation to violations of local independence and undimensionality. In this respect, measurement precision and reliability of the DKAS were compromised. However, Rasch calibration did uncover valuable measurement properties of the DKAS items that could not have been captured using only classical test theory methods.

Chapter 8 – Discussion and conclusions

In this final chapter, results from this PhD thesis are collated and discussed in the context of the wider literature; this discussion frames the contribution to knowledge that this thesis provides. Implications for research and practice within dementia education are discussed in this chapter, and statements are made about the strengths

and limitations of the work contained in this thesis. Following this, discussion around the direction for future work leads to some relevant recommendations for further research. Finally, a conclusion section is offered, to summarise the key points and outcomes from this thesis.

CHAPTER 2

THEORETICAL FRAMEWORKS OF MEASUREMENT

2.1 Introduction

Theoretical frameworks of measurement underpin the development and evaluation of measurement scales in psychology, healthcare and medical education, and a host of other health-related disciplines. Any test must contain a selection of items that represent the phenomenon of interest accurately and comprehensively; these items must also be capable of measuring the phenomenon with effective precision in given targeted population samples.

This chapter introduces the concept of measurement and provides discussion on the two theoretical frameworks that underpin measurement of knowledge in health and related research: *Classical Test Theory* (CTT) and *Item Response Theory* (IRT). The intention is not to provide an exhaustive description of all aspects of CTT and IRT, but instead to explain the background and principles that form the basis of both frameworks. This chapter provides a comprehensive overview of the specific methods and models applied to dementia knowledge datasets during this PhD research, within the wider context of measurement theory.

2.1.1 What is Measurement?

The act of measurement is to assign a number or value to a variable, predominantly to enable comparison within or across variables (De Champlain, 2010). Here, the term *variable* represents any object, trait, or event that can change, and can therefore be measured and potentially manipulated. Measurement of observed, or 'seen', variables is conducted using calibrated measurement instruments, which, when calibrated correctly, are *known* to provide accurate and comparable measurement values. For example, the volume of a liquid might be measured in millilitres or litres, using a pipette, graduated cylinder, or beaker; wherein one could easily compare volumes by counting the number of pipettes it would take to fill one beaker, or comparing the capacity of each in millilitres. A health-related example might be

when health-visitors conduct repeated measures of the height and weight of an infant, the outcomes of which are evaluated against evidence-based parameters to confirm adequate growth over time, allowing for intervention in the event of insufficient growth.

The act of measurement becomes somewhat more complicated when the variable in question is an abstract concept such as emotion, pain, intelligence, or attitudes. *Psychometrics* is the science of measuring such psychological traits and processes (McDonald, 2013). These *unobserved* variables are referred to as *'latent'*, with latent being derived from the Latin term for *hidden*. Where latent variables are concerned, measurement still occurs by assigning a value to a variable, however these values are generated through participant responses to a test, or a single question within a test. Under these circumstances, the value assigned is *assumed* to accurately measure the construct in question (De Champlain, 2010). This is where theoretical frameworks of measurement demonstrate their true value, by providing the statistical models that facilitate understanding of the relationship between the observed variable (a test score) and the unobserved concept of interest (the latent trait).

2.1.2 Theoretical frameworks of measurement

The sub-section of measurement theory called psychometrics is generally accepted to encompass two frameworks that address test development and scoring procedures for the measurement of latent variables (Embretson and Reise, 2013; Steiner and Norman, 2008; DeMars, 2018). Classical Test Theory (CTT) covers the older and more traditional test theory models, whereas Item Response Theory (IRT) is more modern and more psychometrically powerful (Embretson and Reise, 2013), as would be expected in line with advances in statistical methods and software in recent decades. CTT and IRT should not be considered as antithetical theoretical foundations of measurement, but instead as overlapping families of models that can provide a range of useful information about tests, and items they contain, when used appropriately (Hulin, Drasgow and Parsons, 1983; McDonald, 1999).

2.1.3 Points on tests and test development

Measurement of latent variables is conducted using sets of carefully constructed items (questions or statements) that warrant a response; terms used to describe a collective group of items might include (but are not limited to): questionnaire, measure, scale, quiz, instrument, or test. Throughout this chapter, the term ‘test’ will be used in reference to groups of measurement items, unless otherwise specifically stated.

Tests must provide a suitable option for response, of which there are several possible formats. Common response format options include:

i) dichotomous – two response options, with one correct and one incorrect
ii) dichotomous multiple-choice – several response options (usually 3-5) with one correct and all others incorrect
iii) polytomous multiple-choice – several response options (usually 3-5) with more than one correct
iv) Likert scaling – a scale with (usually) 5-7 points on which respondents rate their level of agreement with the statement

The appropriateness of test development and evaluation procedures often relies on item-type; as in, certain models are an appropriate fit for dichotomous items, and others for polytomous items (Rust and Golombok, 1999). As such, it is important to note this distinction before any discussion on the models themselves.

2.2 Classical Test Theory

Within the context of psychometric testing and evaluation, the primary purpose of classical test theory is to ascertain and understand the reliability and validity of tests by examining the relationship between characteristics of the test and the overall score of respondents to the test.

2.2.1 Historical basis of CTT

Classical Test Theory has provided a strong foundational base for test development and the measurement of latent variables since the early 20th Century. CTT is widely thought to have been pioneered through the work of Charles Spearman (1863-1945), who developed the now-famous Spearman correlation coefficient (Spearman, 1987), with this being the first rank correlation capable of accounting for measurement error, due to the monotonic, as opposed to linear, function. The product-moment correlation previously pioneered by Karl Pearson (1857-1936) models *linear* relationships between variables; linear relationships are those in which the observed change in one variable is *proportionally* related to the observed change in another variable. In *monotonic* relationships however, although the two variables change together, they do not necessarily change proportionally to one another, therefore the rate of change is not constant. With this contribution, as well as his seminal contributions to early models of common factor analysis, Spearman helped to pave the way for quantitative representation of latent variables (Lovie and Lovie, 1996).

Many defining achievements for CTT took place in subsequent years, including the work of L. L. Thurstone (1887 – 1955), who expanded on Spearman's factor analysis theory to develop a new exploratory model capable of examining multiple factors under one construct (McDonald, 1999), as in, multiple subscales that contribute towards one latent trait. Factor analytic methods continued to evolve throughout the early 20th Century, despite the increasing computational demands. However, development and practical application of these methods would not be complete until post-1950's, when the computer revolution provided psychologists with powerful statistical software capable of processing the complex mathematics required (Traub, 1997).

CTT methods such as test-level indices of reliability and validity continue to be widely applicable and valuable, though the interpretability of results can be hindered where the methods have been misused or misreported. Such examples can be found in the dementia knowledge testing landscape, where tests have been used in populations with different characteristics to those for whom the test was intended. See Chapter 3: Literature review for further discussion on this.

2.2.2 Foundational assumptions underlying CTT

Estimation of a respondent's position on a latent trait continuum using CTT methods is achieved by summing the item responses to achieve a total score. As an illustrative example, possible dichotomous responses to a 4-item test might include 1100, 1010, 0110, 0101, and 0011, with there being 2 to the power 4 possible patterns in total (sixteen patterns in this case). Scoring in CTT is a simple summation of endorsed items, which equates to the total number of items a participant answered correctly (since correct answers would be scored *1* and incorrect answers scored *0*). As such, the 16 possible response patterns are reduced to only five possible scores, wherein the respondent might score 0, 1, 2, 3, or 4 out of 4. Using the CTT scoring framework on this example dataset, all participants with *any two items* correct would achieve equal score values of 2 (50% correct) and would therefore be assigned to the same position on the latent trait continuum.

No test, examination, or measure is perfectly accurate; observed scores are all subject to some degree of measurement error that is unrelated to the latent variable under consideration (Embretson and Reise, 2013). This is the primary assumption that underlies CTT; that a score assigned to a respondent (X) is in fact a combination of true score (T) and a component of random measurement error (E), as expressed in *Equation 2.1*:

$$T = X + E \tag{2.1}$$

For this reason, CTT is also sometimes referred to as True Score Theory. A true score would be the expected value of the observed score if a candidate took a test an infinite number of times under the exact same conditions (DeMars, 2018). In other words, true score is the mean of infinite Xs for one participant. Common sources of measurement error include poorly constructed test and/or items; inconsistencies in a respondent (as in, having a bad or a good day); or testing conditions and equipment (unintended distractions, or malfunctioning equipment).

Other basic assumptions of CTT are that (a) error variance across respondents is expected to be random and (b), the error is uncorrelated with other variables

including true score, error for scores on other tests, and true scores on other tests (Embretson and Reise, 2013, p42-43).

All CTT analyses are based on the *total scores* for a test in a *defined sample*, and all psychometric evaluation of a test is related to the entire test, in this sample only. Hence, to remove a section of the test, or change the characteristics of the sample would require complete re-evaluation of psychometric parameters and strength (Embretson and Reise, 2013).

Discussion now follows on the CTT methods used to examine dementia knowledge data over the course of this PhD research project; methods include examination of indices of test *reliability* and *validity*, CTT *item parameters*, and inspection for sufficient *unidimensionality*.

2.2.3 Test reliability in CTT

Reliability coefficients (r_{xx}) allow us to estimate how precisely an observed score reflects the true score of the respondent (DeMars, 2018), and can be calculated as the proportion of observed score variance ($\sigma^2 x$) attributed to the variance of the true score ($\sigma^2 T$), as demonstrated in *Equation 2.2*:

$$r_{xx} = \sigma^2 T \,/\, \sigma^2 X \qquad (2.2)$$

The reliability coefficient can range from 0 (all variance is due to measurement error) to 1 (no measurement error detected). Higher reliability coefficients yield responses that are closer to the true score; as all tests contain some amount of error, reliability cannot be reported as 1.0, and as a general rule, estimates of >0.70 are deemed sufficient (Streiner and Norman, 2008). If the reliability estimate is lower (<0.5), then most of the observed variance is due to chance, meaning that the test does not measure the construct it was intended for (De Champlain, 2010).

Two common types of reliability estimates in CTT include test-retest and internal consistency reliability. Test-retest is an estimate of longitudinal stability; it involves administering the test on two or more occasions, in the same sample, with no

intervention in-between. The stability of responses can then be examined; for adequate test-retest reliability, no significant difference between values would be observed (Embretson and Reise, 2013).

Internal consistency reliability is a measure of homogeneity across items within a test, calculated as a function of average inter-item correlation and the number of items in a test (Streiner and Norman, 2008). Low internal consistency estimates (generally below 0.70) indicate that a summed score of a test (e.g. a score of 1/2/3/4 out of 4) would provide limited useful information, since the items do not necessarily measure the same general construct.

One iteration of internal consistency reliability is Cronbach's alpha (α); indeed, this is often the sole type of reliability reported in test development articles (Dunn, Baguley and Brunsden, 2014), and was indeed the case in KoD instruments (Chapter 3). The limitations associated with α, however, have been widely discussed (McDonald, 1999). Cronbach's alpha depends on the assumption of tau-equivalence in a test; tau-equivalence means that all variables contribute equally to the construct of interest, which is rarely the case. Where the assumption of tau-equivalence is violated, α is known to underestimate true reliability. In contrast, α will overestimate reliability in cases where a single factor does not account for all of the common variance across items (Dunn, Baguley and Brunsden, 2014). Further, test length is known to influence reliability estimates, meaning that simply adding more items can increase the reliability when using Cronbach's α (Streiner and Norman, 2008). McDonald's hierarchical omega (ω) coefficient has been cited as an effective alternative to overcome the limitations of α, as omega does not require tau-equivalence or for error variances to be uncorrelated (Zinbarg, Yovel and Revelle, 2006).

2.2.4 Item parameters in CTT

When developing a test using CTT, it is common practice to begin with a large pool of items which is then reduced based on item characteristics and how well they perform in the intended population (Streiner and Norman, 2008). Item analysis involves examination of item parameters and is a principal method for item-pool

reduction. Although primarily a test-based theory, CTT can estimate item parameters of difficulty and discrimination for any group of respondents. Item parameters in CTT can be estimated relatively easily; item difficulty is simply the proportion of correct responses to an item (De Champlain, 2010), this is referred to as the item mean and is represented by the value *P*. Higher difficulty values indicate an easier item, since more respondents correctly endorsed the item, whereas lower difficulty estimations indicate a more difficult item. For continuously scored items (those with unbound response formats), the discrimination parameters are generated through Pearson product-moment correlations between sample score on an item and sample scores on the entire test. Dichotomously scored items instead use the point-biserial correlation coefficient for difficulty parameters (Embretson and Reise, 2013). These item-total correlations indicate whether particular items are likely to be endorsed by respondents with different ability levels (De Champlain, 2010), as in, high correlation between an item score and test score would indicate an item that is likely to be failed by those with low ability and passed by those with high ability. Low correlation would suggest the opposite; such items ought to be flagged for examination or removal from the item-pool, since low correlation estimates here indicate that the item is not functioning as intended. It is important to note that item parameters in CTT are entirely dependent on respondent population characteristics.

2.2.5 Validity in CTT

Evidence of validity of test scores arises where it can be proven that the test indeed measures the concept it purports to measure, in the population it is measuring. A test cannot be deemed valid if evidence of reliability is not sufficient, even though reliability can hold without validity (McDonald, 1999). In other words, consistency of results can render a test reliable, but if it is not actually measuring what it claims to measure, then it cannot be valid, and any results generated are meaningless. As such, test validity is established in relation to a specified purpose/setting/sample and will not be readily generalisable. Test validation often requires both qualitative and quantitative methods; commonly reported forms of validity include content and criterion validity (Embretson and Reise, 2013).

Content validity (otherwise referred to as face validity) is generally established using qualitative methods; topic experts will judge to what extent the items relate to the latent trait, and whether the concept is represented in a comprehensive manner (Streiner and Norman, 2008).

Criterion validity is established when scores on a test correspond to a relevant external criterion. Criterion validity can be readily assessed where a gold standard test already exists. As in, validity can be assumed if the results of a test are in agreement with (concurrent validity) or predict (predictive validity) those of the criterion gold standard (Guyatt et al, 1993). This does however raise questions about the need for a new/additional test if there already exists a gold standard; justification in cases such as these is often that the new test is shorter or more economically viable (Streiner and Norman, 2008).

Like test reliability, the establishment of validity is not a one-off treatment. It has been suggested that it may be inappropriate to claim an instrument/test has been validated (Guyatt et al, 1993), given there are instead varying degrees of confidence in validity estimates in relation to the sample and study setting. This confidence can be built on only with repeated administration and evaluation of a test.

2.2.6 Dimensionality

Examining the number of latent dimensions that underlie responses to a test is an important aspect of test validity; it is also important to a key assumption of common IRT models. A unidimensional test is one in which all items measure the same single trait; where multiple traits underly test items, multidimensionality would be assumed (Embretson and Reise, 2013). In publications that report test development using CTT methods, it is still common to find Cronbach's alpha reliability estimates reported as evidence of unidimensionality (Schmitt, 1996), when in fact Cronbach's alpha is an estimate of internal consistency (or in other words, inter-relatedness) amongst items, and is therefore irrelevant to dimensionality (Streiner and Norman, 2008). As such, if using coefficient alpha as proof of sufficient internal consistency reliability, it should be estimated only after dimensionality testing has been performed, and it must be estimated for each identified dimension.

Factor analysis is one common method of dimensionality testing, in which the structure of the test is examined and the relationships between items (or groups of items) and the defined latent trait are scrutinised (McDonald, 2014). Two common iterations of factor analysis include exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).

Very briefly, factor analysis estimates: a) how highly items load on the first factor, in other words, how highly correlated each item is with a single trait, or with multiple aspects of a trait (each trait would be referred to as a *factor*), and b) the amount of variance that is accounted for by each factor. Such estimates of variance are represented by eigenvalues, which are in turn estimated from a matrix of tetrachoric correlations. For a test to be unidimensional, the eigenvalue (and therefore proportion of variance) for the first factor should be significantly higher than those of all succeeding factors (Streiner and Norman, 2008). For example, in a dementia knowledge test, all items would ideally contribute strongly to the latent trait of general dementia knowledge, therefore establishing unidimensionality. If some items, for example, were to contain attitudinal aspects, then these items would likely group together and load more strongly on a second factor than the first.

Establishing dimensionality is not a straightforward task; there are numerous methods in the literature that have been proposed and investigated in relation to dimensionality assessment (Hulin, Drasgow and Parsons, 1983; McDonald, 2014). However, similarly to the overuse of coefficient alpha, many of these methods continue to be misunderstood and are often misused (Embretson and Reise, 2013). Concepts and methods related to establishing dimensionality, including tetrachoric correlation estimates, factor analysis, and the concept of sufficient or 'essential' unidimensionality will be referenced throughout this thesis in relation to meeting the assumptions of the IRT models applied.

2.2.7 Limitations of CTT

A significant limitation of CTT is the issue of sample dependency; to maintain validity and reliability, *all items* must be administered to *all respondents* (McGrory et al, 2014) which significantly limits the generalisability of tests. Psychometric

analysis of CTT instruments is based on whole test scores, and as a rule, results are not readily generalisable across populations due to variations in respondents' characteristics (Embretson & Reise, 2013). Further, short-form tests and tests targeted at specific ability levels are challenging to develop under the CTT framework which can lead to a lack of precision in testing.

Other limitations associated with CTT include the assumption that all items in a scale contribute equally to the overall score; the assumption that response options have equal intervals, as in, if the options for response were 1 - 4, then the interval between 1 and 2 is the same as the interval between 2 and 3, etc.; and finally, the assumption that measurement error applies equally to all possible scores. As in, respondents with low levels of dementia knowledge will have the same measurement error as those placed at the centre or top end of the trait continuum (De Champlain, 2010). Discussion later in this chapter on IRT reliability indices will illustrate that measurement error varies with more extreme scores.

Now follows an introduction and overview of IRT, including discussion of the statistical models under this framework that have been applied to dementia knowledge datasets during the PhD.

2.3 Item Response Theory

Item response theory models score tests based on participant responses to individual items. Whereas CTT models attempt to explain *collective observed responses* as a function of an unobserved latent trait, IRT models can explain *responses to individual items* as a function of the unobserved trait (Reise and Revicki, 2015). Single items contain additional information; for example, two people with the same score on a test may not have answered the same questions correctly, hence information gained from the whole test score may be limited, unless all items possess equal characteristics (McGrory et al, 2014). IRT methods facilitate the examination of item parameters and patterns of response across datasets, with reliability indices for items and respondents, hereby adding value to test scores when compared (or combined) with CTT analyses.

2.3.1 A brief history of IRT

Classical test theory served well as a theoretical framework for latent measurement well into the 20th Century, and continues to do so, however, item response theory methods have rapidly expanded the potential for what is possible in test development and evaluation (Embretson and Reise, 2013). Tests based on IRT provide more information about items and respondents, and as such are capable of highlighting sets of items that fall within the effective (precise) measurement range of groups of respondents. Further, IRT-based tests are suited to adaptive (computerised) administration, which is becoming the norm in an increasingly digital world (Magis, Yan, and Von Davier, 2017)

The principals of IRT originated from the seminal work of Lord, Novick and Birnbaum (1968) in item parameter estimation methods (Lord and Novick, 2008), and Georg Rasch (1960) with the Rasch model, in contribution to the educational testing landscape as models that facilitated test score comparisons. IRT-based tests and testing expanded beyond educational settings and into health and related research only in the past decade or two (Reise and Revicki, 2015).

2.3.2 Foundational assumptions of IRT

All IRT models express the relationship between a respondent's position on a latent trait continuum (such as dementia knowledge) and the probability that they will endorse an item (as in, score 1 instead of 0) in the form of a logistic model. There are three underlying assumptions of IRT models: (1) unidimensionality – all items measure one dominant latent trait, (2) monotonicity – there is a monotonically increasing relationship between the latent trait and the probability of item endorsement, and (3) local independence – item responses are not correlated when the effect of the latent trait is controlled for (Embretson and Reise, 2013).

IRT models can be categorised by their appropriateness for dichotomous or polytomous items, and whether they estimate parameters for unidimensional or multidimensional data. This thesis employed unidimensional dichotomous models only, and as such, description has been limited to these models. For further information on IRT models for polytomous and/or multidimensional data, readers are

directed to the following texts and their references: *Item Response Theory for Psychologists* by Embretson and Reise (2013); and *Test Scoring* (Thissen and Wainer, 2001).

To explain the IRT models from the perspective of this PhD project, the formal mathematical models will be outlined in the context of dementia knowledge measurement. The symbol for the Greek letter *theta*, Θ, is the common notation in IRT for the latent trait. Θ will be used throughout the following sections to describe the latent trait of dementia knowledge.

2.3.3 IRT modelling – item parameter estimation

IRT models characterise items based on the *Item Characteristics Curve* (ICC) which is a non-linear regression relating the *probability* of correct response (a score of 1 versus 0) as a function of θ (DeMars, 2018). It is important to note the relationship between true score (*T*) in CTT and θ in IRT: when the assumptions of IRT are withheld, *T* and θ are essentially the same ability estimated using different metrics. The key difference between the two metrics is that, where the mean (average) and variance (distribution) of the θ continuum are established/calibrated, a respondent's θ estimate is not dependent on one specific measurement instrument but can be compared across other sets of items that have been calibrated on the same latent trait continuum (Hulin, Drasgow and Parsons, 1983).

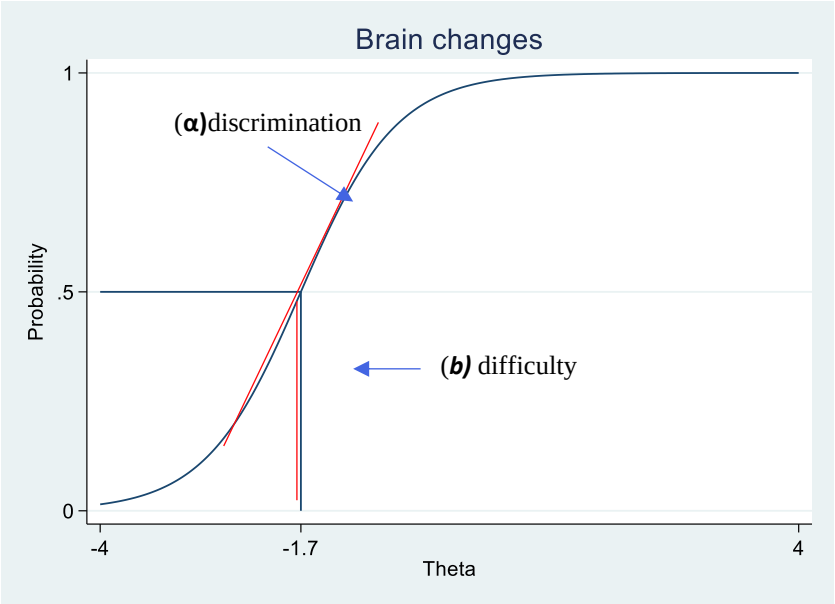
In IRT, measurement precision varies in line with ability level, for example, when a respondent's ability increases after an educational intervention, so will their probability of correct response to more difficult items (Embretson & Reise, 2013). Two useful interpretations of the ICC are item *difficulty* and *discrimination*. Item difficulty is the θ estimate that is associated with a 50% probability of scoring one, rather than zero. Examination of item difficulty indices is important in matching the suitability of a test to the population target (Mair, 2018). As in, test items should be not so difficult to cause frustration, and not so easy to cause boredom or inattention.

Item discrimination indices identify how well an item can differentiate between respondents/groups with different levels of the trait in question, such as knowledge of dementia expressed by a consultant versus a member of the public. (McGrory et

al, 2014). Items that are more highly discriminating are characterised by steeper slopes on their ICC. Such items are weighted more heavily in a test and can render IRT scores more reliable than CTT (DeMars, 2010).

As an example, the ICC for a dementia knowledge item "Permanent changes to the brain occur in most types of dementia" can be seen below in *Figure 2.1*, where the x-axis denotes the Θ estimate (zero being the sample average), and the y-axis demotes the probability of correct score, from 0 to 1. Given the Θ of zero is the sample average, the item is relatively easy, being 1.7 standard deviations below the mean; this is graphically represented in the curve being situated to the left of the plot. This example item discriminates well between respondents with low and high knowledge levels, as can be seen by the relatively steep slope of the curve. The *a* parameter estimate for the item is 1.83, with estimations of $a > 1$ being desirable.

Figure 2.1. Item characteristics curve (ICC) for the item "Permanent changes to the brain occur in most types of dementia", short code name: 'Brain changes'



2.2.4 Test information and the standard error of measurement

Whereas the ICC is the graphical representation of item parameter estimates under an IRT model, the *item information curve* (IIC) (also called information *function*) represents the Θ range over which an item provides the most information, as in, the item would be of limited use in respondents out with the information range. The standard error of measurement (SEM) indicates how precise this information is. Information functions can be generated for each item within a test; individual information values can then be summed to generate the information function for the test overall.

An advantage of IRT is that the SEM acts as an index of precision for test scores (DeMars, 2010); importantly, the SEM can be calculated for each item within a test, as well as the test itself. Throughout the empirical sections of this thesis, the SEM is often abbreviated further to 'standard error' or SE.

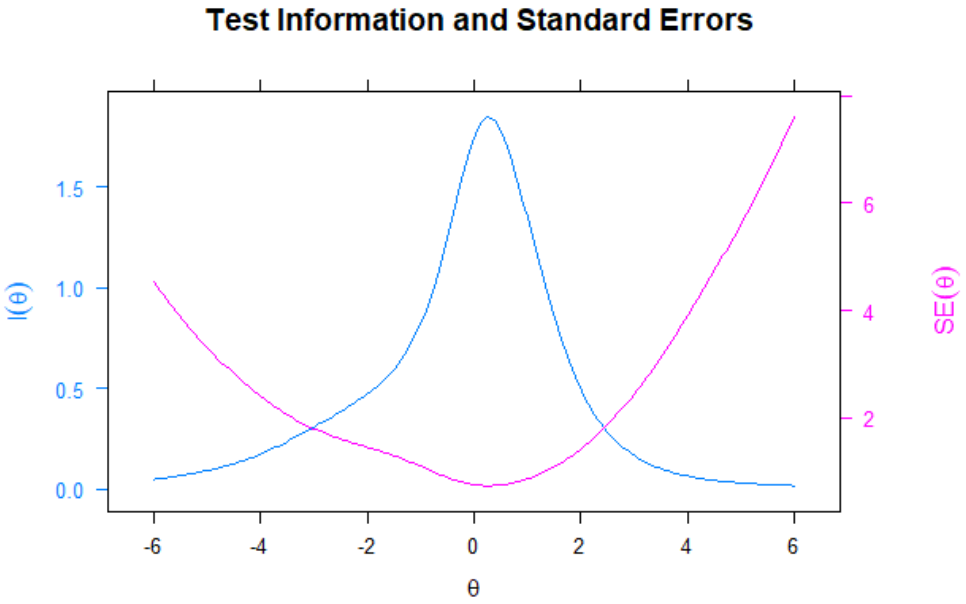
The standard error is an estimation of the expected variations in score due to measurement error (Embretson and Reise, 2013, p.16). The SEM is a principal feature in the evaluation of the psychometric quality of a test, and the values can be used as confidence intervals to guide the interpretation of scores. Whereas the mean of infinite tests in one respondent would be equal to the true score, the standard deviation (SD) of scores in CTT would correspond to the SEM in IRT (De Champlain, 2010). The standard error (σ_E) can be calculated as the square root of 1 minus reliability (1 – ρ_{XX}), times the SD of the test (σ_X), as expressed in *Equation 2.3*:

$$\sigma_E = {}^{\sigma_X}\sqrt{1 - \rho_{XX}} \tag{2.3}$$

The test information function is used to calculate the SEM and therefore the effective measurement range of the test, that is, the items in a test where the SEM is sufficiently low to facilitate precision of measurement. The *test information function* is a graphical representation of where on the dementia knowledge continuum the test provides the most and least information (Embretson and Reise, 2013). This function is used to calculate the SEM, which is the inverse of the square root of the curve.

Both test information (blue curve) and standard errors (red curve) can be seen in *Figure 2.2*, where the x-axis denotes Θ (zero being the sample average) and the left y-axis denotes information value, increasing from low to high information. The right y-axis denotes the standard error of measurement, with the lowest values representing the greatest precision. This plot shows that this example test provides the most information with sufficient measurement precision over the Θ range of approximately -1 to 1.

Figure 2.2. An example of test information and standard errors.



IRT information functions here provide an advantage over CTT, where an amended test would require a repeated administration to estimate updated reliability indices. In IRT, information functions are generated for all items, and then summed, therefore different combinations of items can be tested to find the group with the most precision. (DeMars, 2010). New test information functions can be generated for subsets of items with ease.

2.3.5 IRT models for dichotomous data

A selection of unidimensional IRT models facilitate the estimation and examination of item parameters for dichotomously-scored items; common models include the one-parameter logistic (1PL) model, the Rasch model, and the two-parameter logistic (2PL) model; estimation of item parameters is also referred to as IRT *calibration*. Unidimensional models are based on the assumption that all items in a test correspond to a single common factor.

2.3.6 Rasch model and One Parameter Logistic model (1PL)

The IRT models for dichotomous data with the least number of parameters are the Rasch model and the one-parameter logistic (1PL) model (Finch and French, 2015). *Equation 2.4* expresses the Rasch model, where the probability (*P)* of a correct (*1*) response from person (*s*) to item (*j)* (denoted as $P(x_{sj} = 1)$) is determined by the item's difficulty (b_j) and the respondent's position on the latent trait (Θ_s). Additionally, the following constant is included in these unidimensional models: (*e*) is the base of the natural logarithm scale (2.718) (since these are *logistic* models).

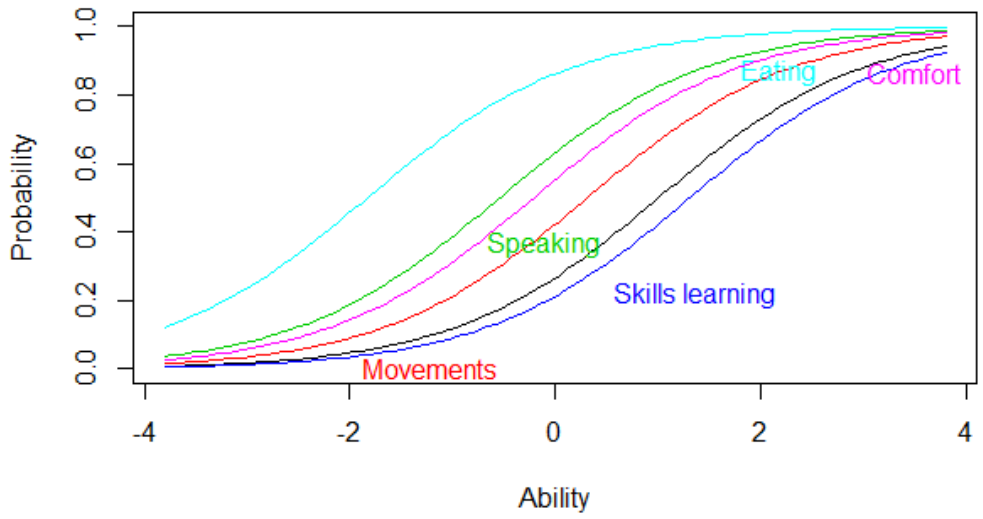
$$P(x_{sj} = 1|\theta s, b_j) = \frac{e(\theta s, -b_j)}{1 + e(\theta s, -b_j)} \tag{2.4}$$

The Rasch model is a special case of the 1PL model in which item difficulty parameters are unique, but item discrimination parameters are fixed to 1.0. In the 1PL model, item discrimination parameters are also constrained to be equal across items; a key difference between 1PL and Rasch is that the constraint value for item discrimination is determined from the data. *Equation 2.5* denotes the 1PL model, with (*a*) expressing the common item discrimination value.

$$P(x_{sj} = 1|\theta s, b_j) = \frac{e^a(\theta s, -b_j)}{1 + e^a(\theta s, -bj)} \tag{2.5}$$

When graphically represented as ICCs, item discrimination parameters are characterised by the slopes of the curves, and therefore the Rasch and 1PL model (with constrained discrimination parameters) generate equally steep curves for all items, meaning that the ICCs will converge, but never cross (Andrich and Marais, 2019). An example can be found in *Figure 2.3*, where item difficulty parameters vary and therefore the curves cover a range of Θ. The curves toward the left show less difficult items and the curves further to the right show more difficult items.

Figure 2.3. Example ICCs generated under the Rasch model.



In the wider context of measurement theory, and usually in item response theory, the terms Rasch modelling and 1PL modelling are often used interchangeably (Andrich and Marias, 2019). However, it is important to note that, although all IRT models are concerned with the estimation and examination of item and person parameters, there are fundamental conceptual differences in Rasch measurement theory to IRT more generally (Andrich, 1988). The Rasch model is seen as a formal representation of proper measurement and as such, misfit to the Rasch model indicates weakness in the data, not weakness of the model (Boone, Staver and Yale, 2013). In the case of misfit, defective data cannot be improved by the inclusion of an additional parameter

(as discussed below in section 2.3.7) and must therefore be examined and revised. In IRT however, misfit to the 1PL model points to a weakness in the model, not the data, therefore the data should be re-examined using a different model, without necessarily altering the dataset (Embretson and Reise, 2013).

For the purpose of this thesis, Rasch measurement theory is considered to be an important framework of measurement that sits within the wider framework of item response theory. There is however considerable debate in the literature about the appropriateness of this due to the theoretical discrepancies discussed above. For further reading around this interesting topic, see chapter three in Thissen and Wainer's (2001) text 'Test Scoring', and David Andrich's (2004) article "Controversy and the Rasch model: a characteristic of incompatible paradigms?". Currently, in the field of dementia knowledge measurement (discussed in Chapter 3 of this thesis), and in nursing research more generally, neither Rasch measurement theory methods nor item response theory methods is widely employed or understood (Hagquist, Bruce and Gustavsson, 2009).

Extended Rasch models are available, for example the partial credit model and the rating scale model, however, since the data used in this thesis are dichotomous, the Rasch model in its original form, as described by Georg Rasch, (1960) was an appropriate model. As such, all references to the Rasch model throughout this thesis are referring to the dichotomous Rasch model, unless otherwise stated.

2.3.7 Two Parameter Logistic model (2PL)

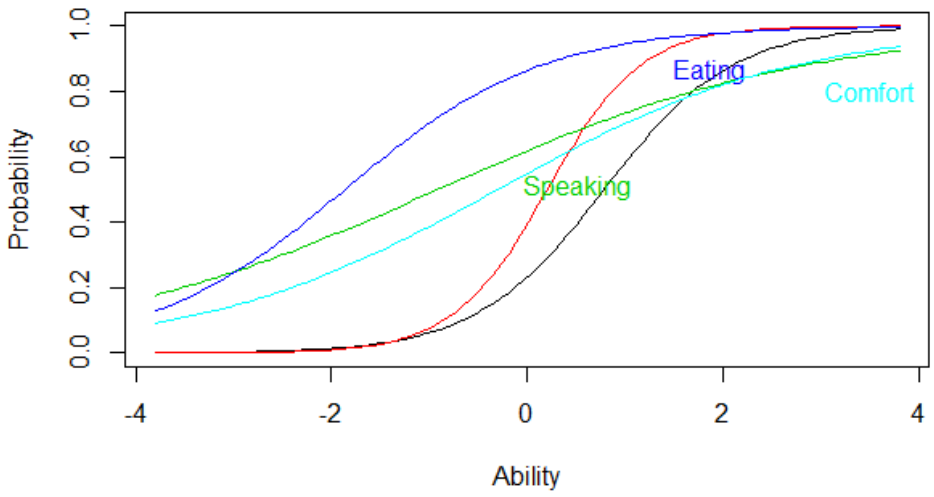
In addition to item difficulty parameters, the 2PL model allows for unique discrimination (a_j) parameters for each item, as expressed in *Equation 2.6*:

$$P(x_{sj} = 1|\theta s, b_j) = \frac{e^{a_j}(\theta s, -b_j)}{1 + e^{a_j}(\theta s, -b_j)} \qquad (2.6)$$

In contrast to the Rasch / 1PL models, the ICC slopes may cross where discrimination parameters vary across items. A primary advantage of using the 2PL

model is that it facilitates greater precision scoring since more item parameters are taken into account (Embretson and Reise, 2013). An example of ICCs generated under the 2PL model can be found in *Figure 2.4*, where item discrimination parameters vary, meaning the curves have different slopes and therefore intersect with one another.

Figure 2.4. Example of ICCs generated under the 2PL model.



As a general rule, for model selection, the most parsimonious model is the model that demonstrates the best fit with the least number of item parameters; if the Rasch model showed adequate fit to the data, then Rasch should be used over, for example, a 2-parameter-logistic (2PL) model (Meijer and Tendeiro, 2018).

2.3.8 Scoring respondents on a latent trait using IRT

Once item parameters have been estimated under an appropriate model, IRT-based scoring algorithms can estimate a respondent's position on a latent trait continuum using patterns of item response and parameter estimates (Embretson and Reise, 2013). An individual pattern of response is generated for each participant who takes a test; for binary-response data, response patterns form a matrix of 0s and 1s from which response patterns and their frequencies can be examined.

Returning to the example dataset of binary responses to a 4-item test (1100, 1010, 0110, 0101, and 0011 with 16 possible patterns of response); IRT-based scoring recognises each unique pattern and can assign each to a position on the latent trait continuum, based on the characteristics of the correctly endorsed items (Bartholomew, Knott, and Moustaki, 2011). Therefore, using the example above, respondents with a CTT score of 2 can be scaled more accurately (and individually) on the trait continuum.

An example of how IRT-based scoring can extract additional information is represented in *Table 2.1*, using an excerpt from results generated in the early stages of this PhD. A selection of response patterns to a 16-item dementia knowledge test are displayed alongside their associated test scores (CTT method) and estimated Θ value (IRT method). The first and final rows (highlighted in bold) show patterns of responses that generated equal CTT scores of 10, whereas 2PL calibration of the items illustrated that items were not equally weighted, therefore the score of '10' might indeed place a respondent at multiple positions on the dementia knowledge continuum.

Table 2.1. A selection of response patterns to a 16-item test, with associated Θ estimates (ranging from lower knowledge to higher knowledge), standard deviations, and total test score.

Θ estimate	**SD**	**Test score**	**Response pattern**
-1.551	**0.258**	**10**	1 1 0 1 1 1 1 1 1 1 0 0 0 0 0 1
-1.523	0.262	8	1 0 0 0 0 1 1 1 1 0 1 0 1 1 0 0
-1.437	0.280	8	1 1 0 1 1 0 0 0 1 1 1 1 0 0 0 0
-1.424	0.283	9	0 1 0 0 1 1 0 1 1 0 1 1 1 0 1 0
-1.423	0.283	8	0 1 0 0 1 0 0 1 1 1 1 1 0 0 1 0
-1.409	0.286	10	0 1 1 0 1 1 0 0 1 0 1 1 0 1 1 1
-1.189	0.334	11	1 1 0 0 1 1 1 1 1 0 1 1 1 0 0 1
-1.004	**0.369**	**10**	0 1 0 1 1 0 1 1 1 1 1 1 0 0 1 0

2.3.9 Differential item functioning

One important application of IRT that has yet to be deployed in dementia knowledge measurement is that of differential item functioning (DIF). Items that display DIF perform differently across sets of respondents and therefore scores obtained in a test with such items cannot be compared (Embretson and Reise, 2013); as such, the presence of DIF is a threat to the validity of any test. As an example, it is not uncommon for student nurses to have work experience in healthcare settings prior to beginning their degree programme, particularly in the form of a year-long access course which is common in the UK. Such respondents would likely have come into clinical contact with people with dementia and as such, knowledge items may be biased in favour of this group if assessing KoD in Year one student nurses.

When items are calibrated under an IRT model, item parameters should be the same across all groups, within the range of measurement error. Two types of DIF can be assessed for: *uniform* DIF is present where item *difficulty* parameters differ between groups; *non-uniform* DIF is present where item *discrimination* parameters differ across groups (Mair, 2018).

2. 4 Conclusions

This chapter has provided discussion on the two theoretical frameworks of measurement, CTT and IRT. These overlapping frameworks contain numerous methods and models that are useful in test development and evaluation. Though by no means an exhaustive overview of these elements of measurement theory, description here has been framed in the context of this PhD thesis, in an attempt to set the scene for the chapters that follow.

Dementia knowledge measurement is currently underpinned by CTT methods of test-level evaluation of validity and reliability indices. This PhD employed IRT methods including Rasch model calibration to score respondents on the latent trait of dementia knowledge. This facilitated in-depth examination of item parameters and measurement precision of previously validated KoD instruments. Potential bias in item performance across groups was also examined using methods to assess for

differential item functioning. This extent of IRT modelling is currently absent in the field of dementia knowledge measurement.

The following chapter details a systematic and psychometric review of dementia knowledge tests, where a reliance on CTT methods in test development procedures will become evident, further justifying the exploration of IRT-based methods in dementia knowledge tests and testing.

CHAPTER 3

A SYSTEMATIC AND PSYCHOMETRIC REVIEW OF DEMENTIA KNOWLEDGE INSTRUMENTS

The work in this chapter was initially completed in 2016. A manuscript titled "Measurement properties of instruments evaluating Knowledge of Dementia: A methodological review**"** was submitted for publication towards the end of 2018 and the searches have since been updated in response to comments from reviewers and the journal editor. The keyword searches were conducted again in 2019 to bring the review up to date, and second and third reviewers were invited for title/abstract screening, full text screening, data extraction, and quality appraisal. To reflect this, the study design has been updated to 'systematic and psychometric review', and the manuscript was undergoing major revisions at the time of writing this chapter.

3.1 Introduction

Measurement of dementia knowledge is necessary in the development and evaluation of educational interventions and public awareness campaigns. With the increasing burden of dementia, it has become important that instruments exist that accurately assess and quantify levels of Knowledge of Dementia (KoD) between populations. This is especially true for settings where people will routinely come across people with dementia, and will be required to provide treatment, support, and guidance (such as healthcare environments). However, given recent healthcare policy shifts advocating for those affected to live in their own communities for as long as possible (Scottish Government, 2017), understanding community-level trends in dementia awareness has become equally important.

There is no standardised approach to the measurement of dementia knowledge. Currently, researchers and educators have a choice between several scales for varying target populations, and with various response formats and reported psychometric strengths (Spector *et al.,* 2012). These instruments have struggled to achieve widespread adoption; many are used infrequently beyond the study in which they were developed. This might be explained, in part, by the limited psychometric

evidence and justification provided for available instruments in existing reports and development studies. This lack of standardisation of measurement instruments can render it challenging to compare scores across or within populations (Sunderland *et al.*, 2018), and to achieve harmonization of datasets where different scales have been used (Griffith *et al.*, 2013).

An earlier review of dementia knowledge instruments examined the reported psychometric properties and appropriateness of administrations of five instruments across populations including healthcare professionals, informal caregivers and lay populations. Spector *et al.'s* (2012) search strategy was robust and data on reported psychometric strength were extracted not only from the instrument development studies but also from subsequent administrations of the measures. However, although their chosen quality appraisal criteria were transparent, these were limited in number and not based on any pre-defined framework. Acceptability of face and content validity were based on subjective expert judgement. The term 'construct validity' was used to cover concurrent, convergent, and divergent validity, but also sensitivity to change; parameters for Spearman or Pearson's coefficients were specified here in relation to 'acceptable' levels of reported validity. Reliability here covered internal consistency reliability and test-retest reliability, with parameters identified for acceptable reporting of Cronbach's alpha and Cohen's kappa. The data extracted by Spector *et al.* (2012) in relation to psychometric performance of the instruments were informative, however the use of a guiding framework may have facilitated a more robust examination of the quality of instruments.

This current systematic review strengthens and provides an update to the previous review by Spector *et al.* (2012). The aim of the present study was to identify existing instruments and report on their content and psychometric analysis and evaluation. As a result, this field will have an updated source and summary of instruments designed to assess Knowledge of Dementia across or between populations - and if used longitudinally - over time.

3.2 Methods

This systematic and psychometric review employed a range of methods to explore the field of research around dementia knowledge tests; some methods were

traditional whereas others may be viewed as more novel. All methods associated with this review are detailed in the following sections, with the aim of providing clarity and transparency.

3.2.1 Literature search

To identify KoD instruments for appraisal, an approach based on article citation tracking was utilised. This (relatively novel) approach was possible following Spector *et al's* (2012) review identifying five KoD tools for which citation tracking searches and alerts were set up and then prospectively monitored. Since there has been considerable debate about the effectiveness of citation tracking as a sole method in literature searches (Wright, Golder and Rodriguez-Lopez, 2015; Janssens and Gwinn, 2015) a standard systematic database search approach was also performed. Hence two methods of searches were conducted with the intention of comparing results, which are detailed below in sections 2.2 and 2.3.

3.2.2 Citation searches and alerts

A staged, two-set scientific citation search designed following Spector *et al's* (2012) systematic review of 'knowledge of dementia' outcome measures was performed in August 2019 in Web of Science (webofknowledge.com). Citation index functions in such databases can be used for forward citation tracking, where *all articles that cite a selected study* are identified, but also backward citation tracking, where all articles *cited by* the selected study are found (reference list checking) (Kuper, Nicholson, & Hemingway, 2006). These functions facilitate the option to prospectively monitor future citations of prior instruments that will likely capture all new instrument development studies, as well as studies that detail additional psychometric properties of existing instruments. The five instruments identified in Spector et *al's* (2012) review were used as an anchor from which to base a new round of evidence retrieval, seeking citations of said instruments.

Set one involved forward citation tracking from the five instruments identified in the original review to identify newly developed instruments, which were expected to

refer to/reference one of these. Another round of forward citation tracking was then carried out from any newly identified instruments, until this approach yielded no new results. After three rounds of searches, no further instruments were identified. Additionally, forward citation tracking was performed on Spector *et al's* (2012) review.

The second set of citation searches involved forward citation tracking from all included instrument development studies (meaning the original five and all newly identified instruments). This stage differed from the previous stages in that the purpose was not to identify recently developed KoD instruments, but instead to source all empirical studies in which any of the included measures of knowledge have been administered. The purpose of this stage was to build on the quality assessment of each instrument, by extracting data on the psychometric testing and evaluation beyond those reported in the development studies.

Monthly citation alerts for each included KoD instrument were set up to keep track of any developments in psychometric test reporting, or new instrument development studies. Although the original searches were performed in 2016 and then repeated in 2019, the citation alert protocol facilitates rapid retrieval of any new articles that are relevant to this field of research and practice.

3.2.3 Protocol-based keyword searches

A traditional systematic review protocol-based search was conducted in August 2019 to provide a comparison method. The systematic searches were conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher *et al.,* 2009) and were equivalent to *Set one* of citation tracking. Set two citation searches were a follow-on that were dependent on the results of Set one and the systematic searches.

Electronic database searches were conducted in CINAHL Plus, PsychINFO, ProQuest, Pubmed, Embase and Google Scholar using the search components listed:

(dementia OR alzheimer's) AND knowledge AND (questionnaire OR measure OR instrument OR test OR quiz) AND (care staff OR nurs* OR general public OR general physician OR healthcare prof*').

Further to the electronic searches, hand searching and cross-checking of reference lists of seminal papers pertaining to knowledge of dementia and knowledge of dementia outcomes was performed.

3.2.4 Inclusion criteria

All included KoD instruments (*set one citation searches*) were required to be published (a) in English, (b) in a peer-reviewed journal, and (c) from 1988 onwards (consistent with the review by Spector *et al.*, (2012)). No geographical limitations were placed on the searches. Further, to be included, articles had to meet the following inclusion criteria, adapted from the original review:

- Have published psychometric properties (regardless of whether they are reported to be strong or poor)
- Describe the development process of the instrument, including discussion on how the initial item pool was generated, as a minimum standard
- Be developed for further use, ie. not only for the study alongside which it is published
- A focus on general Knowledge of Dementia over all other outcomes

There were no geographical limitations, and there were no population target limitations (unless the instrument was designed for non-English speaking groups – the instrument should be *validated* in English, not only reported in English). Only original instruments were included, therefore adaptations of instruments or translated versions of existing instruments were excluded.

For *set two citation searches* (identifying administration of instruments in empirical studies), the following criteria were required:

- Knowledge of Dementia assessed using one or more of the instruments identified for inclusion
- No more than 50% adaptions to the original instrument

3.2.5 Establishing quality indicators of instruments

The quality of included instruments was assessed using criteria recommended by Terwee *et al.* (2007) who guide critical appraisal of 8 psychometric properties, as follows:

i. content validity (extent to which the domain of interest is comprehensively sampled by the items in the questionnaire)
ii. internal consistency reliability (extent to which items in the (sub)scale are intercorrelated)
iii. criterion validity (extent to which scores on a particular questionnaire relate to a gold standard)
iv. construct validity (extent to which scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured)
v. reproducibility (agreement: extent to which scores on repeated measures are close to each other
vi. reliability: extent to which respondents can be distinguished from each other, despite measurement errors)
vii. responsiveness (ability of the questionnaire to detect statistically important changes over time)
viii. floor and ceiling effects (number of respondents who achieved the lowest or highest possible scores)
ix. interpretability (degree to which one can assign qualitative meaning to quantitative scores).

Data were fully extracted using a data extraction form informed by these guidelines, and each property was rated as positive (+), intermediate (?), negative (-), or omission (0). Quality assessment was initially conducted based on psychometric properties extracted from the instrument development studies, then additional data relating to psychometric properties revealed in subsequent testing were extracted from later studies. Detailed guidance on the statistical tests and parameters for quality appraisal can be found in the original Terwee *et al*. (2007) article.

3.3. Results

3.3.1 Search results – set one

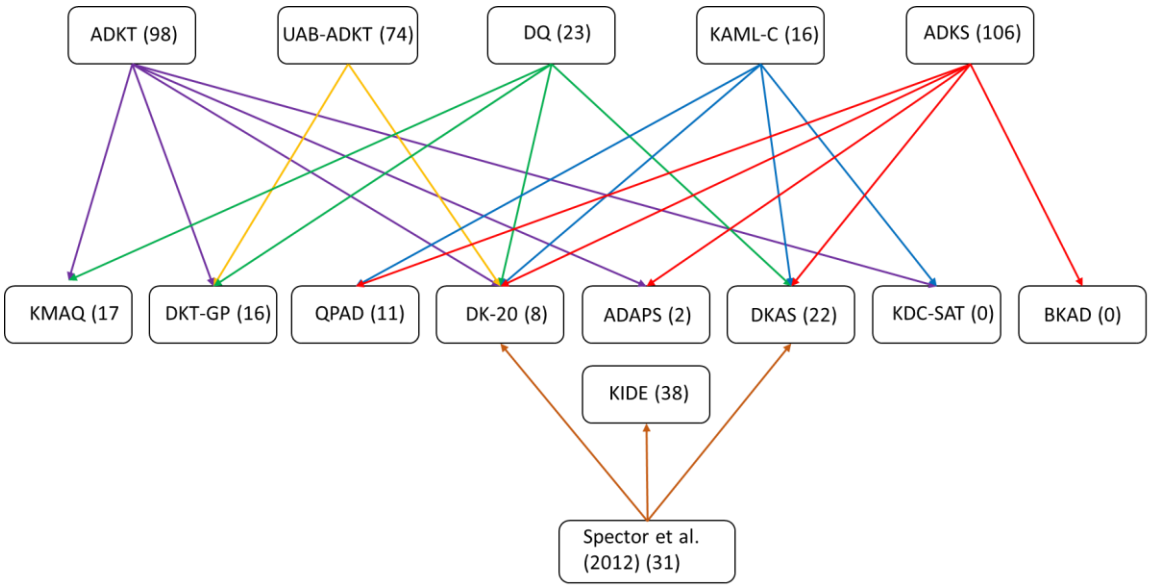
During set one, 462 titles and abstracts were screened to reveal *nine* KoD instruments. Forward citation tracking from the five anchor instrument development articles plus the original systematic review identified 348 articles that were screened; from these, nine new instrument development articles met the inclusion criteria. Forward citation tracking from the nine new instrument development articles identified 114 articles that were screened; from these, no further KoD instruments were discovered. The nine new instruments were retained alongside Spector *et al.'s* (2012) original five, resulting in 14 articles for inclusion. Search results for the citation tracking approach are presented in *Table 3.1*.

Table 3.1. Results from Set one citation searches

Set one citation search	**Anchor articles**	**Number of citations**	**Outcome**
Step 1	5 instruments + Spector et al. (2012) article	317 + 31	**9 new instruments**
Step 2	9 instruments	114	**0 new instruments**

Figure 3.1 is a graphical representation of Set one citation searches. Spector *et al*.'s (2012) five anchor instruments (using their abbreviated labels – see Table 3.2 for full information) are shown across the top, with the Spector *et al*. (2012) article at the very bottom. Connecting lines identify which new instruments were identified from the citations of each article (both from the anchor articles and the original review). The number of citations from each instrument development article is in brackets alongside the instrument abbreviation. The anchor instrument development articles and the 2012 review article all identified between two to five new instruments.

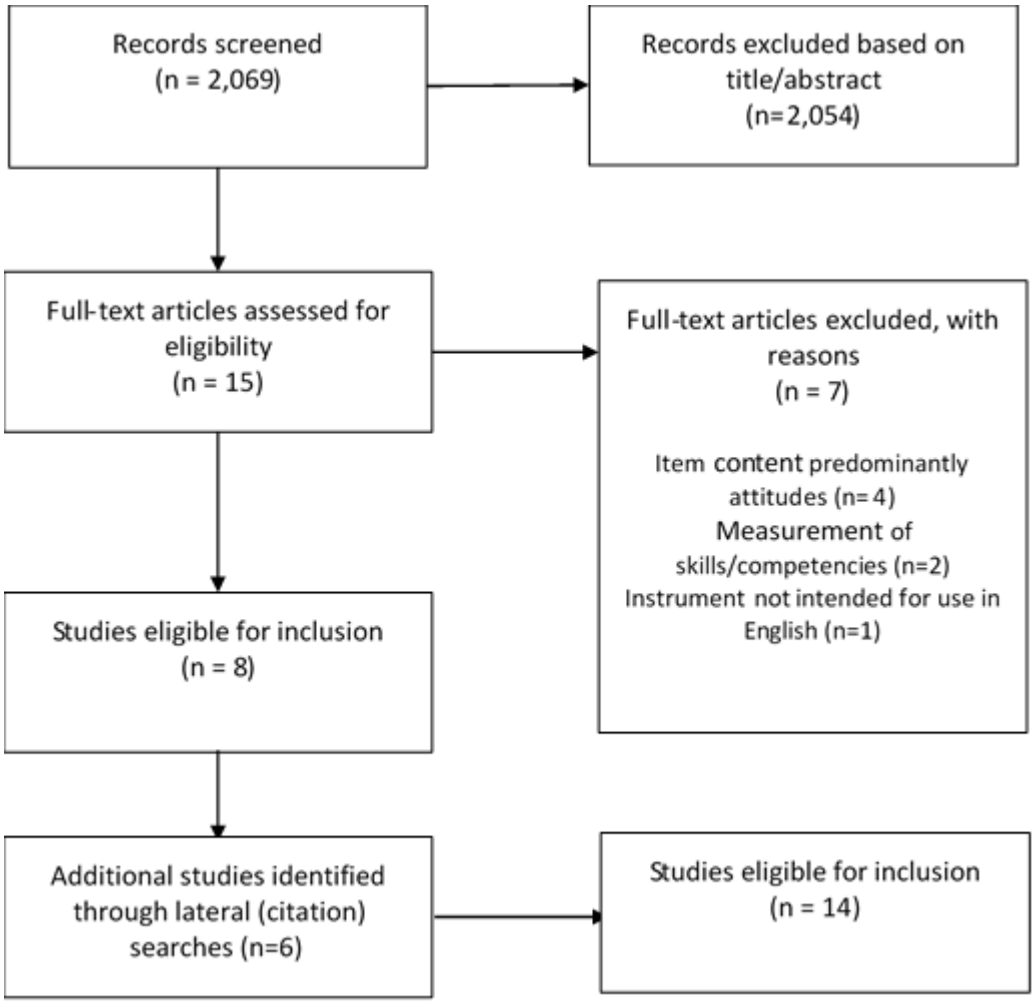
Figure 3.1. Set one citation tracking process from the five anchor instrument development articles and the original systematic review.



3.3.2 Keyword-based searches

The keyword-based searches identified 2,069 records that were screened to reveal *eight* KoD instrument development articles. Of these eight, three were anchor articles from Spector *et al.*'s (2012) review, and the other five were included in the pool of new articles identified by Set one citation searches. The keyword-based searches failed to yield four articles that had been identified during citation tracking, and two of the anchor instruments. *Figure 3.2* shows the PRISMA flowchart for the keyword-based searches.

Figure 3.2. PRISMA flowchart describing the process of article selection for the keyword-based search results



3.3.3 Characteristics of instrument development studies

In total, 14 KoD instruments were identified and reviewed (five anchor and nine new). The instruments are: Alzheimer's Disease Knowledge Test (ADKT) (Dieckmann et al., 1988); Dementia Quiz (DQ) (Gilleard & Groom, 1994); UAB Alzheimer's Disease Knowledge Test (UAB-ADKT) (Barrett et al., 1997); Knowledge of Memory Ageing Questionnaire (KMAQ) (Cherry et al., 2000); Knowledge About Memory Loss and Care (KAML-C) (Kuhn et al., 2005); Alzheimer's Disease Knowledge Scale (ADKS) (Carpenter et al., 2009); Dementia Knowledge test for GPs (DKT-GP) (Pentzek et al., 2009); Questionnaire on Palliative Care in Dementia (qPAD) (Long et al., 2012); DK-20 (Shanahan et al., 2013); Alzheimer's Disease and Perception Scale (ADAPS) (Bettens et al., 2014); Knowledge in Dementia Scale (KIDE) (Elvish et al., 2014); Dementia Knowledge Assessment Scale (DKAS) (Annear et al., 2015a); Knowledge of Dementia Competencies Self-Assessment Tool (KDC-SAT) (Curyto & Vriesman, 2016); Basic Knowledge of Alzheimer's Disease (BKAD) (Wiese et al., 2017). Overall instrument characteristics are as follows:

- The 14 included studies ranged in publication date/year from 1988 to 2017.
- Most of the instruments originated from the USA (n=8) followed by the UK (n=3) and Australia (n=2), with one measure having been developed in Germany.
- The most common response format was multiple choice (n=7) followed by True/False (n=4) but other formats included Agree/Disagree (n=1), an adapted four-point Likert scale with additional 'I don't know' option (n=1), and "multiple choice combined with true/false format" questions (n=1).
- Heterogeneity was evident, with several different populations sampled: e.g. professional and non-professional caregivers, students, older adults, lay people (members of the public), non-university-educated staff, long-term care staff, general practitioners and hospital staff.
- Eight instruments were developed to assess knowledge of all/any dementia without any specific stage of illness being identified; three pertained to Alzheimer's Disease (AD) only. Two assessed knowledge of early AD, and one concerned advanced dementia.

- The number of subscales identified within instruments ranged from 0 - 8.
- The number of article citations ranged from 0 – 106.
- Pilot samples ranged in size from n=72 – n=1,767. All included articles contained at least some discussion around the development process and psychometric properties of the measures.

Reported instrument development processes for the 14 KoD instruments are presented in *Table 3.2*.

Table 3.2. Knowledge of dementia instruments: characteristics summarised from published studies.

	Author, date, country, instrument (acronym)	**Test format, reported item reduction process, target populations**	**Sample populations/settings**	**Reported development process**	**Citations /administrations (accurate September 2019)**
1	Dieckmann *et al.*, 1988, USA. Alzheimer's Disease Knowledge Test (ADKT)	20 multiple choice items (initially 36 – reduced in 2 stages). For caregivers, mental health professionals, nursing home staff	Total participants *(n=96)* Undergraduate students *(n=34)* Undergraduate gerontology students *(n=23)* Graduate students *(n=18)* MH professionals *(n=21)*	1.Item generation, 2. item selection, 3. content validity, 4. item analysis and PPDI, 5. final item selection, 6. internal consistency reliability, 7. construct validity, 8.test-retest reliability	98/30
2	Gilleard & Groom, 1994, UK. Dementia Quiz (DQ)	25 multiple choice items (initially 36 – reduced in 2 stages) For caregivers and professionals	Total participants *(n=298)* Experienced carers *(n=194)* Naïve carers *(n=45)* Aware professionals *(n=21)*	Unpublished questionnaires assembled, draft versions subject to expert review – minimal information given.	23/5

	Author, date, country, instrument (acronym)	Test format, reported item reduction process, target populations	Sample populations/settings	Reported development process	Citations /administrations (accurate September 2019)
			Less aware professionals *(n=16)* Ex-carers *(n=22)*		
3	Barrett *et al.*, 1997, USA. UAB Alzheimer's Disease Knowledge Test (UAB-ADKT)	12 multiple choice items (initially 75 – reduced in 3 stages) For health professionals	Total participants *(n=610)* Specialists *(n=9)* Generalists *(n=148)* Experts *(n=108)* Experts *(n=116)* Generalists *(n=179)* Undergrad nurses *(n=50)*	1.articulation of domains, 2. choice of test format, 3. item generation, 4. review of items, 5. administration to representative sample, 6. item analysis, 7. revised administration, 8. empirical analysis of final items	74/7
4	Cherry, 2000, USA	28 true/false items (initially 34 – reduced in 1 stage)	Total participants *(n=134)*	No discussion on where original items came from – content	17/5

	Author, date, country, instrument (acronym)	Test format, reported item reduction process, target populations	Sample populations/settings	Reported development process	Citations /administrations (accurate September 2019)
	Knowledge of Memory Aging Questionnaire (KMAQ)	For students, older adults, service providers	Clinical and cognitive psychologists *(n=14)* Undergraduate students *(n=120)*	validity modelled on Dieckmann et al's development of ADQ	
5	Kuhn, King and Fulton, 2005, USA. Knowledge about Memory Loss and Care (KAML-C)	15 multiple choice items (initially 31 – reduced in stages – no information on how many stages of reduction) For family caregivers	Total participants *(n=121)* Caregivers *(n=45)* Experts (medical, nursing, social work, research, psychology, gerontology) *(n=37)* Medical students *(n=39)*	1.articulation of knowledge domains, 2. choice of test format and item generation, 3. administration to experts, 4. administration to sample populations, 5. item analysis	16/0
6	Carpenter *et al.*, 2009, USA.	30 true/false items (initially 57 – reduced in 2 stages)	Total participants *(n=865)* College students and older adults *(n=unknown)* College students *(n=26)*	Review of 21 existing instruments. 1.articulation of domains, 2. choice of format, 3. item	106/40

	Author, date, country, instrument (acronym)	Test format, reported item reduction process, target populations	Sample populations/settings	Reported development process	Citations /administrations (accurate September 2019)
	Alzheimer's Disease Knowledge Scale (ADKS)	For laypeople, patients, caregivers and professionals	Individuals of any age with no CI* *(n=40)* Students *(n=36)* HCPs* *(n=75)* Senior staff *(n=61)* Caregivers *(n=54)* Older adults with no CI *(n=89)* College students *(n=484)*	generation, 4. pilot testing, 5. analysis	
7	Pentzek *et al.*, 2009, Germany. Dementia Knowledge Test for GPs (DKT-GP)	20 multiple choice items (initially 59 – reduced in 2 stages) For General Practitioners	Total participants *(n=308)* Dementia experts *(n=9)* GPs *(n=7)* GPs *(n=292)*	37 items chosen in previous study (Pentzek et al, 2006) – most items developed specifically for DKT-GP, others taken and translated from prior instruments	16/1

	Author, date, country, instrument (acronym)	Test format, reported item reduction process, target populations	Sample populations/settings	Reported development process	Citations /administrations (accurate September 2019)
8	Long *et al.*, 2012, USA. Questionnaire on Palliative Care for Dementia (qPAD)	23 true/false items (initially 50 – reduced in 2 stages) For long-term care staff	Total participants *(n=85)* Consisting of: HCAs*, caregivers, registered nurses, activities staff, social workers, dieticians, admin staff, enrolled nurses and others.	1. Literature search, 2. Initial test developed from 20 item KAT (University of Iowa) and expanded to 50 items incorporating key palliative care concepts. 3.pilot testing, 4. analysis and revision, 5. further analysis and expansion	11/2
9	Shanahan *et al.*, 2013, UK. DK-20	20 multiple choice items (initially 39 – reduced in 3 stages) For unqualified care staff	Total participants *(n=256)* Pilot 1: *(n=45)* Dementia experts *(n=7)*, Care staff *(n=38)* Pilot 2: *(n=211)* Care staff *(n=153)*, AHPs* *(n=32)*, management staff *(n=20*), non-clinical staff *(n=6)*	1.literature search, 2. focus group, 3. expert review, 4. item generation and choice of test format, 5. expert commentary, 6. pilot testing	8/2

	Author, date, country, instrument (acronym)	Test format, reported item reduction process, target populations	Sample populations/settings	Reported development process	Citations /administrations (accurate September 2019)
10	Bettens *et al.*, 2014, Australia. Alzheimer's Disease and Ageing Perception Scale (ADAPS)	25 multiple choice items (initially 60 – reduced in 3 stages) For general community and aged care professionals	Total participants *(n=252)* General community *(n=196),* Aged care professionals *(n=56)*	1.item generation, 2. piloting, 3. item analysis, 4. validation	2/0
11	Elvish *et al.*, 2014, UK. Knowledge in Dementia Scale (KIDE)	16 agree/disagree items (initially 27 – reduced in 1 stage) For staff caring for people with dementia in general hospital settings.	Total participants (n=72) Nurses (n=21) AHPs (n=17) FY1s (n=10) HCAs (n=6)	1.item generation, 2. piloting, 3. item analysis and revision, 4. further analysis	38/3

	Author, date, country, instrument (acronym)	Test format, reported item reduction process, target populations	Sample populations/settings	Reported development process	Citations /administrations (accurate September 2019)
12	Annear *et al*., 2015(a), Australia. Dementia Knowledge Assessment Scale (DKAS)	27 modified Likert scale items (initially 52 – reduced in 3 stages) For 'Those who provide care and treatment for people with dementia'	Total participants *(n=1767)* Nurses *(n=495)* Professional care workers *(n=467)* Other: students, retired persons and other health workers *(n=805)*	1.item generation 2.pretesting 3.pilot testing 4.psychometric evaluation 5.principal component analysis	22/4
13	Curyto and Vriesman, 2016, USA. Knowledge of Dementia Competencies Self-Assessment Tool (KDC-SAT)	82 items – combination of true/false and multiple choice (initially 100 – reduced in 1 stage) For 'Direct care workers'	Total participants *(n=159)* Direct care workers	1.item generation 2.pilot testing 3.psychometric evaluation 4.analysis	0/0

	Author, date, country, instrument (acronym)	Test format, reported item reduction process, target populations	Sample populations/settings	Reported development process	Citations /administrations (accurate September 2019)
14	Wiese *et al.*, 2017, USA. Basic Knowledge of Alzheimer's Disease (BKAD)	20 items, T/F response format (reduced from 39 through content analysis and cognitive interviews with nurse researchers) For 'Rural, underserved populations'	Initial pilot (n=200) older rural adults. Additional sample of (n=20) healthcare providers for discriminant validity	Followed the development process of Polit and Yang (2016), includes cognitive interviews. Rasch modelling for internal consistency and stability. Test retest (n=20) after 3 weeks (separate from pilot sample)	0/0

3.3.4 Psychometric appraisal of instrument development

Criterion validity could not be assessed since there is currently no ‘gold standard’ (for validity confirmation) and no benchmark instrument against which to compare newly developed tools. Further, none of the included articles published sufficient information with which to assess the property of responsiveness. For the remaining 6 properties outlined in the Terwee *et al*. (2007) framework, psychometric reporting was generally consistent across all instruments, with a marked absence of reported testing for some domains; the results of psychometric appraisal are presented in *Table 3.3*.

Table 3.3. Results of psychometric appraisal of KoD instruments using the Terwee et al. (2007) framework.

	Content validity	Internal consistency	Construct validity	Reproducibility - Agreement	Reproducibility - reliability	Floor or ceiling effects	Interpretability
ADKT	+	?	+/?	0	?	?/+	0
DQ	+	-	?/-	0	0	0	?
UAB-ADKT	+	?	?	0	?	?/+	?
KMAQ	?	-	?/+	0	0	?	?
KAML-C	+	-	?/+	0	0	?/+	?
ADKS	+	+	?/+	0	?	?	?
DKT-GP	+	+	0	0	0	?	?
qPAD	?/+	-/+	?/0	0	0	?/+	0
DK-20	+	-	?/+	?	+	?/+	?
ADAPS	+	-	?/+	0	0	?/+	?
KIDE	?/+	?	0	0	0	0	?
DKAS	+	-/+	?/+	?	?	?	?
KDC-SAT	+	?	?	?	+	?	?
BKAD	+	?	?	0	0	+	+

Key: positive (+), intermediate (?), negative (-), or omission (0).

3.3.5 Search results – set two

Set two searches (citations of instruments) identified 462 articles spanning 30 years, of which 99 reported on an administration of the KoD instruments. Of the 14 included instruments, the number of administrations post-development spanned from 0 – 40, with the ADKT and the ADKS accounting for the majority of these, having 30 and 40 published accounts of administration, respectively. Predominantly, the 99 administration studies did not report on any further psychometric evaluation of the instruments.

3.4 Discussion

The psychometric properties of 14 KoD instruments were considered based on results reported in development and subsequent studies. Properties were directly compared, and the results aggregated to form overall conclusions on the application of KoD instruments within intended population groups.

3.4.1 Citation tracking methods

These findings highlight the efficacy of citation tracking as a potential method for updating systematic reviews of instrument development studies. In this instance, the citation tracking method identified six more studies for inclusion than the keyword-based searches, whilst producing less than a quarter of the number of records to be screened (462 vs 2,069).

3.4.2 Psychometric properties of KoD instruments

The psychometric properties of 14 KoD instruments were considered based on results reported in development and subsequent studies. Properties were directly compared, and the results aggregated to form overall conclusions on the application of KoD instruments within intended population groups. KoD instruments are not well characterised psychometrically, with none having achieved positive ratings for even half of the properties referred to by Terwee *et al*. (2007).

Collectively, the lack of further use of KoD instruments following their administration in development studies is an important finding, given that the process of instrument development begins with, but is not complete after just one study (McDonald, 2013). Rather, cumulative evidence of psychometric quality is required and built upon through further administrations, or practical application, in suitable populations (Embretson and Reise, 2013). This leads to somewhat of a conundrum, as researchers/educators may be less inclined to apply a lesser-used scale due to lack of verified psychometric quality, yet repeated practical application facilitates data collection for robust evidence of established psychometric properties.

Across the KoD item sets in this review, only two studies report concurrent administrations outside of any instrument development study (between the ADKT, DK-20 and ADKS; and DKAS and ADKS) (Sullivan & Mullan, 2016; Annear *et al*., 2016). Indeed, even during the development of new instruments, concurrent administration with already established KoD instruments has been uncommon, based on the articles included in this review. In future application of KoD tests, sets of instruments might be administered concurrently more often than has been the case as this helps to establish dimensionality as well as in validation (Bannigan and Watson, 2009).

3.4.3 Practical application of the KoD instruments

To assess dementia knowledge in healthcare staff and students, the ADKS and DKAS appeared to demonstrate greater psychometric strength than other instruments. Instruments for other populations including family caregivers and GPs are available, but could benefit from further application and psychometric validation, as detailed below.

Instruments intended for general use in healthcare staff and students

The ADKT is the oldest and most widely cited instrument; it was also the most frequently used until 2009, when the ADKS was published. Item content of the ADKT is now outdated. Consequently, the ADKS has taken over as the most frequently used instrument.

The DQ and UAB-ADKT also appeared to contain somewhat outdated item content. They lacked any psychometric evaluation studies beyond their development. Outdated content of these older instruments along with the availability of the ADKS, DKAS and DK-20 as more contemporary instruments suitable for use in healthcare staff, indicated that the ADKT, DQ and UAB-ADKT may have limited value in future evaluations when viewed in the context of the wider set of available measures.

The ADKS, DK-20 and DKAS underwent robust development in large samples of students and healthcare staff. The ADKS and DKAS appeared to be the most psychometrically robust instruments, having performed better than most other instruments under Terwee *et al*.'s (2007) framework. Psychometric testing post-development (including translation) has been reported for both. Further assessment of the DK-20's performance has been reported in relation to the ADKS (Sullivan & Mullan, 2017); although the ADKS demonstrated poorer internal consistency reliability, the DK-20 scale had not been subject to factor analysis. Until this testing occurs, the ADKS and DKAS are likely to remain the stronger instruments.

The KMAQ and qPAD require further psychometric evaluation to establish adequate content validity and are therefore not recommended for use in current evaluations without further critical testing.

Instruments intended for general use in untrained care staff and lay populations

To assess dementia knowledge in wider populations of untrained healthcare staff and laypeople, the ADAPS, KIDE, KDC-SAT and BKAD contain potentially useful items. However, further published studies that report on aspects of psychometric evaluation are required, particularly to examine scale structure and dimensionality. Although the KIDE was developed for hospital staff, further administration (Lorio *et al.*, 2017) suggested the presence of ceiling effects in this population and as such it may prove more useful as a basic KoD test for lay populations.

Instruments intended for specific use in family caregivers and GPs

The KAML-C and DKT-GP are unique in their target populations, being family caregivers and GPs, respectively; as such, their item content is valuable. Although

these instruments have had limited use out with their development studies, they have the potential to be updated and expanded for use in similar populations, with further psychometric validation.

3.4.4 Position of this review and resulting recommendations

This review supports and updates the findings of Spector *et al*.'s (2012) review, which concluded that all KoD instruments had weaknesses regarding psychometric properties and the confirmation of these in their target population: application of item response theory methods is absent, and factor analytic methods are limited in frequency and scope. Regarding the lack of reporting of psychometric properties; the current results echo findings from the wider literature on assessment tools (see Trevena & Waters, 2014; Yang *et al*., 2015; Davies, Waters & Marshall, 2016; Clari, *et al.,* 2016; Lui, Kim and Alessio, 2020); each of these reviews identifies widespread omissions in the reporting of psychometric properties. This lack of reporting is of concern as it may lead to instruments being disregarded for use due to seemingly limited psychometric strength. Simply, judgement on the suitability of instruments becomes challenging when all of the required information is not available.

Current consensus on the importance of dementia knowledge in guiding education and awareness interventions is clear (Cahill et al., 2015; Sullivan & Mullan, 2017; Surr et al., 2017); as such it is important to highlight the lack of published studies that seek to establish how KoD instruments perform across populations. Further, longitudinal validation studies are absent, as are empirical studies demonstrating how newer instruments maintain consistency with historical item-sets/datasets for benchmarking purposes.

Based on the quality appraisal of development process, item content, and reported psychometric properties for these dementia knowledge tests, suggestions for research and further empirical use can be offered.

Recommendations for further research

One recommendation to come from this review is to further test the efficacy of citation tracking when updating reviews of measurement instruments. Further applications of citation-based searches and keyword-based searches alongside one another would contribute to this limited evidence base.

A recommended route for future research using KoD instruments is for further *concurrent* administrations of instruments; this would allow for joint calibrations, assessment and comparison of dimensionality, and item level analyses. This would facilitate more robust psychometric appraisal of the relative efficiency of item sets.

Recommendations for practical application of the instruments

Further evaluation of the ADKS, DKAS and DK-20 as part of joint KoD instrument validation studies could further establish dimensionality, sensitivity, and measurement range in populations of healthcare staff. Likewise, administration of these item-sets alongside the older ADKT, DQ and UAB-ADKT would work towards strengthening the empirical evidence on the performance of items across populations, and benchmarking against historical item-sets. However, if administering the ADKS, DKAS, and/or DK-20 concurrently, one must be mindful that they were developed in different countries (USA, Australia, and UK, respectively) which, although all Western and English-speaking, may account for any cultural differences in item content.

Further administration of the ADAPS, KIDE, KDC-SAT and BKAD would ideally be in samples of untrained healthcare staff and wider lay populations, with a focus on examining dimensionality of the item-sets to establish structural reliability. Additionally, the ADAPS may benefit from a more standardised response format. Currently participants are asked to respond with ‘A – characteristic of normal ageing’ or ‘B – characteristic of Alzheimer’s disease; a ‘true/false’ response format may widen the potential of this instrument.

Regarding the KAML-C and DKT-GP, and as mentioned previously, further administration could help to widen the scope and establish usefulness of these item sets in family caregivers, and medical/specialist nursing staff, respectively.

3.4.5 Strengths and limitations

One of the main strengths of this review is the concurrent application of citation searching and traditional search methods. In this instance, citation tracking proved to be the more efficient method; this finding may be of interest to researchers wishing to update systematic reviews of measurement instruments. A related limitation may be that the keyword-based search protocol was not strong enough to yield the same results as the citation tracking method. However, all methods have been reported thoroughly in an attempt to render them transparent and reproducible. As a further limitation, it should be acknowledged that the reported psychometric properties of measurement instruments may have been limited by journal-imposed word limits, and this in turn would influence our ability to appraise the instrument to its true extent.

3.5 Conclusions

Robustly developed, contemporary item-sets with established and reproducible psychometric properties are a desirable and achievable goal. Their development assists in the identification of knowledge levels, and hence gaps, across populations, but also in establishing the effectiveness of educational or staff training interventions. The question of instrument selection in any study is ultimately down to the researcher/educator, however, with a range of seemingly 'validated' KoD measures currently available, this decision could benefit from more detailed psychometric characterisation and acceptability of use/evaluation. In their current form, many of the KoD instruments appear to have limited usefulness in healthcare practice and policy due to insufficient reporting of psychometric testing and results, and outdated item content. This review has provided suggestions on the use of instruments (or sets of instruments) based on appraisal of content and psychometric strength.

The following chapter introduces the empirical sections of this thesis, the work of which was heavily influenced by the results of this systematic and psychometric

review, in which an evident lack of IRT methods in the field of dementia knowledge research and measurement instruments was revealed.

CHAPTER 4

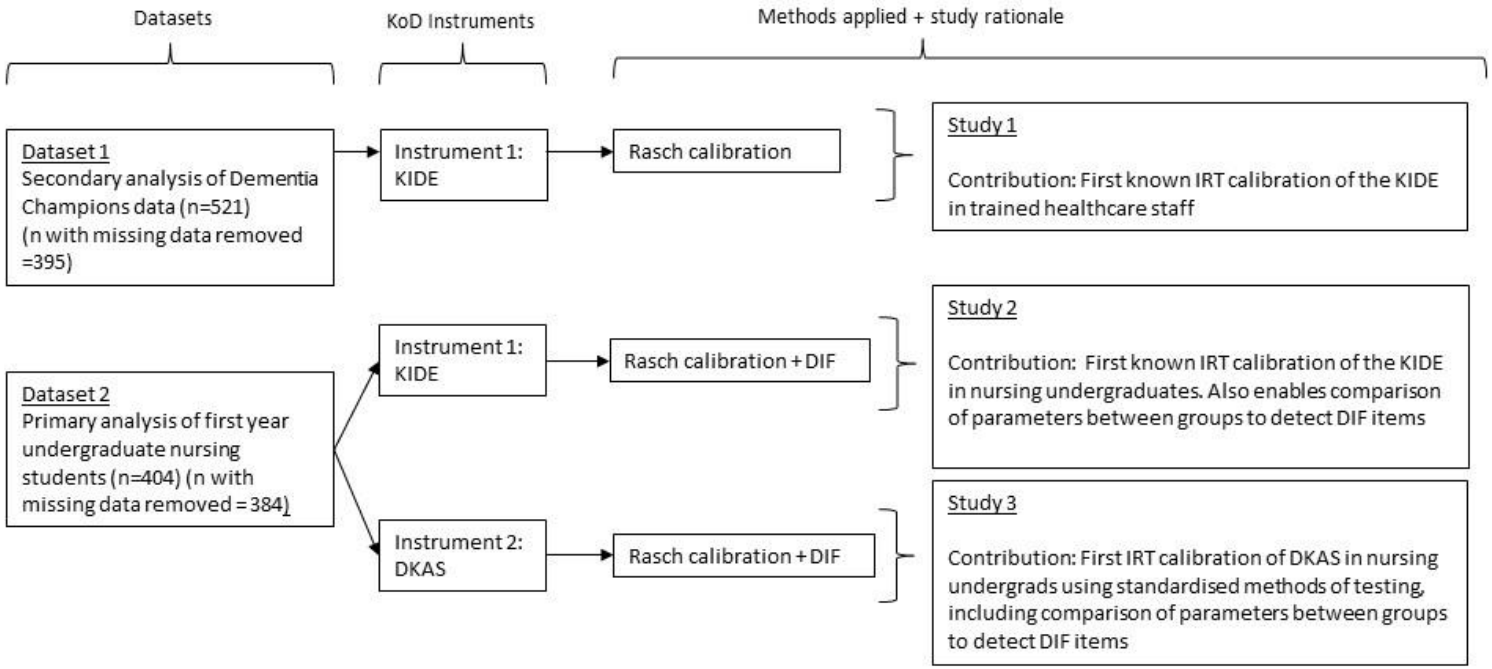
MEASUREMENT INSTRUMENTS AND DATASETS

4.1 Introduction

As per the methods described in *Chapter 2, theoretical frameworks of measurement*, item response theory (IRT) modelling techniques were applied to two datasets containing participant responses to knowledge of dementia (KoD) items.

This chapter describes the two datasets used in the empirical sections of this thesis; methods of data collection and data entry are detailed, and the two datasets are described in order to set the scene for the analytical chapters 5-7. The two dementia knowledge measurement instruments used are also described here. The instruments used are the Knowledge in Dementia (KIDE) scale (Elvish *et al*., 2014) and the Dementia Knowledge Assessment Scale (DKAS) (Annear *et al*., 2017). Description of the instruments includes information on their content, methods of development, and any post-development psychometric analysis. *Figure 4.1*, shown below, shows the datasets, KoD instruments, and methods applied in this thesis, in relation to one another. This schematic will be used throughout the results chapters (5-8), with relevant sections highlighted to signpost the reader to the section being discussed at each stage.

Figure 4.1. Methods schematic showing datasets, instruments, methods, and overall contributions



4.2 Datasets

The first dataset was acquired through a collaboration that was formed in the early stages of the PhD, and the second dataset was generated based on the results and recommendations that arose from the secondary analysis of dataset one. The methods involved in acquiring and generating both datasets are described here.

4.2.1 Dataset one – qualified health and social care staff

The first dataset analysed in this thesis was acquired on forming an academic collaboration, one of the terms of which being that the dataset would be shared for secondary analysis. The background to this collaboration was as follows.

Background to dataset one – formation of academic collaboration

A cross-institutional academic collaboration was formed between staff in the School of Health Sciences, University of Dundee (SHS, UoD) and the School of Health and Life Sciences, University of the West of Scotland (UWS) after both parties met whilst networking at a dementia-related conference. The UWS team were in the planning stages for cohort 10 of their Dementia Champions programme, which was a dementia education programme for qualified health and social care staff. The team had been experiencing difficulties with their evaluation of previous cohorts, including an increase in ceiling effects and a lack of sensitivity, therefore a reduced ability to measure increases in dementia knowledge post-intervention (Jack-Waugh *et al*., 2018).

The author of this thesis was attending the conference to present the results of a systematic review of the psychometric properties of dementia knowledge measurement instruments (as per Chapter 3) and was therefore approached by the UWS team to discuss the potential availability of (a) dementia knowledge instrument/s that may be suitable for use in the Dementia Champions programme. An academic collaboration was formed to bring the author onto the Dementia Champions team for cohort 10 in an attempt to investigate and potentially address the ceiling effects noted when measuring dementia knowledge pre- and post-

intervention. In exchange, the team would share the dataset of participant responses from previous cohorts of the programme for secondary analysis. A Memorandum of Understanding (MoU) was then generated between the universities with the specific objectives and activities of the co-operative relationship established as follows:

a) Create an agreement for sharing previous Dementia Champions data collected using the Knowledge in Dementia Scale (Elvish *et al*., 2014) with the Dundee team to be utilised in analysis of empirical data sets and updating of items (questions). *(Cohorts 6-9)*

b) Subject to an agreement, to share results of the above activities with UWS to inform the evaluation methods for Alzheimer Scotland Centre for Policy and Practice (ASCPP) education activities, starting with Champions Cohort 10 in March 2019.

c) For Dundee to conduct further analysis on Cohort 10 data in line with validation.

d) Generate shared narrative and plan future activities to be coordinated in relation to Dundee thesis submission and REF 2021.

e) Plan for joint dissemination activities including local seminars/training for UWS staff on item level analyses, academic publications and conference presentations.

As such, anonymised data from cohorts 6-10 of the Dementia Champions programme were made available for secondary analysis as part of this PhD project. The full MoU document can be viewed in full in *Appendix 1*.

Dementia Champions programme

The Dementia Champions programme was a dementia education intervention, developed in response to Scotland's first National Dementia strategy which contained a commitment to improve the quality of care and treatment for people with dementia in hospital settings (Scottish Government, 2010). The primary purpose of the intervention was:

> *"…to enable staff to lead and support change in the workplace, to improve care and treatment outcomes for people with dementia" (Jack-Waugh et al, 2018)*

Participants enrolled on the Dementia Champions programme were professionally qualified health and social care staff, including registered nurses and allied health professionals (AHPs) who were working in NHS or social care settings in Scotland. At the time of writing, the intervention had been implemented for ten cohorts and more than 1,000 Dementia Champions had been trained through the programme (MacRae *et al*., 2019). Planning and preparation for cohort 11 was underway, with the intervention being redesigned for online delivery due to the Covid-19 pandemic.

The Dementia Champions programme is a blended learning intervention delivered over eight months. It consisted of self-directed distance learning, five face-to-face teaching days delivered across multiples sites, one half-day spent in community settings, and three written assignments. See Brown *et al.* (2018) and Jack-Waugh *et al*. (2018) for further information on the Dementia Champions programme. Evaluation of the intervention consisted of self-completed measures of dementia attitudes, dementia knowledge, and self-efficacy on the first and final teaching days. The pre-programme questionnaire can be viewed in *Appendix 2*.

The complete dataset consisted of pre- and post-intervention responses to three measurement instruments (one attitudes, one knowledge, and one self-efficacy) for cohorts 6-9, and responses to four measurement instruments (as before with an additional dementia knowledge instrument) for cohort 10. Primary analysis of this dataset (up to cohort 9, by the UWS team) consisted only of descriptive statistics and repeated measures t-tests to assess the differences in participant responses pre- and post-programme. No further statistical methods had been used on the data.

This thesis details a secondary analysis of the data relating to pre-intervention dementia knowledge scores for one instrument only; the KIDE (Elvish *et al*., 2014). The second knowledge instrument (the DKAS [Annear *et al*., 2017]) was introduced only during cohort 10, and therefore the sample size was too small to be sufficient for item response theory analysis. It is important to note that no demographic data

were collected as part of the intervention, as such the sample could not be analysed in relation to age groups, or by sex to address issues of fairness.

Ethical approval

Ethical approval for the Dementia Champions programme had been granted by the University of the West of Scotland Research Ethics Committee. An ethics amendment was submitted by the UWS team based on the terms and conditions outlined in the MoU. The amendment was granted and no further ethical approval was required (for example on the UoD side) in light of the active MoU.

Sample and data collection

The dataset consisted of pooled responses to one dementia knowledge instrument from 5 cohorts. Data were collected from participants during the first and final study days of the Dementia Champions programme. Cohorts included here were those numbered 6-10, these being held over a three-year period between 2017-2019. During this period, the training was delivered by the same team of nurse academics (AJW, RM, LR[1]), with the addition of the author (CG) for cohort 10. Survey packs were administered in paper and pencil format in a lecture theatre setting. Data collected included participant responses to 45 items (16 being dementia knowledge items) during cohorts 6 - 9, and responses to 65 items (41 being KoD items) during cohort 10.

Data entry

Data entry for cohorts 6-9 of the programme was completed by a member of the Dementia Champions team (LR). Data entry for cohort 10 of the programme was conducted by the author (CG) on-site at the UWS campus where the final training day took place. For cohort 10 data entry, pre- and post-intervention survey packs were matched by health board and postcode details, data were entered using SPSS

[1] Dr. Anna Jack-Waugh (AJW), University of the West of Scotland (UWS); Dr. Rhoda MacRae (RM), UWS; Dr. Louise Ritchie (LR), UWS; Clair Gamble (CG), University of Dundee.

v.25, and 10% of entries were double checked by a member of the Dementia Champions team for accuracy.

4.2.2 Dataset two – undergraduate nursing students

The second dataset analysed in this thesis was generated the thesis author, for primary analysis of two dementia knowledge instruments in undergraduate student nurses.

Background to dataset two

The design of the dataset two was based on recommendations that arose from the secondary analysis of Dementia Champions data. The two KoD instruments used in cohort 10 were administered in a sample of undergraduate nursing students to explore, using IRT modelling techniques, how the instruments (and the items they contain) performed when administered in samples of healthcare students, compared to registered health and social care staff.

Ethical approval

Ethical approval for this study was granted by the School of Health Sciences (SHS) (formerly School of Nursing and Health Sciences [SNHS]) Research Ethics Committee. The letter of confirmation of ethical approval can be viewed in *Appendix 3*.

Sample

First-year nursing undergraduates were recruited, using a purposive sampling approach, from a nursing programme in one university in Scotland. Upon ethical approval, permission to access students was sought via the head of undergraduate studies as the gatekeeper. The sample consisted of 479 students who were enrolled in their first year of the 2019/2020 adult nursing, mental health nursing, or child nursing

pathway, across two campuses. The undergraduate students had been on their BSc course for four weeks and had undertaken no dementia-specific education at the time.

Data collection

With permission, participants were approached during a lecture (one at each campus) and were provided with verbal and written information about the PhD study, with a request that they complete the survey pack if they chose to participate.

Data were collected in the form of pencil and paper surveys for two reasons: i) to ensure the test conditions were comparable to the Dementia Champions evaluations, and ii) to maximise sample size, since data collection took place in lecture theatres that did not facilitate individual IT access. Further, requesting that participants use smartphones/personal devices to complete the survey may have resulted in bias or exclusion of those did not own (or bring) a smart device.

Survey packs were separate to the participant information sheets, and contained:

- Front cover with instructions
- Consent form
- 41 dementia knowledge questions (2 instruments)
- 5 questions on demographics and personal experience with dementia

As an incentive to participate, respondents were offered the chance to enter a prize draw for one of four £25 Amazon vouchers. To enter, respondents were required to provide an email address on the consent form. It was made clear both in verbal and written form that the consent forms would be detached and stored separately from the survey responses, and therefore confidentiality of responses would not be breached. The participant information sheet, consent form, and survey pack can be viewed in *Appendices 4, 5, and 6*.

Data entry

Survey packs were assigned a unique code number and checked for complete consent forms. Consent forms were then detached and stored in a separate file, away from the

responses. Data were entered using SPSS v.25; 15% of data were double entered to check for accuracy.

4.3 Dementia knowledge measurement instruments

Instrument 1: The Knowledge In Dementia scale (KIDE)

The Knowledge In Dementia scale (KIDE) Elvish *et al*. (2014) is a 16-item self-report instrument for assessing "knowledge of dementia" in healthcare staff. Item content is described in *Table 4.1*. The response for all items in the KIDE is binary: agree (1)/disagree (0). Items are reverse scored where the correct answer is disagree (0). Possible scores on the KIDE range from 0 – 16, with higher scores indicative of greater dementia knowledge.

Table 4.1: KIDE item content, scoring guide, and shortened item codes

	KIDE item content	**Item code**[c]
1	Permanent changes to the brain occur in most types of dementia	Brain changes
2	People who have dementia will usually show the same symptoms [b]	Same symptoms
3	Dementia can be caused by a number of small strokes	Strokes
4	Currently, most types of dementia cannot be cured	Incurable
5	When people with dementia walk around it is usually aimless [b]	Aimless walking
6	People with dementia will eventually lose all their ability to communicate [b]	Communicate
7	People with dementia who are verbally aggressive nearly always become physically aggressive [b]	Aggression
8	Brain damage is the only factor that is responsible for the way people with dementia behave [b]	Brain damage
9	It is possible to catch dementia from other people [b]	Contagious

10	My perception of reality may be different from that of a person with dementia	Perception
11	People with dementia never get depressed [b]	No depression
12	Anger and hostility occur in dementia mostly because the 'aggression' part of the brain has been affected [b]	Anger
13	Dementia is a general term which refers to a number of different diseases	Umbrella term
14	A person with dementia's history and background plays a significant part in their behaviour	History/behaviour
15	Physical pain may result in a person with dementia becoming aggressive or withdrawn	Pain
16	A person with dementia is less likely to receive pain relief than a person without dementia when they are in hospital	Analgesia

[a] Response range: agree (1) or disagree (0).

[b] Items 1, 4, 7, 8, 10, 12, 14, and 16 were reverse scored. Disagreement with these statements represents a correct answer/response.

[C] Codes that will be used to refer to items throughout results sections of this thesis

As described in the systematic review (*Chapter 3*), the KIDE scale was not subject to any psychometric validation techniques during development, however it contained potentially useful items. Psychometric evaluation of the scale, post-development, was also limited, with no published studies reporting on dimensionality and comparative validation studies. Although the KIDE was developed for hospital staff, studies by Lorio *et al*. (2017) and Jack-Waugh *et al*. (2018) reported ceiling effects in samples of this population and as such, the systematic review in this thesis recommended that the KIDE may prove more useful as a more basic knowledge test, for populations who would be expected to have lower levels of dementia knowledge than trained healthcare staff, for example, healthcare students, family caregivers, or lay populations.

The second KoD instrument had a more comprehensive background with regard to validation and reported psychometric properties, as follows.

Instrument 2: The Dementia Knowledge Assessment Scale (DKAS)

The dementia knowledge assessment scale (DKAS) (Annear *et al*., 2017) was developed and validated for use in healthcare staff and students, with items generated based on an international Delphi survey (Annear *et al*., 2015) and a literature review. It contains 25 statements that cover four domains of dementia knowledge, as follows: 1. Causes and characteristics; 2. Communication and engagement; 3. Care needs; and 4. Risks and health promotion.

The DKAS was originally developed with a modified Likert scale with five options for response: false, probably false, probably true, true, don't know. In order to reduce burden and for the response format to be consistent with the KIDE, response options were reduced to: True, False, Don't know. Possible scores on the DKAS range from 0 – 25, with higher scores indicative of greater dementia knowledge.

The DKAS is the only KoD instrument in current use to have examined and reported scale structure using exploratory and confirmatory factor analysis (EFA/CFA). A strong four-factor model (GFI = .974: RMSEA = .040) was reported in a large, diverse sample (n=3649) of health and care staff, healthcare students, family carers, and general population (Annear *et al*., 2017). The item content of the DKAS is reported in *Table 4.2*, and the confirmed four-factor structure as reported by the developers is displayed in *Figure 4.2*.

Post-development, the DKAS has been administered in predominantly Australian samples of healthcare staff as part of a massive open online course (MOOC) and has been subject to more rigorous psychometric testing and evaluation than the KIDE. In a comparative validation study (Annear *et al*., 2016) between the DKAS and the Alzheimer Disease Knowledge Scale (ADKS) (Carpenter *et al*., 2009) (see *Chapter 3, Table 3.3* for the reported psychometric properties), the DKAS was reported as superior with regards to parameters of response, minimisation of ceiling effects, and ability to discriminate between pre-and post-intervention scores.

Table 4.2. Item content and codes for the 25-item DKAS

Item	Item wording	**Item code**
1	Most forms of dementia do not generally shorten a person's life	Lifespan
2	Blood vessel disease (vascular dementia) is the most common form of dementia	Vascular
3	People can recover from the most common forms of dementia	Recover
4	Dementia is a normal part of the ageing process	Normal ageing
5	Dementia does not result from physical changes in the brain	Brain changes
6	Planning for end of life care is generally not necessary following a diagnosis of dementia	Planning
7	Alzheimer's Disease is the most common form of dementia	Alzheimer's
8	It is impossible to communicate with a person who has advanced dementia	Communication
9	A person experiencing advanced dementia will not generally react to changes in their physical environment	Environmental changes
10	It is important to correct a person with dementia when they are confused	Correcting dementia
11	People experiencing advanced dementia often communicate through body language	Body language
12	Uncharacteristic behaviours in a person experiencing dementia are usually a response to unmet needs	Unmet needs
13	Medications are the most effective way of treating behavioural symptoms of dementia	Medication
14	People experiencing dementia do not generally have problems making decisions	Decisions
15	Movement is generally affected in the later stages of dementia	Movement
16	Difficulty eating and drinking generally occurs in the later stages of dementia	Eating

Item	Item wording	**Item code**
17	People with advanced dementia may have difficulty speaking	Speaking
18	People experiencing dementia often have difficulty learning new skills	Skills learning
19	Daily care for a person with advanced dementia is most effective when it focuses on providing comfort	Comfort
20	Having high blood pressure increases a person's risk of developing dementia	Hypertension
21	Maintaining a healthy lifestyle does not reduce the risk of developing the most common forms of dementia	Lifestyle
22	Symptoms of depression can be mistaken for symptoms of dementia	Depression
23	The sudden onset of cognitive problems is characteristic of common forms of dementia	Sudden onset
24	Exercise is generally beneficial for people with dementia	Exercise
25	Early diagnosis of dementia does not generally improve quality of life for people experiencing the condition	Early diagnosis

Figure 4.2. Reported four-factor structure of the DKAS, taken from Annear et al. (2017)

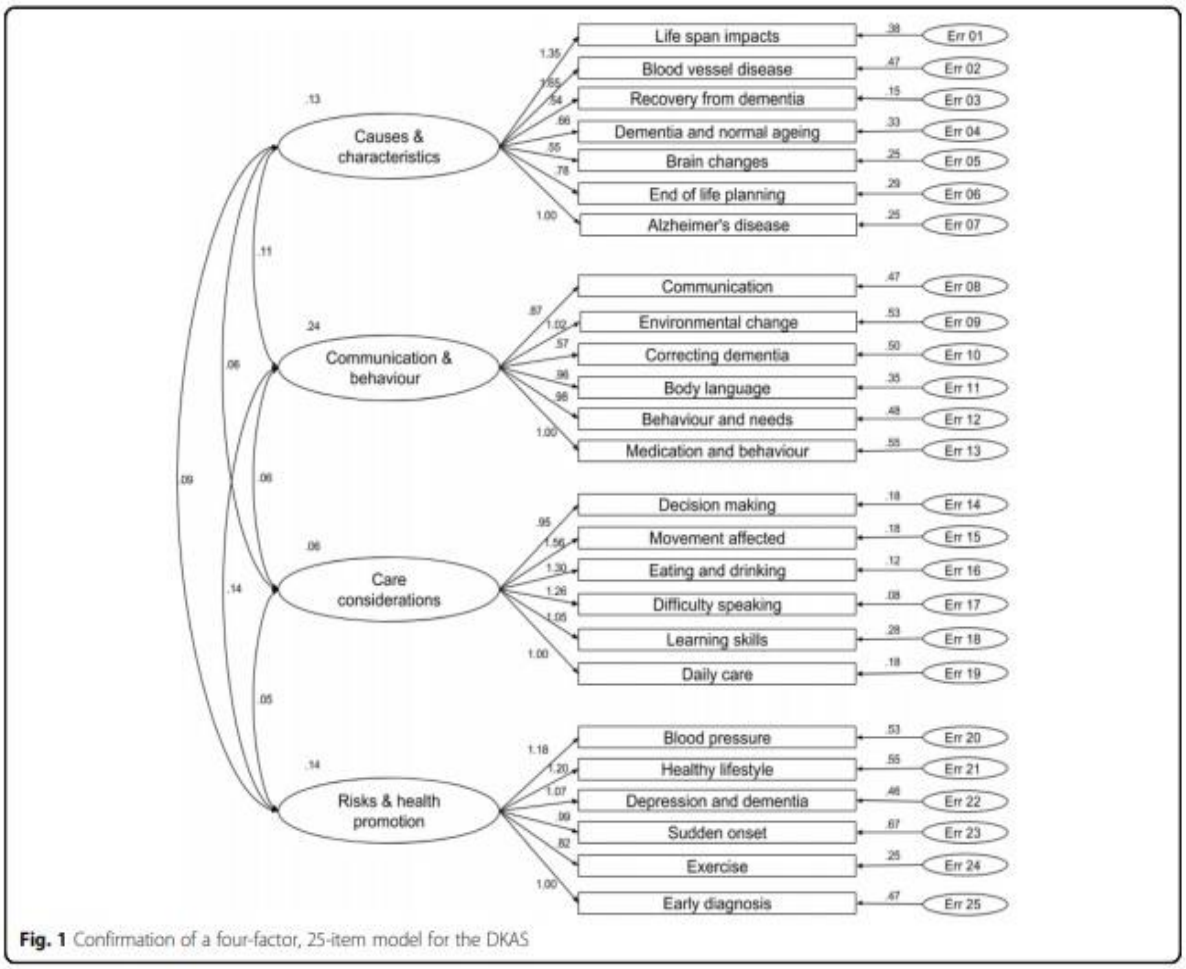


Fig. 1 Confirmation of a four-factor, 25-item model for the DKAS

4.4 Classical test theory results - Descriptive statistics

Traditional descriptive statistics for datasets one and two are reported here, to set the scene for the IRT analyses reported in chapters 5-7. Descriptive statistics are presented here under the classical test theory (CTT) framework of measurement, as was described in *Chapter 2* (Theoretical frameworks of measurement). Distributions of score frequencies were mapped graphically using SPSS v.25. Item analysis and reliability estimates were examined using R version 3.6.1.

4.4.1 Dataset one - Registered health and social care staff

KIDE responses from Dementia Champions participants

Across the five cohorts of the Dementia Champions programme, pre-intervention data were collected from 521 participants. Sample sizes of the individual cohorts are shown in *Table 4.3*; cohort data were pooled to form a combined sample for this PhD project. All cohort members/programme participants were qualified health and social care staff at the time of completing the Dementia Champions programme. The content of the programme was not changed or adapted between cohorts, and the intervention was conducted by the same team of academics in the same settings for each, therefore it was deemed acceptable to pool the data for all initial analysis.

Rows of data with any KIDE responses missing were removed. All analyses of this dataset are based on the pooled sample of five cohorts, pre-intervention, with missing data removed (n=395).

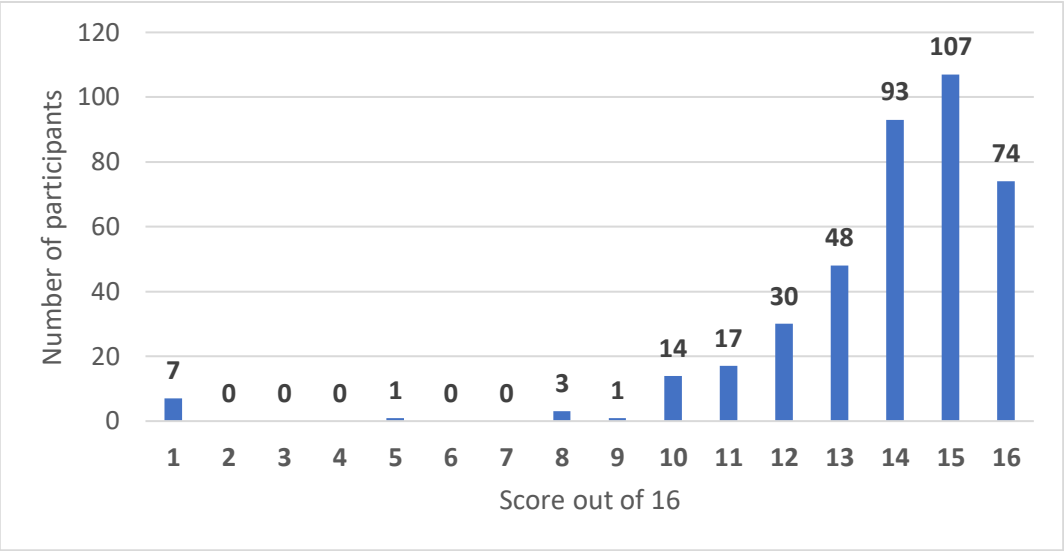
Table 4.3. Dementia Champions participants, cohorts 6 - 10, pre-intervention

Cohort 6	**Cohort 7**	**Cohort 8**	**Cohort 9**	**Cohort 10**
122	121	77	82	117

Figure 4.3 shows the distribution of total scores (out of a possible highest score of 16) from Dementia Champions responses to the KIDE. There were very few low

scores, apart from the seven scores of zero. Most scores were in the range 10 to 16 i.e., lay predominantly towards the higher end of the scale (with 74 at the maximum value). This is consistent with the ceiling effects reported in previous studies (Lorio *et al*., 2017; Jack-Waugh *et al*., 2018).

Figure 4.3. Distribution of KIDE scores for Dementia Champions participants (n=395)



Traditional item statistics are shown in *Table 4.4.* In the Dementia Champions sample, item difficulty indices for the KIDE ranged from 0.625 – 0.977. Items were relatively easy and UWS Dementia Champions answered many - but not all - correctly. Eight of the sixteen items had high (>0.90) endorsement frequencies, meaning they were very easy for almost all participants. Respondents with higher summed scores on the KIDE were more likely to answer more items correctly. Item discrimination indices were predominantly low, indicating very little difference in item response between the poorest and best performing respondents. Item correlations were acceptable except for those between three items: these were Communicate, History/behaviour, and Analgesia.

Table 4.4. Descriptive statistics for Dementia Champions participant responses to the KIDE

Item	Biserial correlation of item with total score	Biserial correlation of item with total score (when item removed) [a]	Item difficulty [b]	Item discrimination [c]
Brain changes	0.542	0.448	0.896	0.237
Same symptoms	0.440	0.347	0.919	0.145
Strokes	0.525	0.376	0.709	0.603
Incurable	0.537	0.437	0.884	0.267
Aimless walking	0.579	0.516	0.949	0.107
Communicate	0.412	0.262	0.780	0.298
Aggression	0.591	0.522	0.937	0.160
Brain damage	0.522	0.428	0.901	0.244
Contagious	0.675	0.639	0.975	0.069
Perceptions	0.543	0.468	0.934	0.176
No depression	0.733	0.704	0.977	0.069
Anger	0.432	0.307	0.851	0.282
Umbrella term	0.465	0.329	0.808	0.382
History/behaviour	0.436	0.271	0.696	0.519
Pain	0.580	0.508	0.932	0.176
Analgesia	0.403	0.225	0.625	0.466

[a] Guidance on minimum criterion for item-total correlation varies, though > 0.3 is generally acceptable

[b] Item difficulty indices should be between 0.5 – 0.95 to contribute to effective measurement range

[c] Item discrimination indices are ‘good’ > 0.3, ‘fair’ 0.1-0.3, ‘poor’ <0.1

Internal consistency reliability of the KIDE in this sample was within acceptable range with a KR-20 estimate of 0.77.

4.4.2 Dataset two – Undergraduate nursing students

Data were collected from 404 undergraduate students across two campuses. The participants had been on their BSc course for one month, and had undertaken study on general nursing theory only, no clinical placement. Across campuses there were a total of 479 students, 363 at campus one and 116 at campus two. Permission was granted for the author of this thesis[2] to access the students at the end of two lectures (one at each campus), where the lecture content ended 30 minutes before the class was scheduled to finish. At campus one, 342 students attended the lecture and 318 agreed to participate in the study. Four survey packs contained incomplete consent forms and were therefore not useable, leaving 314 completed survey packs. At campus two, 98 students attended the lecture and 93 agreed to participate. There were nine survey packs with either no consent form or more than half of the questions unanswered; these were removed, leaving 84 completed survey packs from this campus. Overall, the response rate was very strong, at 92% (93% for campus one, and 95% for campus two).

The participants were predominantly female (92%)[3], aged between 18-25 (60%), and did not have degree-level education prior to their current course (83%). Further, 64% of participants reporting knowing someone with dementia, and 52% reported having previously worked with people living with dementia. See *Table 4.5* for demographic information and sample characteristics.

Consistent with dataset one, cases with missing data were removed since a number of the statistical analyses used throughout this thesis do not accept missing data. Therefore, the final sample was (n=384).

[2] It should be noted that at the time of writing, the author (CG) is a staff member at the University where these data were collected, however at the time of data collection she was a PhD student with no formal connections to the undergraduate student body.
[3] This was expected due to the nature of nursing as a predominantly female profession

Table 4.5. Undergraduate nursing student sample characteristics

Sample characteristics	**Undergraduate nurse sample (n=384)**	
Age		
18-25	232 (60%)	
Over 25	144 (37%)	
Prefer not to say	8 (3%)	
Gender		
Female	353 (92%)	
Male	31 (8%)	
Prefer not to say	Nil	
Qualifications		
School level	320 (83%)	
Prior **University degree**	62 (16%)	
Prefer not to say	3 (1%)	
	Yes	**No**
Do you know or have you known someone who lives with dementia?	246 (64%)	137 (36%)
Do you work or have you ever worked with people who live with dementia?	199 (52%)	185 (48%)

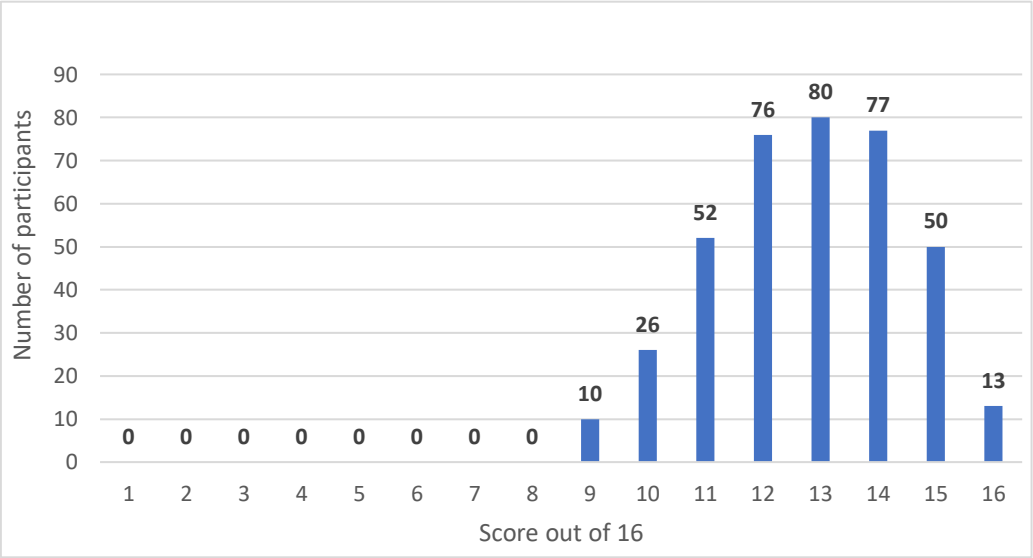
Undergraduate nursing students' responses to the KIDE

KIDE – Descriptive statistics

Figure 4.4 shows the distribution of total scores (out of a possible highest score of 16) from undergraduate nursing student responses to the KIDE. The distribution of scores was marginally more normal than KIDE responses in dataset one, though

again predominantly towards the higher end of the scale. All scores were in the range of 9-16 out of 16. Every participant scored higher than 50% on the KIDE scale.

Figure 4.4. Distribution of KIDE scores for undergraduate nursing participants (n=384)



Traditional item statistics are shown in *Table 4.6.* In the undergraduate nursing student sample, item difficulty indices for the KIDE ranged from 0.297 – 0.995. Five of the sixteen items had high (>0.90) item difficulty estimates, meaning they were very easy to answer. Only one item had a very low endorsement frequency, being item 16: "*A person with dementia is less likely to receive pain relief than a person without dementia when they are in hospital*" (code: analgesia). Interestingly, item 16 was also the most difficult item in the Dementia Champions sample, therefore perhaps raising questions about item wording and comprehension, given the majority of other items were found to be very easy in both samples. Item discrimination indices were predominantly low; all fell into the 'fair' or 'poor' categories. Biserial correlations between items were generally low, with the majority falling below the acceptable value of 0.3.

Table 4.6. Descriptive statistics for undergraduate nurse participant responses to the KIDE

Item	Biserial correlation of item with total score	Biserial correlation of item with total score (when item removed) [a]	Item difficulty[b]	Item discrimination[c]
Brain changes	0.066	0.031	0.953	-0.060
Same symptoms	0.332	0.258	0.818	0.107
Strokes	0.345	0.352	0.747	0.091
Incurable	0.134	0.055	0.977	0.044
Aimless walking	0.344	0.266	0.836	0.130
Communicate	0.327	0.336	0.638	0.043
Aggression	0.341	0.320	0.833	0.125
Brain damage	0.288	0.172	0.865	0.087
Contagious	0.155	0.031	0.990	0.095
Perception	0.144	0.063	0.974	0.050
No depression	0.120	0.016	0.995	0.077
Anger	0.308	0.273	0.557	0.012
Umbrella term	0.460	0.531	0.688	0.203
History/behaviour	0.381	0.367	0.729	0.124
Pain	0.266	0.180	0.883	0.076
Analgesia	0.409	0.414	0.297	0.148

[a] Guidance on minimum criterion for item-total correlation varies, though > 0.3 is generally acceptable

[b] Item difficulty indices should be between 0.5 – 0.95 to contribute to effective measurement range

[c] Item discrimination indices are ‘good’ > 0.3, ‘fair’ 0.1-0.3, ‘poor’ <0.1

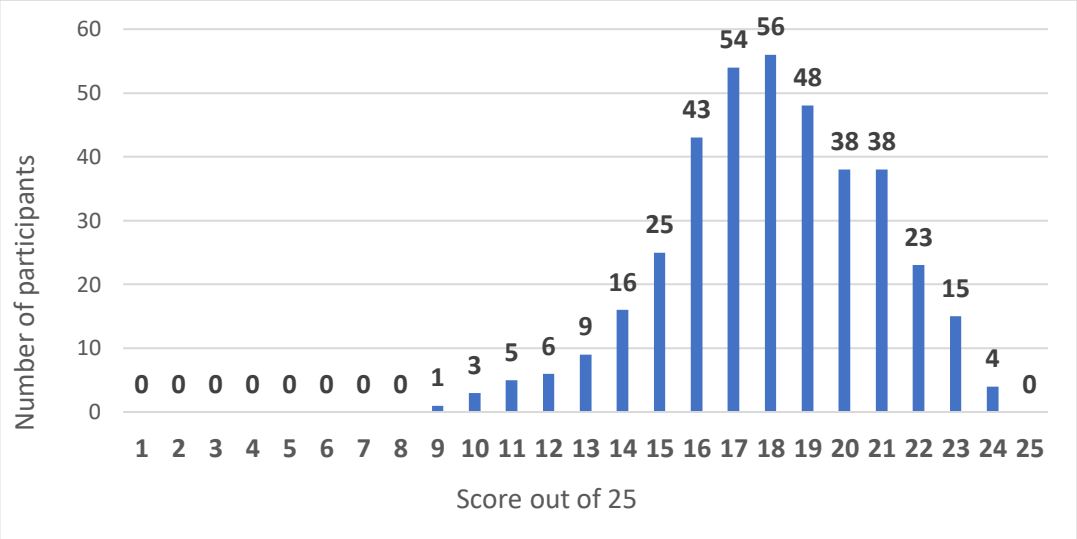
Internal consistency reliability of the KIDE in this sample was unacceptably low with a KR-20 estimate of 0.29.

Dataset two – undergraduate nursing students' responses to the DKAS

DKAS - Descriptive statistics

Figure 4.5 shows the distribution of total scores (out of a possible highest score of 25) from undergraduate nursing student responses to the DKAS. The distribution of scores was reasonably even, though predominantly towards the higher end of the scale. Most scores were in the range of 15 – 22 out of 25. The lowest score achieved was 9 out of the possible 25, and no participants achieved 100% correct responses.

Figure 4.5. Distribution of DKAS scores for undergraduate nursing students (n=384)



Traditional item statistics are shown in *Table 4.7.* In the undergraduate nursing student sample, item difficulty indices for the DKAS ranged from 0.240 – 0.964. Only three of the 25 items had high (>0.90) item difficulty estimates, meaning that only a few items were very easy to answer. Further, three of the 25 items had low (<0.50) difficulty estimates, meaning they very difficult to answer in this sample. As with the KIDE responses, all item discrimination indices fell into the 'fair' or 'poor'

categories, and correlations between items were predominantly too low to be deemed acceptable.

Table 4.7. Descriptive statistics for undergraduate nurse participant responses to the DKAS

Item	Biserial correlation of item with total score	Biserial correlation of item with total score (when item removed)[a]	Item difficulty[b]	Item discrimination[c]
Lifespan	0.385	0.398	0.615	0.226
Vascular	0.140	0.203	0.513	-0.036
Recover	0.078	0.031	0.964	0.013
Normal ageing	0.188	0.141	0.870	0.071
Brain changes	0.251	0.242	0.841	0.126
Planning	0.281	0.242	0.656	0.117
Alzheimer's	0.186	0.172	0.729	0.030
Communication	0.199	0.227	0.797	0.058
Environmental changes	0.372	0.359	0.729	0.226
Correcting dementia	0.241	0.203	0.799	0.102
Body language	0.362	0.242	0.865	0.251
Unmet needs	0.405	0.375	0.701	0.258
Medication	0.224	0.203	0.732	0.070
Decisions	0.289	0.180	0.870	0.176
Movement	0.333	0.258	0.818	0.205
Eating	0.348	0.211	0.893	0.247
S**peaking**	0.356	0.203	0.906	0.262
S**kills** learning	0.240	0.227	0.818	0.107
Comfort	0.166	0.109	0.917	0.070
Hypertension	0.402	0.445	0.299	0.254
Lifestyle	0.390	0.430	0.432	0.228
Depression	0.253	0.266	0.625	0.085

Sudden onset	0.124	0.109	0.240	-0.026
Exercise	0.382	0.266	0.831	0.261
Early diagnosis	0.347	0.391	0.544	0.180

[a] Guidance on minimum criterion for item-total correlation varies, though > 0.3 is generally acceptable

[b] Item difficulty indices should be between 0.5 – 0.95 to contribute to effective measurement range

[c] Item discrimination indices are 'good' > 0.3, 'fair' 0.1-0.3, 'poor' <0.1

Internal consistency reliability of the DKAS in this sample was on the low end of adequate, with a KR-20 estimate of 0.504.

4.5 Summary

This chapter has introduced the two datasets used in the empirical chapters of this thesis. The first dataset was acquired through a cross-institutional academic collaboration, and the second was generated based on results and recommendations from dataset one. Methods of data collection, data entry, and ethical approval have been described, for transparency. The development procedures and content of each of the two KoD instruments have been detailed, as well any psychometric evaluation that took place post-development. Finally, descriptive statistics have been reported under the classical test theory (CTT) framework for both datasets, in order to set the scene for the item repsonse theory analyses reported throughout chapters 5-7.

Despite being developed specifically for use in professional healthcare staff, the KIDE scale performed poorly in the Dementia Champions educational programme, with ceiling effects pre-intervention evident to the extent that increases in dementia knowledge could not be measured post-intervention. A hypothesis was formed that the KIDE scale may perform better in those who would be expected to have less knowledge, however results from dataset two demonstrated that this was not the case, with ceiling effects prominent in undergraduate nurisng students, despite the sample being only one month into their degree course, and having had no dementia-specific education at the time. One explanantion for this may be that the majority of

undergraduate sample had both known and worked with people with dementia in the past, with 64% and 52% having known and worked with this population, respectively.

The DKAS instrument was introduced to the Dementia Champions programme to support and compare with the KIDE results. The DKAS was selected based on the results of the systematic review of psychometric properties of dementia knowledge measurement instruments (*Chapter 3*). The findings from dataset two showed that the DKAS did indeed perform better in the undergraduate nurse sample, with no evidence of ceiling effects. This was also the case in cohort 10 of the Dementia Champions programme[4].

All findings here have been reported under the CTT frawework of 'test level' analysis, to examine how the KIDE and thr DKAS performed as *entire measurement instruments* in samples of healthcare professionals and undergraduate student nurses. As described in *Chapter 2*, item repsonse theory (IRT) facilitaites the examination of relationships between levels of knowledge and participants' responses to *individual questions* within a test; this facilitates a more in-depth analysis the measurement precision and effective measurement range of tests and the items they contain.

The following three chapters (5-7) will report on the application and calibration of the Rasch model to the two dementia knowledge datasets that have been introduced here.

[4] Data not included in this thesis. Manuscript accepted for publication pending minor revisions, due to be resubmitted by 29/09/21.

CHAPTER 5

Rasch modelling of dichotomous dementia knowledge data: five cohorts of dementia champions responses **to** *the knowledge in dementia (KIDE) scale*

5.1 Introduction

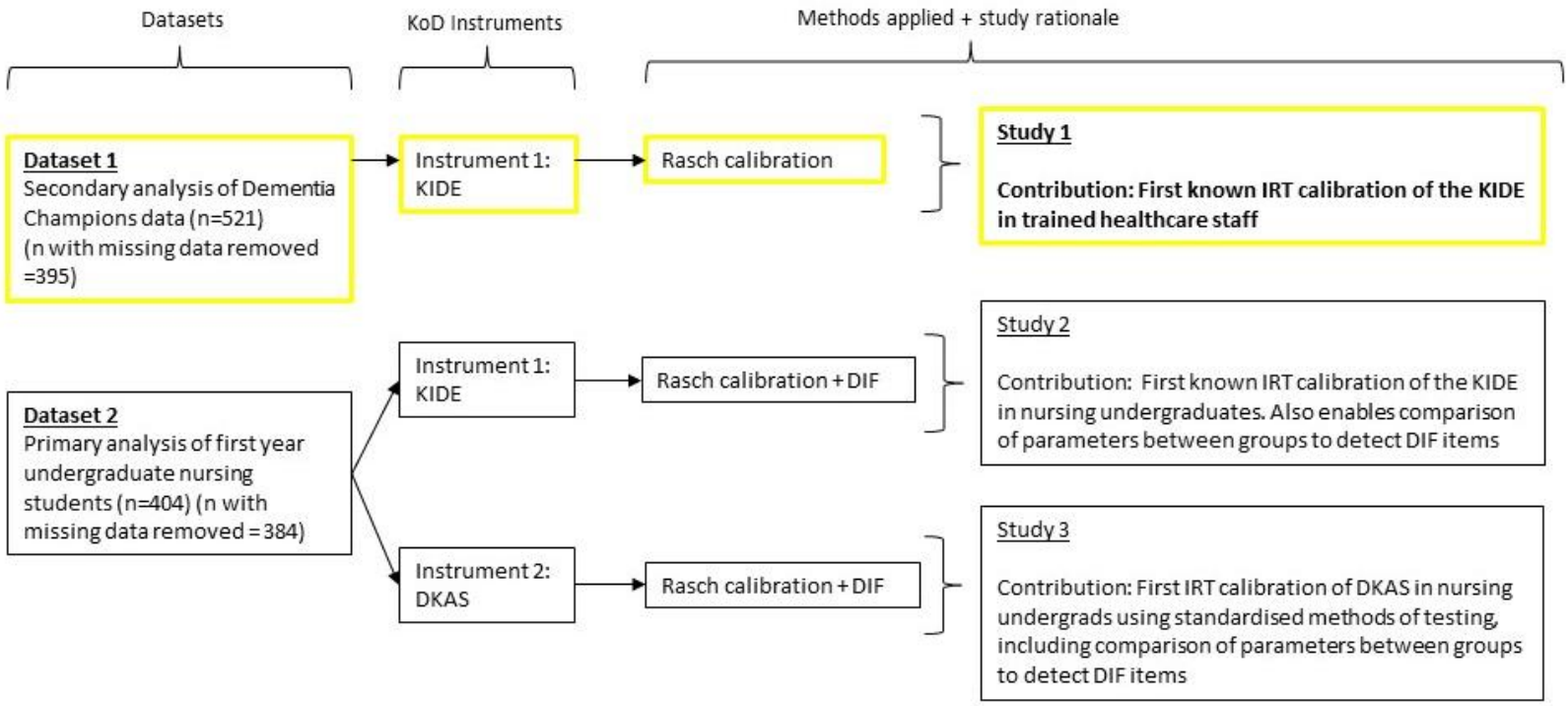
Treatment and care for people living with dementia is provided by a wide range of healthcare staff. With the increasing prevalence of dementia in primary, secondary and community settings, workforce education for post-qualifying staff is required, though current opportunities are scarce (Macrae et al., 2019). Key to the perceived success of educational interventions is the availability and application of appropriate measurement instruments to ascertain whether interventions change knowledge levels, and by how much.

As discussed in *Chapter 3*, knowledge of dementia (KoD) instruments in current use have been developed using classical test theory (CTT) methods whereby responses to dichotomous items are scored 0 and 1, and the sum of responses is *assumed* to represent the respondent's true score (meaning their ability in relation to KoD). Item response theory (IRT) models can be used to facilitate more rigorous development and examination of measurement instruments, as well as more sophisticated estimation of parameters for persons (respondents) and items. The simplest and strongest IRT model for unidimensional binary response data is arguably the Rasch model (Andrich and Marias, 2019).

5.1.1 Chapter Aim

The KoD instrument central to this study was developed using CTT (sum score) methods; this analysis is the first examination of the Knowledge in Dementia (KIDE) scale (Elvish *et al*., 2014) item responses using IRT scoring methods, specifically Rasch model calibration. The methods used in this chapter are highlighted in the schematic below (*Figure 5.1*).

Figure 5.1. Methods schematic. The highlighted pathway shows the dataset, instrument, and methods covered in this chapter.



This chapter details the application and calibration of the Rasch model to estimate item parameters, person parameters, and model fit of the Dementia Champions data, a multi-sample initiative that recruited separate groups of post-qualifying healthcare staff and developed their knowledge. The objective was to apply and evaluate a simple but powerful psychometric model: the Rasch model, for its suitability to calibrate KIDE dementia knowledge items on a unidimensional continuum in collaboration with the UWS team who organised and ran the Dementia Champions Programme.

The results form the first empirical contribution in this thesis and provide information in relation to the effective measurement range and targeting of the KIDE items on the continuum measured by the instrument and in relation to the individuals who responded to all the KIDE questions. To the author's knowledge, this is the first application of an IRT model to the KIDE scale.

5.2 Methods - Rasch calibration of dataset one

The Rasch model was fit to the KIDE data using a conditional maximum likelihood (CML) approach using the eRM package in R (Mair, Hatzinger, and Maier, 2020) and Stata version 16.1. In the Rasch model, total scores are sufficient to estimate a person's position on the latent scale. CML is an appropriate estimation technique when using the Rasch model as it facilitates estimation of item difficulty parameters independent of theta, but instead from the probabilities of a variety of response patterns that amount to the same total score (Irwing, Booth, and Hughes, 2018). Item and person parameters were estimated independently. The magnitude of the p-values for Andersen's LR tests (model fit) and Wald tests (item fit) were used to examine for misfit. Graphics were generated using the eRM package and the TAM package (Robitzsch, Kiefer & Wu, 2019).

The fundamental assumptions underlying the Rasch model include unidimensionality, parallel and non-intersecting item characteristics curves, and local independence (see *Chapter 2, section 2.3* for more information on these assumptions). If data fit the Rasch model sufficiently then all three assumptions can be considered fulfilled (Mair, 2018).

5.3 Results

As described in Chapter 4, the pooled sample of Cohorts 6-10 of the Dementia Champions programme comprised of (n=521) healthcare professionals. All analyses in this section are based on the pooled sample with missing data removed (n=395).

5.3.1 Rasch model calibration – CML estimation results

Rasch model calibration was performed using the CML method. Estimation using the CML method converged after 26 iterations of computation with a conditional log-likelihood value of -1250.66. These results were then replicated in the Stata-based analyses.

5.3.2 Calibration of item parameters

The first step in Rasch model calibration was to determine KIDE item parameters, to examine the distribution of items across the latent trait continuum of dementia knowledge.

Table 5.1 reports item parameter estimates and standard errors for KIDE items under a Rasch model. The difficulty parameter (*b*) estimates covered a wide range of the latent trait, from -2.92 (least difficult) to 2.05 (most difficult). The easiest item in the KIDE was 'People with dementia never get depressed' (item code: depression) with 98% correct. The most difficult item in the KIDE was 'A person with dementia is less likely to receive pain relief than a person without dementia when they are in hospital' (item code: analgesia), with 63% correct. *Table 1* also includes standard errors (se) for all parameters. The standard errors were acceptable for the majority of items, suggesting that there were likely no problems with the model estimation.

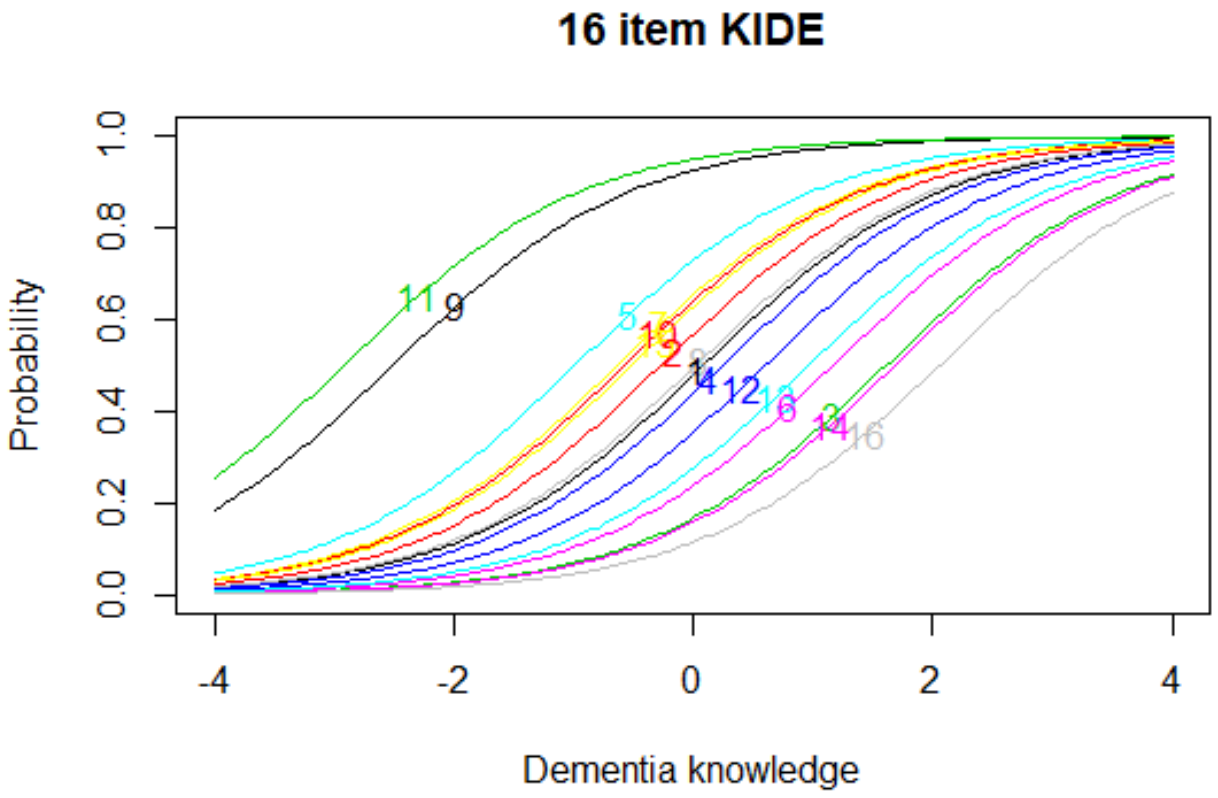
However, two items, 'It is possible to catch dementia from other people' (item code: contagious) and 'People with dementia never get depressed' (item code: depression) had markedly larger standard errors than other items. These were the two easiest KIDE items and were answered correctly by almost all participants. These large standard errors reflect the limited information provided by these items in the context of the KIDE scale.

Table 5.1. CML estimations of item difficulty with associated standard errors

	Kide item	*b* Difficulty (se)
1	Permanent changes to the brain occur in most types of dementia	0.09 (0.20)
2	People who have dementia will usually show the same symptoms	-0.27 (0.21)
3	Dementia can be caused by a number of small strokes	1.61 (0.13)
4	Currently, most types of dementia cannot be cured	0.25 (0.18)
5	When people with dementia walk around it is usually aimless	-0.99 (0.28)
6	People with dementia will eventually lose all their ability to communicate	1.17 (0.14)
7	People with dementia who are verbally aggressive nearly always become physically aggressive	-0.64 (0.24)
8	Brain damage is the only factor that is responsible for the way people with dementia behave	0.01 (0.19)
9	It is possible to catch dementia from other people	-2.51 **(0.55)**
10	My perception of reality may be different from that of a person with dementia	-0.58 (0.24)
11	People with dementia never get depressed	-2.92 **(0.67)**
12	Anger and hostility occur in dementia mostly because the 'aggression' part of the brain has been affected	0.60 (0.16)
13	Dementia is a general term which refers to a number of different diseases	0.97 (0.15)
14	A person with dementia's history and background plays a significant part in their behaviour	1.68 (0.13)
15	Physical pain may result in a person with dementia becoming aggressive or withdrawn	-0.52 (0.23)
16	A person with dementia is less likely to receive pain relief than a person without dementia when they are in hospital	2.05 (0.13)

The item difficulty parameters are displayed graphically in *Figure 5.2*. Examination of the joint ICCs shows pairs of overlapping curves, for example, items 3 and 14, shown in green and lilac towards the right-hand side of the plot. These items (strokes and history/behaviour) had similar difficulty parameters, at 1.61 and 1.68 respectively. Overlap suggests redundancy of item content. Note also overlapping items 7, 10 and 15, here shown in red and yellow near the centre of the plot. These items (Aggression, Perceptions, and Pain) had very similar difficulty parameter estimations, estimated to be calibrated at -0.64, -0.58, and -0.52, respectively.

Figure 5.2. Joint ICCs for the 16 item KIDE. Items are labelled by number as per the running order of the KIDE.



Further examination of the joint ICCs (*Figure 5.3*) showed a proportion of the latent trait that was unrepresented by item content in the KIDE, with no items representing theta levels of approximately -2.0 to -3.5. In this respect, the measurement precision of the KIDE was reduced in this population.

Figure 5.3: An unrepresented section of the latent trait can be seen in the gap between items 9 and 5 at the probability of 0.5.

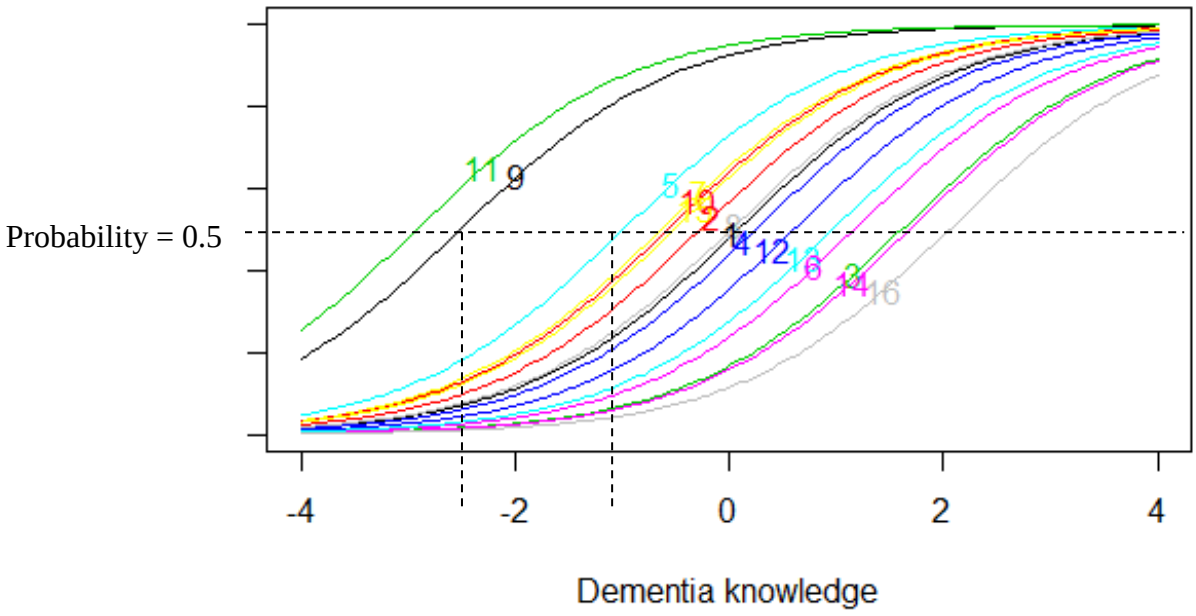
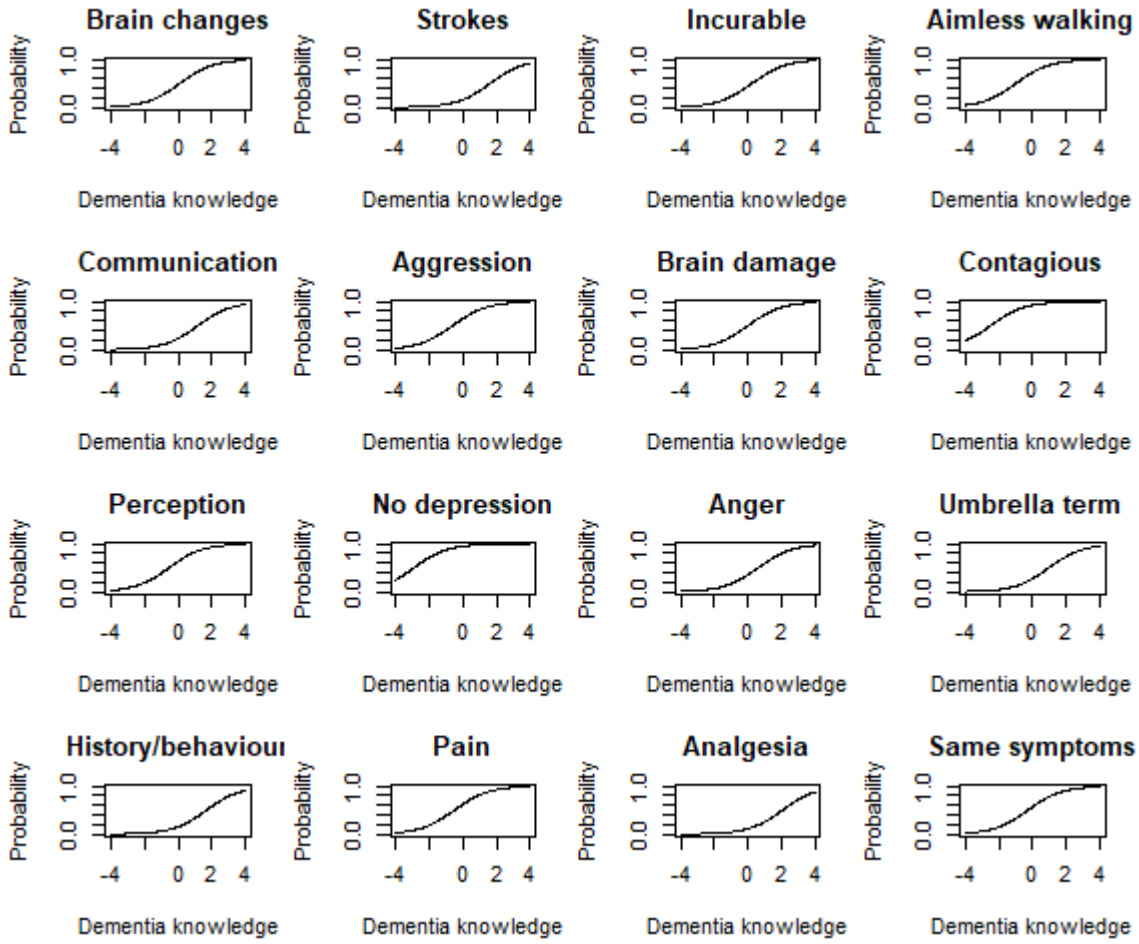


Figure 5.4 shows the ICCs individually. The majority of items had clear s-shaped curves, as would be expected. Two items (Contagious and No depression) had curves that were missing the left-hand tail of the s-shape due to the ceiling effects evident in this population; this also resulted in the larger standard errors discussed above.

Figure 5.4: Individual ICCs for the 16-item KIDE.



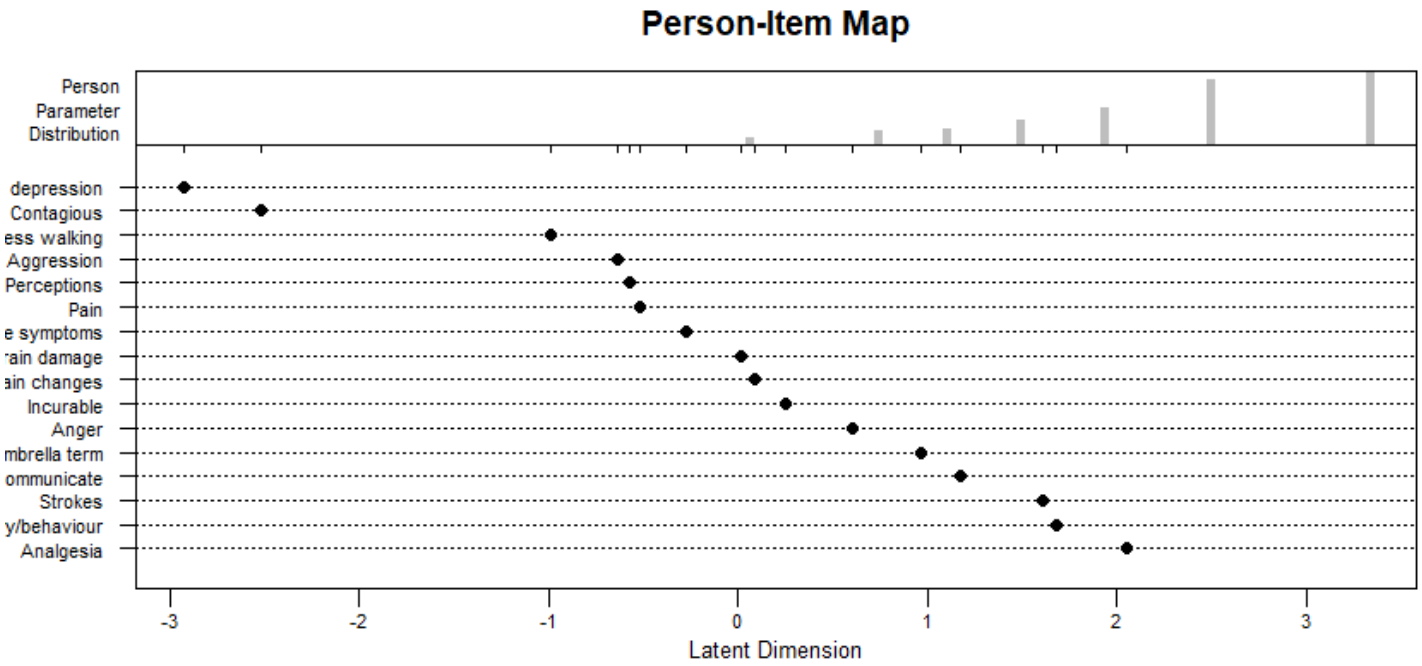
Rasch calibration of item parameters resulted in parallel and non-intersecting curves; this was the first indication that the data had sufficient fit to the model. KIDE item locations, as seen above, cover a significant portion of the latent trait of dementia knowledge, however there were notable gaps in the coverage of the latent trait continuum, with no items representing theta estimates of -2.0 to -3.5, or above $\theta = 1$.

The next step in the sequence of Rasch model calibration was to estimate the latent distribution, or person parameters, to examine the distribution of person ability along the latent trait continuum, and the relationship between person ability and item parameters for the KIDE.

5.3.3 Calibration of person parameters

The person-item map *(Figure 5.5)* shows the distribution of persons and items along the KoD latent trait continuum. Item distribution is shown in the lower panel and person distribution in the upper panel. Examination of the location of items relative to persons showed that ten of the KIDE items (those to the left of the person parameter bars, θ estimates of < 0.75) were not appropriate to the ability range of the respondents, therefore suggesting poor targeting of measurement. The lack of item thresholds within the person estimates at the positive end of the scale ($\theta = 2$ - 3.5 on the latent dimension) demonstrated further that the KIDE scale was not sufficient to capture the true ability of the Dementia Champions participants.

Figure 5.5: Person and item parameters for the 16-item KIDE



5.3.4 Fit statistics

To test for the assumption of local independence, the non-parametric testing framework of Ponocny (2001) was used. 13 of 105 item pairs tested showed local dependence. When this was evaluated globally, the results were less favourable. Local independence did not hold at a global level ($p<0.05$), indicating that by strict statistical testing the Rasch model would not fit the full KIDE-16 dataset.

5.3.5 Model fit

Parameters based on the fit of separate subgroups must be approximately the same to verify model fit. Given this dataset did not contain demographic variables, examination had to be performed on groups defined by the sum score (i.e. based on median and mean subgroup splits). Application of the LR test (Andersen, 1973) splitting the sample according to the mean score gave a significant result ($p < 0.05$) indicating that likelihoods differ across these two groups, therefore violating the assumptions of the Rasch model. One item (No depression) was excluded by the test due to its inappropriate response pattern.

The LR test with a median score split gave a non-significant result ($p = 0.181$). This assumption of the Rasch model was upheld. However, to achieve this result it was necessary to reduce the scale, as four items were removed (Aggression, Contagious, No depression, Pain) due to inappropriate response patterns.

5.3.6 Item fit

Wald tests were performed to examine the fit of items and identify which item/s were responsible for the misfit of the model; the same split criteria were used as in the Andersen LR tests above. Using the mean score split, two items (Same symptoms and Communicate) did not fit which can be seen by the significant p values highlighted in *Table 5.2*. Using the median score split, two items did not fit the model (Same symptoms and Strokes). These misfitting items were in addition to those items already removed by the Andersen LR tests due to inappropriate response patterns.

Table 5.2: Item fit indices using mean and median split Wald tests

Item	Mean split		Median split	
	z-statistic	p-value	z-statistic	p-value
Brain changes	-0.920	0.358	-0.405	0.686
Same symptoms	2.532	**0.011**	2.200	**0.028**
Strokes	-1.842	0.065	-2.083	**0.037**
Incurable	-1.138	0.255	-0.718	0.473
Aimless walking	0.913	0.361	0.704	0.482
Communicate	2.137	**0.033**	1.024	0.306
Aggression	-0.608	0.543	NA	NA
Brain damage	-1.478	0.140	-0.872	0.383
Contagious	0.291	0.771	NA	NA
No depression	NA	NA	NA	NA
Perceptions	-0.756	0.450	-0.749	0.454
Anger	0.605	0.545	0.809	0.418
Umbrella term	0.202	0.840	0.778	0.436
History/behaviour	0.673	0.501	0.988	0.323
Pain	-0.897	0.370	NA	NA
Analgesia	1.415	0.157	0.010	0.992

Using the mean split criteria, the Rasch model had adequate fit to 13 of 16 items, however the model did not fit on a global level, as evidenced by the Andersen LR test. Therefore, the examination of fit indices was carried forward using the median split criteria which showed global model fit and 10 of 16 items with adequate fit. Of the six misfitting items, five had very low difficulty parameters, these being easy items that were answered correctly by almost all participants, as follows:

Item 2 (Same symptoms) - ‘People who have dementia will usually show the same symptoms’ (**92%** correct)

Item 7 (Aggression) - ‘People with dementia who are verbally aggressive nearly always become physically aggressive’ (**94%** correct)

Item 9 (Contagious) - ‘It is possible to catch dementia from other people’ (**97%** correct)

Item 10 (No depression) - ‘People with dementia never get depressed’ (**98%** correct)

Item 15 (Pain) - Physical pain may result in a person with dementia becoming aggressive or withdrawn (**93%** correct).

The final misfitting item: ‘Dementia can be caused by a number of small strokes’ (item code: Strokes) was a more difficult item, with 71% of participants answering correctly. Therefore this item may have had potential within the KIDE content given the significant ceiling effects evident in almost a third of the items. As such, the item ‘Strokes’ was reintroduced and tests for model and item fit repeated.

The 11-item set showed adequate model fit with an Andersen LR test (median split) p-value of 0.26 and all items fit adequately according to the Wald tests. The item ‘Same symptoms’ (being the most difficult still- discarded item) was then reintroduced for examination of a 12-item set, however when this was reintroduced two items showed misfit according to the Wald tests. Therefore, 11 items that showed sufficient global model fit and item fit were retained and ‘Same symptoms’ was discarded.

Another method used in Rasch model calibration to identify misfit of items is to use graphical illustrations. Graphical checks for item fit were examined for the discarded item ‘Same symptoms’. This is shown in *Figure 5.6*. Items with sufficient fit were expected to fall along the diagonal line and within the confidence bands. The discarded item, ‘Same symptoms’, fell directly on the outer edge of the confidence bands. To demonstrate a comparison, ‘Analgesia’, a well-fitting item, fell directly on the central fit line (*Figure 5.7*).

Figure 5.6: Graphical model check for item fit for discarded item. The grey lines represent a 95% confidence interval.

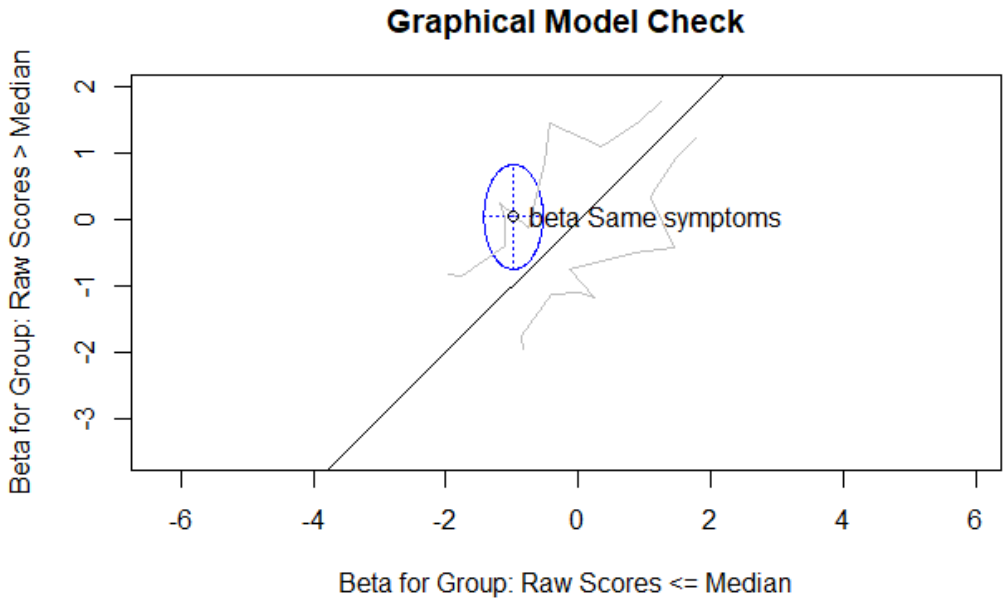
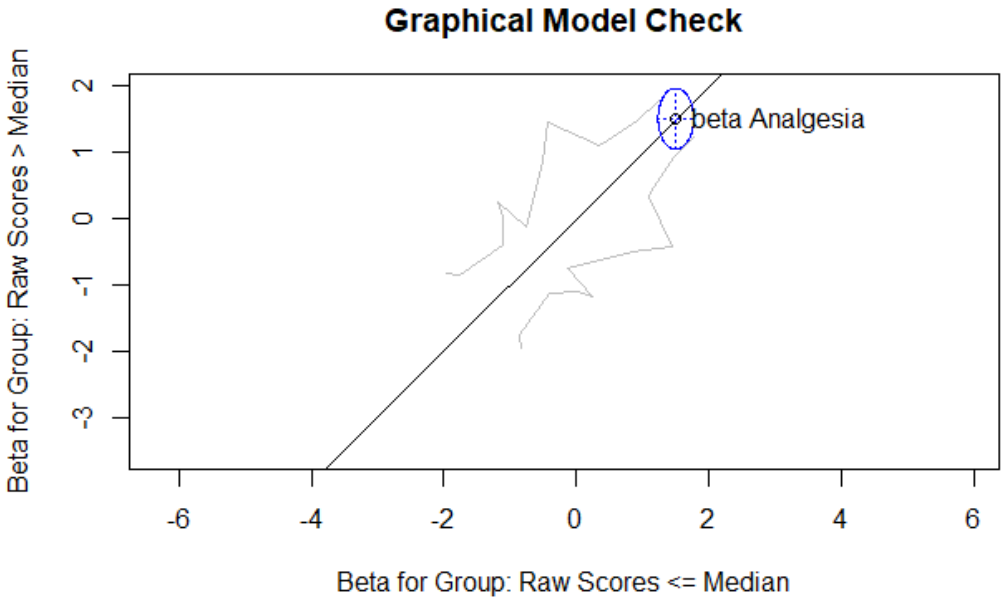


Figure 5.7: Graphical model check for item fit - highly fitting item. The grey lines represent a 95% confidence interval.



The fit of items and persons were also examined using Bond and Fox (2007) pathway maps. The item map shows the location of items against their infit t-statistics. *Figure 5.8* shows that all 11 items were within the recommended -2 to +2 values indicated by the vertical green lines.

Figure 5.8: Pathway map showing items against their infit t-statistics.

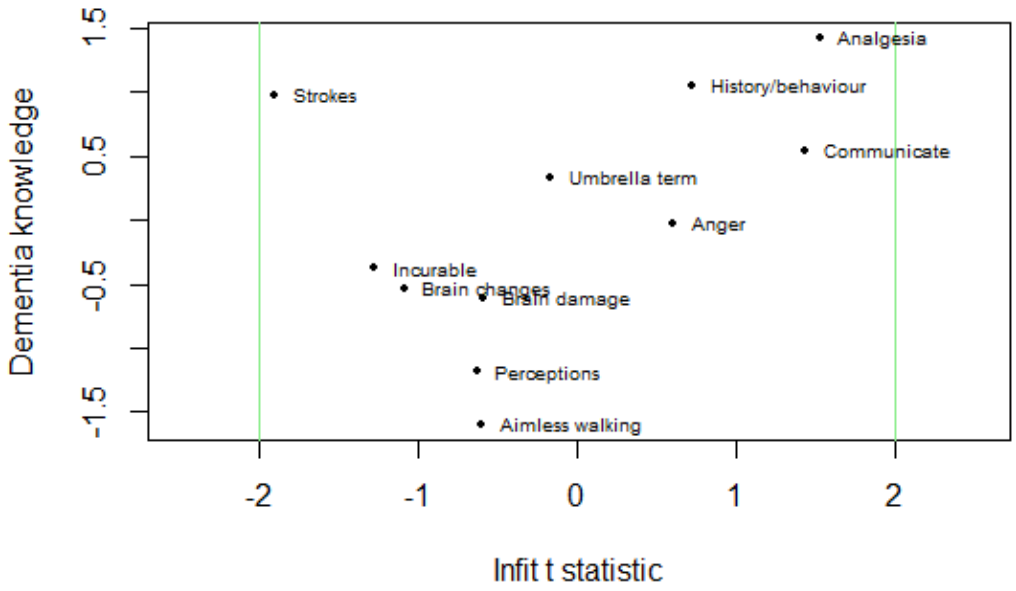
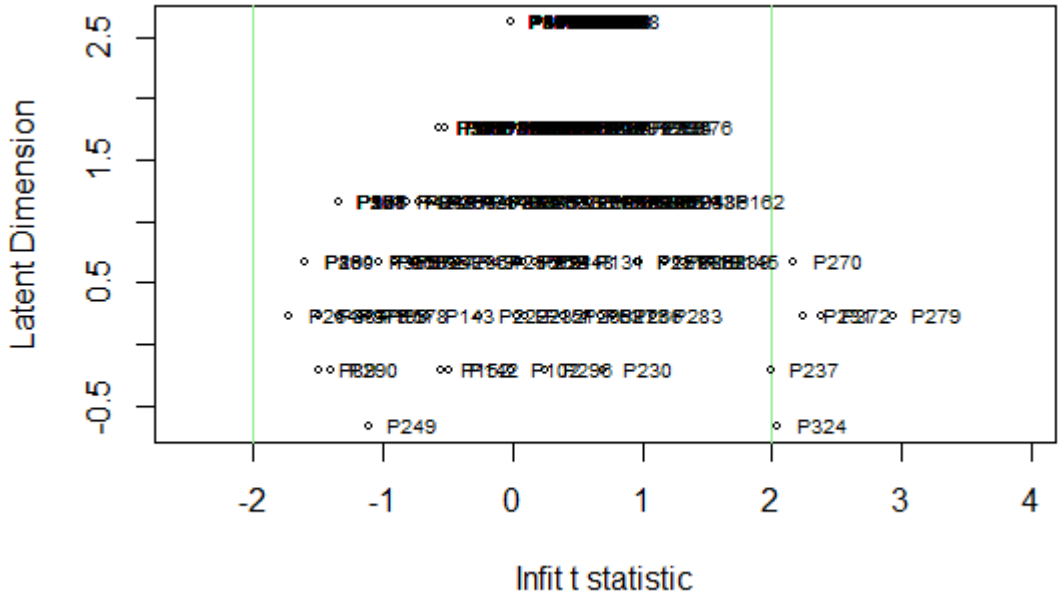


Figure 5.9 shows the pathway map for person fit indices. Of the 395 respondents, only six fell outside of the recommended fit statistics for the 11-item KIDE.

Figure 5.9: Pathway map showing persons against their infit t-statistics.



5.3.7 KIDE-11

Application and calibration of the Rasch model in the Dementia Champions dataset resulted in 11 out of 16 KIDE items being retained, based on the model and item fit statistics. Item content for the KIDE-11 is displayed in *Table 5.3*.

Table 5.3: Item content for the KIDE-11. Items discarded due to misfit to the Rasch model have been greyed out.

	KIDE Item[a]	Item code[c]
1	Permanent changes to the brain occur in most types of dementia	Brain changes
	People who have dementia will usually show the same symptoms [b]	Same symptoms
2	Dementia can be caused by a number of small strokes	Strokes
3	Currently, most types of dementia cannot be cured	Incurable
4	When people with dementia walk around it is usually aimless [b]	Aimless walking
5	People with dementia will eventually lose all their ability to communicate [b]	Communicate
	People with dementia who are verbally aggressive nearly always become physically aggressive [b]	Aggression
6	Brain damage is the only factor that is responsible for the way people with dementia behave [b]	Brain damage
	It is possible to catch dementia from other people [b]	Contagious
7	My perception of reality may be different from that of a person with dementia	Perception
	People with dementia never get depressed [b]	No depression
8	Anger and hostility occur in dementia mostly because the 'aggression' part of the brain has been affected [b]	Anger
9	Dementia is a general term which refers to a number of different diseases	Umbrella term
10	A person with dementia's history and background plays a significant part in their behaviour	History/behaviour
	Physical pain may result in a person with dementia becoming aggressive or withdrawn	Pain
11	A person with dementia is less likely to receive pain relief than a person without dementia when they are in hospital	Analgesia

To examine the KIDE-11 more comprehensively, the Rasch methods detailed for the KIDE-16 were applied again in the 11-item set. Rasch model calibration using the CML method in the KIDE-11 converged after 17 iterations of computation with a conditional log-likelihood value of -1026.60.

Item parameter estimates and standard errors for the Rasch model calibration of the KIDE-11 are displayed in *Table 5.4*. The difficulty parameter (*b*) estimates covered a wide range of the latent trait, from -1.59 (least difficult) to 1.42 (most difficult). The easiest item in the KIDE-11 was 'When people with dementia walk around it is

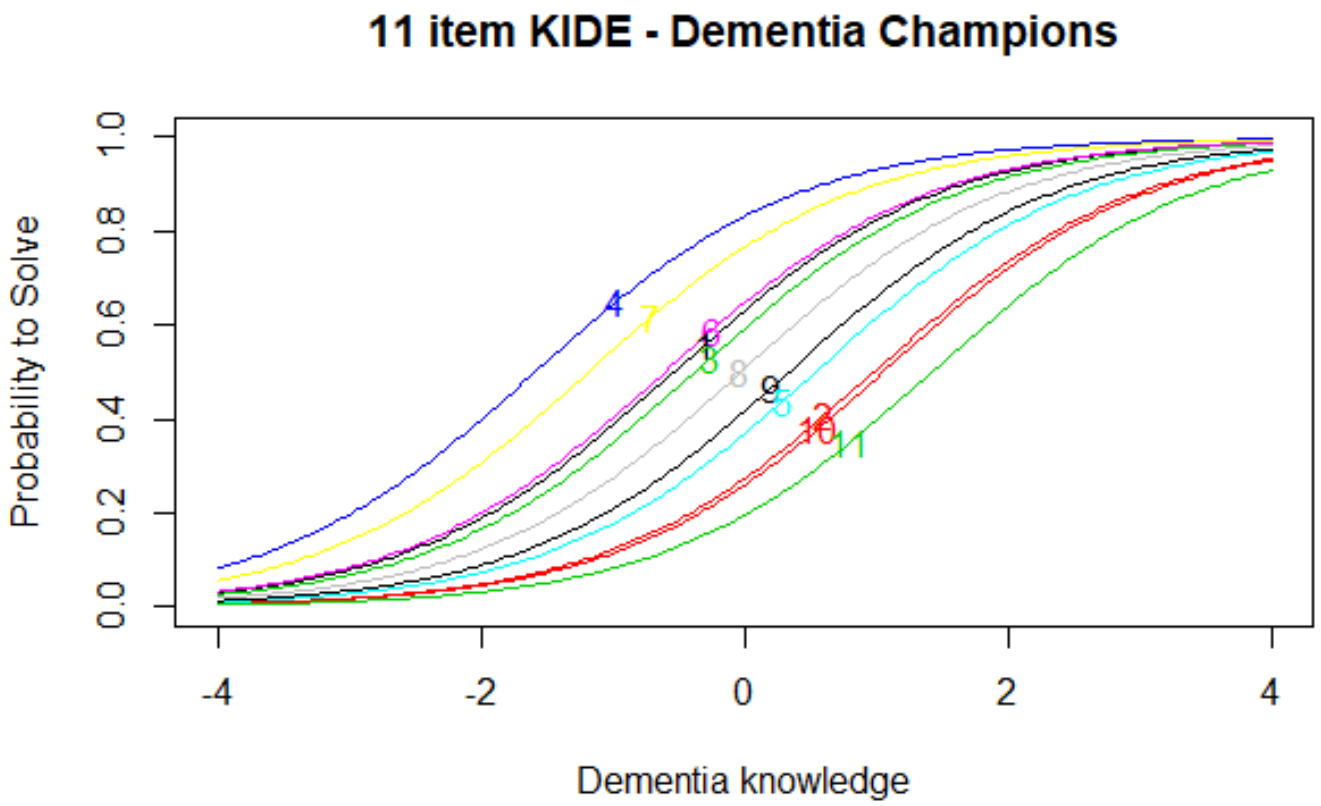
usually aimless' (item code: Aimless walking) with 95% correct. As with the KIDE-16, the most difficult item in the KIDE-11 was 'A person with dementia is less likely to receive pain relief than a person without dementia when they are in hospital' (item code: analgesia), with 63% correct. As also seen in *Table 4*, the standard errors for the difficulty parameters were acceptable for all items, with none of the outliers that were seen with the KIDE-16. Even the least difficult item had a standard error below 0.3.

Table 5.4: CML estimations of item difficulty with associated standard errors for the KIDE-11

Item	***b* Difficulty (se)**
Brain changes	-0.54 (0.18)
Strokes	0.98 (0.12)
Incurable	-0.38 (0.17)
Aimless walking	-1.59 (0.26)
Communicate	0.54 (0.13)
Brain damage	-0.61 (0.18)
Perceptions	-1.19 (0.22)
Anger	-0.03 (0.15)
Umbrella term	0.34 (0.14)
History/behaviour	1.05 (0.12)
Analgesia	1.42 (0.12)

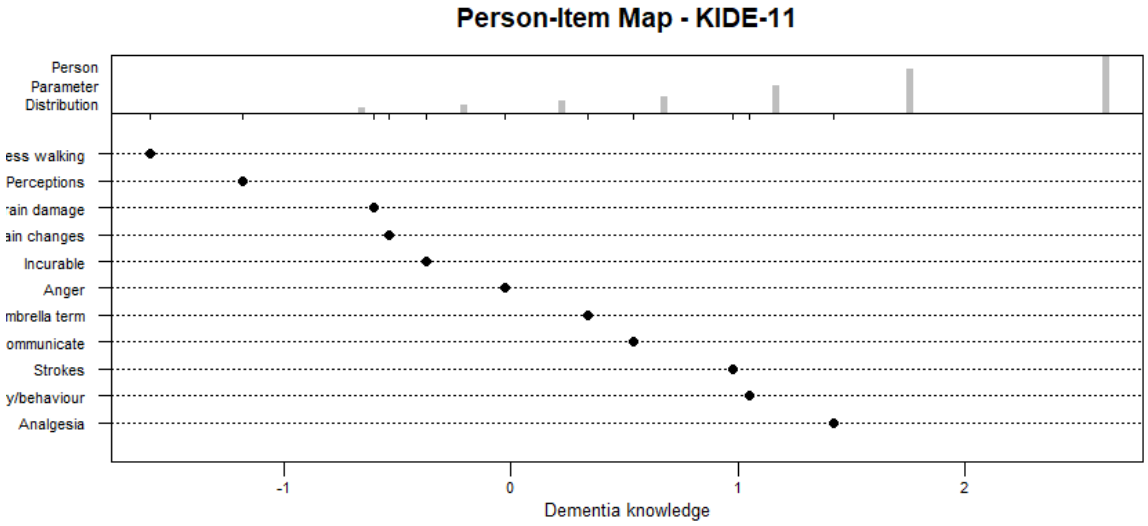
Joint ICCs for the KIDE-11 were plotted and examined. These are displayed below in *Figure 5.10*. Item distribution for the KIDE-11 showed a more even spread than for the KIDE-16, with no notable gaps in representation of the latent trait. As per the KIDE-16, some item redundancy is evident, for example with items 2 (Strokes) and 10 (History/behaviour), seen as almost-overlapping red curves toward the right-hand side of the plot. Rasch calibration of these items within the KIDE-11 showed these items to have very similar difficulty estimates, at 0.98 and 1.05, respectively.

Figure 5.10: ICCs for the KIDE-11



Measurement precision of the KIDE-11 was still limited for those respondents with higher levels of dementia knowledge, ie. those with θ estimates of >1.5, as can be seen in *Figure 5.11*.

Figure 5.11: Location of persons and items for the KIDE-11



The KIDE-11 demonstrated sufficient item fit and global model fit to the Rasch model, suggesting that the 11-item set fulfilled all of the underlying assumptions of the model.

5.4 Summary

The aim of this study was to calibrate a set of dementia knowledge items in a unidimensional latent trait by fitting the KIDE data to the Rasch model. Item and person parameters were estimated, and the fit of the Rasch model was evaluated. This was the first attempt at IRT modelling of the KIDE instrument. The underlying structure of the KIDE had not been examined in any published study; results from this study suggested that 11 of the 16 KIDE items showed sufficient fit to the Rasch model. Five items were discarded due to misfit. The underlying assumptions of the

Rasch model; being unidimensionality, parallel and non-intersecting item characteristics curves, and local independence were all fulfilled in relation to the 11-item KIDE, as evidenced by global model and item fit estimates.

The five discarded items were the least difficult of the KIDE-16. This supports the notion that misfit in relation to Rasch modelling is an indication of weakness in the dataset, rather than weakness of the model itself (Irwing, Booth and Hughes, 2018). Rasch calibration of the KIDE-11 revealed some redundancy of item content, with groups of two items having very similar difficulty parameter estimates. Redundant items add burden to the test taker and do not increase measurement precision, therefore such items should be carefully evaluated and potentially removed (Mair, 2018). In the case of the KIDE-11, redundancy was noted between item 1 (Permanent changes to the brain occur in most types of dementia) and item 6 (Brain damage is the only factor that is responsible for the way people with dementia behave); this was likely due to an overlap of item content, since both items pertain to dementia-related pathological brain changes. In this case, item removal or rewording of content would be recommended (McDonald, 2013). Item 2 (Dementia can be caused by a number of small strokes) and item 10 (A person with dementia's history and background plays a significant part in their behaviour) also had similar difficulty parameters, indicating possible item redundancy. However, there was no overlap of content for these two items since they asked clearly different questions. It may be the case that this redundancy could be resolved if the KIDE-11 were administered in populations with lower overall knowledge of dementia.

Targeting, or effective measurement range, in the KIDE-16 was extremely limited, with the majority of item parameter estimates showing low difficulty, and the majority of person parameter estimates showing above average levels of dementia knowledge. Therefore, the severity of the items did not correspond with the distribution of ability in the population of interest, which limited the usefulness of the test. Lower effective measurement range leads to poorer reliability and an inability to differentiate individuals or groups along the latent trait (Embretson and Reise, 2013), as was the case with the KIDE in this sample (Jack-Waugh, Macrae and Ritchie, 2017).

Future studies might administer the KIDE-16 in populations with lower hypothesised knowledge of dementia, such as healthcare students or public groups, prior to

conducting further psychometric evaluation. Dementia knowledge researchers/educators might administer the KIDE-11 concurrently with other measures of dementia knowledge, to test whether these items load on the same dimension as other knowledge items, and subsequently contribute to the effective measurement range of available instruments.

These findings will be discussed in the wider context of this PhD study in *Chapter 8*.

5.5 Conclusions

This chapter has detailed the application and calibration of the Rasch model to estimate item parameters, person parameters, and model fit of the Dementia Champions data. This study details the first examination of item responses to the 16-item KIDE using IRT model calibration. Item parameters under the Rasch model suggested that around half of the KIDE items were too easy to facilitate precise measurement in populations of healthcare professionals, hence ineffective targeting. The Rasch model had a sufficient fit to 11 of the 16 items, however the KIDE-11 effectively covered only the lower end of ability, as in, those participants with lower-than-average knowledge of dementia. As such, and given it is strong enough to withstand Rasch calibration as demonstrated throughout this chapter, the KIDE-11 may be a suitable dementia knowledge instrument for populations with generally low levels of dementia knowledge.

CHAPTER 6

Rasch modelling of dichotomous dementia knowledge data: Undergraduate nursing student responses to the Knowledge in Dementia (KIDE) scale

6.1 Introduction

The Rasch model is widely understood to be a simple yet powerful application of item response theory (IRT); this model can facilitate rigorous development and examination of measurement instruments (Andrich & Marias, 2019). *Chapter 5* detailed the application and calibration of the Rasch model to estimate item parameters, person parameters, and model fit of the Dementia Champions data. The findings from *Chapter 5* demonstrated that 11 of the 16 KIDE items showed sufficient fit to the Rasch model indicating the potential of the item-set to represent a robust unidimensional scale. Although the KIDE-11 was strong enough to withstand Rasch calibration, most items had difficulty parameter estimations below the average of the sample's ability, meaning that this version of the instrument was too easy for the Dementia Champions participants.

This ineffective measurement range severely limited the ability of the KIDE-11 to provide precise measurement of KoD in trained healthcare professionals and, as such, one recommendation to arise from *Chapter 5* was that the KIDE scale might be administered in populations with lower hypothesised knowledge of dementia, such as healthcare students or public groups, prior to conducting further psychometric evaluation.

Measurement of Knowledge of Dementia (KoD) literature is currently limited in the sense that supporting psychometric studies are predominantly absent and therefore evidence to support robust psychometric characteristics of scales is missing. This represents a missed opportunity in the methods used to develop currently available instruments, for example, in the lack of item response theory modelling in general and more specifically, the absence of Rasch modelling in scale development and evaluation. The literature review in *Chapter 3* mapped the currently available KoD instruments and critically appraised their development methods and reported

psychometric properties. The overwhelming majority of currently available instruments were developed using classical test theory (CTT) methods, and as such there is a marked absence of IRT modelling in this field of research and practice.

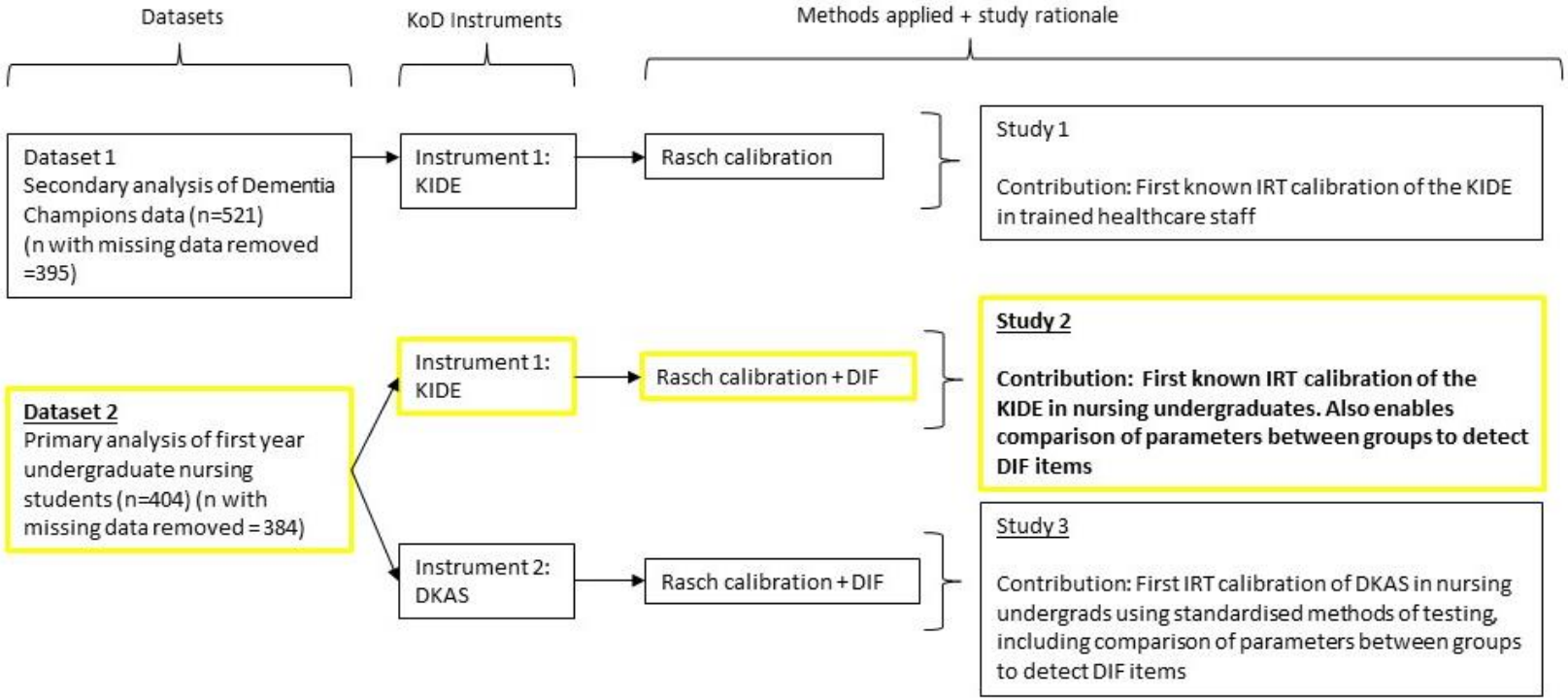
The current chapter addresses this recommendation and missed opportunity by performing Rasch modelling of dichotomous dementia knowledge data collected from undergraduate nursing students who responded to the Knowledge In Dementia (KIDE) scale (Elvish et al., 2014) in a classroom survey across two campuses of a university in Scotland.

6.1.1 Chapter aim

This chapter details the application and calibration of the Rasch model to estimate item parameters, person parameters, and model fit of dataset two; a set of responses to the KIDE from a cohort of first year undergraduate nursing students.

The results form the second empirical contribution to this thesis. The objective of this study was to determine the usefulness of the Rasch model to calibrate the KIDE. It asks whether KIDE items can be calibrated to quantified locations on a unidimensional interval-scaled measurement continuum based on undergraduate students' KIDE responses. If successful, such a Rasch calibration would facilitate the examination of key features such as description and evaluation of effective measurement range and measurement precision/reliability in this sample of undergraduate nursing students in comparison to the Dementia Champions sample from Dataset one. The effective measurement range of a test can be influenced by instrument developers' decisions on whether to aspire to; i) a range of items with narrow difficulty parameter estimates, and therefore the capability for precise scoring (high measurement precision), or ii) items spanning a larger range of theta estimates, but are in turn less able to detect minor, precise changes (lower measurement precision) (DeMars, 2010). The methods used in this chapter are highlighted in the schematic (*Figure 6.1*).

Figure 6.1: Methods schematic for Chapter 6 – Rasch calibration of undergraduate student responses to the KIDE scale.



6.2 Methods – Rasch calibration of dataset two

6.2.1 Rasch analysis

As per *Chapter 5*, the Rasch model was fit to the KIDE data using a conditional maximum likelihood (CML) estimation approach initially using the eRM package in R (Mair, Hatzinger, and Maier, 2020) 'raschtest' commands in Stata version 16.1. Item and person parameters were estimated independently. Statistical tests for model and item fit were conducted i.e., the magnitude of the p-values for Andersen's LR tests (model fit) and Wald tests (item fit) were used to examine the fit of the Rasch model to the data, in two ways. Graphical representations that complement these analyses were generated using the eRM package. Examination for differential item functioning was done using the difR package in R (Magis et al., 2010).

6.2.2 Reminder of rationale for methods

The fundamental assumptions underlying the Rasch model include unidimensionality, parallel and non-intersecting item characteristics curves (monotonicity), and local independence (see *Chapter 2, section 2.3* for more information on these assumptions). If data fit the Rasch model sufficiently then all three assumptions can be considered fulfilled (Mair, 2018). The methods employed in this chapter mirror those detailed in *Chapter 5* to enable examination of how the KIDE item set performed in this sample of nursing undergraduates compared to the Dementia Champions sample. However, given the availability of demographic variables in this undergraduate nurse dataset, the methods have been expanded to include examination for differential item functioning in relation to these variables (characteristics of the student sample).

6.3 Results

As described in *Chapter 4 (section 4.2.2)*, first-year nursing undergraduates were recruited as a whole cohort from a nursing programme in one university in Scotland. Upon ethical approval, permission to access first year students was sought via the

head of undergraduate studies as the gatekeeper. The sample consisted of 404 students who were enrolled in their first year of Adult Nursing, Mental Health Nursing, or Child Nursing pathway, in the year 2019/2020, across each of the two University campuses. All analyses in this chapter are based on the complete sample with missing data removed (n=384).

6.3.1 Rasch model calibration – conditional maximum likelihood estimation

Rasch model calibration was performed using the CML estimation method. Estimation using the CML method converged after 28 iterations of computation with a conditional log-likelihood value of -1692.44. These results were replicated in the Stata-based analyses.

6.3.2 Calibration of item parameters

To examine the distribution of items across the latent continuum of dementia knowledge, the first step in Rasch model calibration was the examination of results to determine KIDE item difficulty parameter estimates and their position on the latent scale.

The difficulty parameter (*b*) estimates covered a wide-ranging latent scale of dementia knowledge, with the least difficult item being ‘People with dementia never get depressed’ (item code: Depression) at 99% correct and a θ estimate of -3.32. The most difficult item was ‘A person with dementia is less likely to receive pain relief than a person without dementia when they are in hospital’ (item code: Analgesia), with a θ estimate of 2.95. *Table 6.1* reports item parameter estimates and standard errors for KIDE items under a Rasch model; larger than average standard errors are highlighted in bold.

Table 6.1. CML estimations of item difficulty with associated standard errors: undergraduate nurse sample.

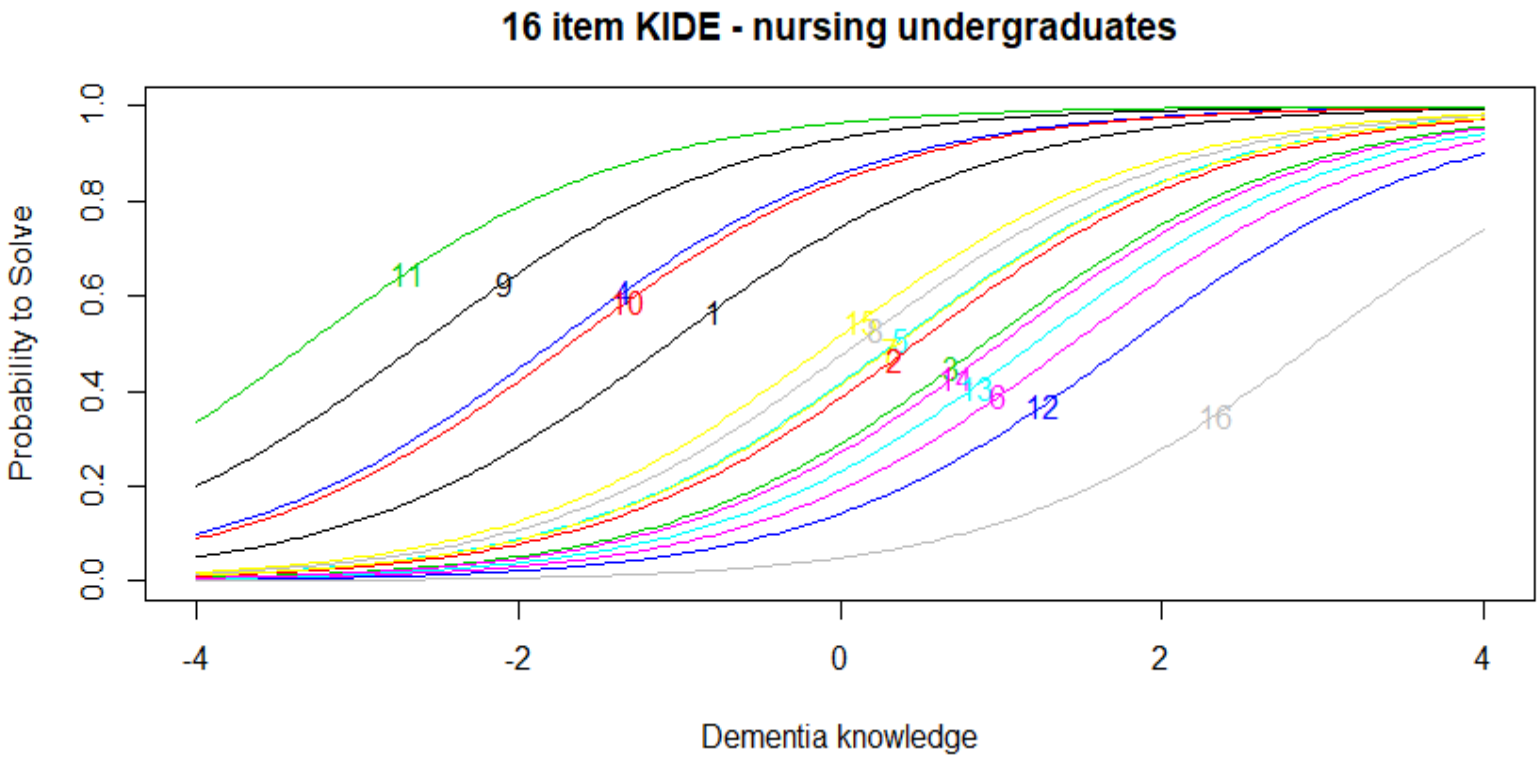
	KIDE items 1-16	***b* Difficulty (se)**
1	Permanent changes to the brain occur in most types of dementia	-1.07 (0.24)
2	People who have dementia will usually show the same symptoms	0.47 (0.14)
3	Dementia can be caused by a number of small strokes	0.90 (0.13)
4	Currently, most types of dementia cannot be cured	-1.79 **(0.32)**
5	When people with dementia walk around it is usually aimless	0.33 (0.15)
6	People with dementia will eventually lose all their ability to communicate	1.44 (0.12)
7	People with dementia who are verbally aggressive nearly always become physically aggressive	0.35 (0.15)
8	Brain damage is the only factor that is responsible for the way people with dementia behave	0.10 (0.16)
9	It is possible to catch dementia from other people	-2.62 **(0.48)**
10	My perception of reality may be different from that of a person with dementia	-1.68 (0.31)
11	People with dementia never get depressed	-3.32 **(0.67)**
12	Anger and hostility occur in dementia mostly because the ‘aggression’ part of the brain has been affected	1.79 (0.12)
13	Dementia is a general term which refers to a number of different diseases	1.20 (0.13)
14	A person with dementia's history and background plays a significant part in their behaviour	0.99 (0.13)
15	Physical pain may result in a person with dementia becoming aggressive or withdrawn	-0.06 (0.16)
16	A person with dementia is less likely to receive pain relief than a person without dementia when they are in hospital	2.95 (0.13)

The standard errors of the parameter estimates were inspected for any large values. SEs were acceptable for most items, however, the three following items had larger standard errors:

i. Item 4: ‘Currently, most types of dementia cannot be cured’ (item code: incurable)
ii. Item 9: ‘It is possible to catch dementia from other people’ (item code: contagious)
iii. Item 11: ‘People with dementia never get depressed’ (item code: depression).

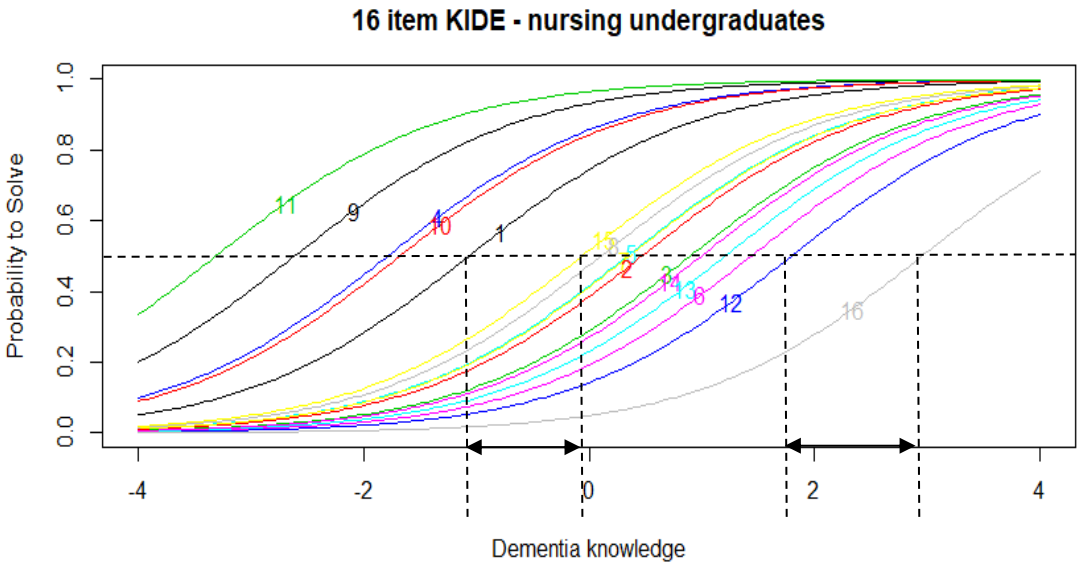
These were the three least difficult KIDE items and were consequently answered correctly by almost all participants, at 98%, 99% and 99%, respectively. The size of the standard errors for these items reflects the low precision of estimation across these (not very informative) items, due to the extreme participant response patterns.

Figure 6.2: Joint ICCs for the 16 item KIDE. Items are labelled by number as per the KIDE running order



Difficulty parameters for all items are displayed graphically in *Figure 6.2 (*above). All item characteristics curves (ICCs) were monotonic and non-intersecting, thus confirming this assumption of the Rasch model. Examination of the joint ICCs shows pairs of overlapping curves, for example, items 4 and 10, shown in blue and red towards the left-hand side of the plot. These items (Incurable and Perceptions) had similar difficulty parameters, at -1.79 and -1.68 respectively. Note also overlapping items 5 and 7, here shown in aqua and yellow near the centre of the plot. These items (Aimless walking and Aggression) had very similar difficulty parameter estimations at 0.33 and 0.35 respectively. Overlapping items, such as those highlighted here, are generally less useful in terms of examining the relationship between participants and items, as two items with the same (or very similar) difficulty parameter estimates will add to participant burden without necessarily increasing measurement precision of the scale (Irwing, Booth and Hughes, 2018).

Figure 6.3: Sections of the latent scale unrepresented by KIDE item content



Further examination of the joint ICCs (*Figure 6.3*) showed two significant sections of the latent scale/ trait that were not covered by KIDE item content, with no items

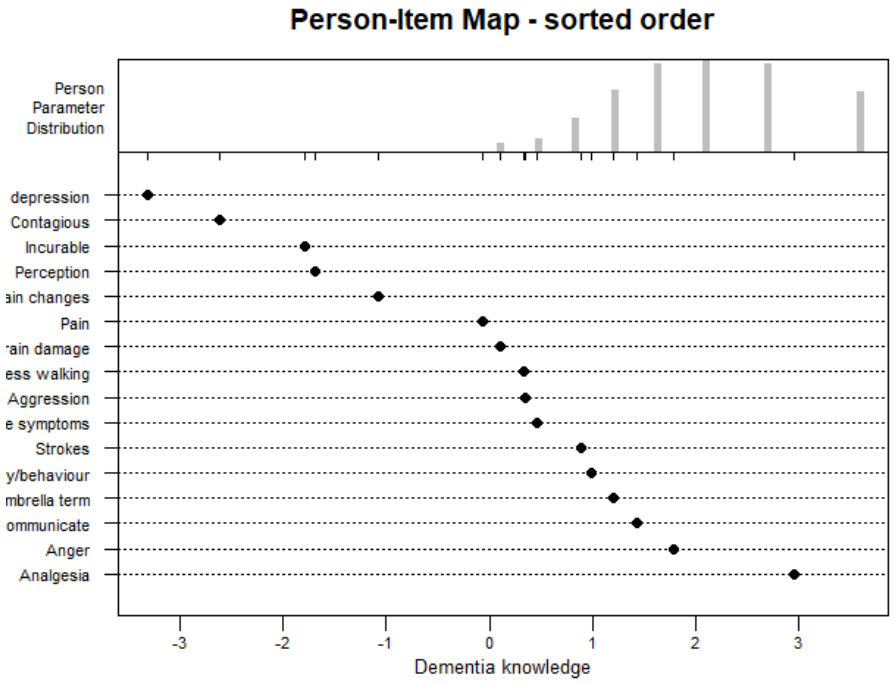
representing theta estimates of -1.07 to -0.06, and 1.79 to 2.95. In this respect, the measurement precision of the KIDE was limited in nursing undergraduates whose ability level fell between these estimates.

The next step in the sequence of Rasch model calibration was to estimate the distribution of person parameters, to examine the latent distribution of person ability along the latent scale continuum, and the relationship between person ability and item parameters for the KIDE.

6.3.3 Person parameters

The person-item map *(Figure 6.4)* shows the distribution of persons and items along the KoD latent trait continuum. The distribution of items is shown in the lower panel, whereas the person distribution is shown in the upper panel. Examination of the location of items relative to persons showed that five of the KIDE items (those to the left of the person parameter distribution bars) were too easy to suit the ability range of the undergraduate nurse respondents, suggesting poor targeting of measurement.

Figure 6.4: Person and item parameters for the 16-item KIDE, sorted from easiest to most difficult item



The unrepresented areas of the latent trait dimension reported above can also be seen clearly in *Figure 6.4.* They appear as (sometimes large) gaps between the spacing of items. Of particular note is the section of the latent scale between θ estimates of 1.79 to 2.95, highlighted below in *Figure 6.5*: the lower panel shows that there are no items in the KIDE that can measure these specific levels of ability, however the upper panel shows that a significant proportion of respondents had ability levels between these parameter estimates. This result demonstrates a lack of measurement precision for this proportion of the sample.

Figure 6.5: Annotated person-item map showing lack of measurement precision of the KIDE

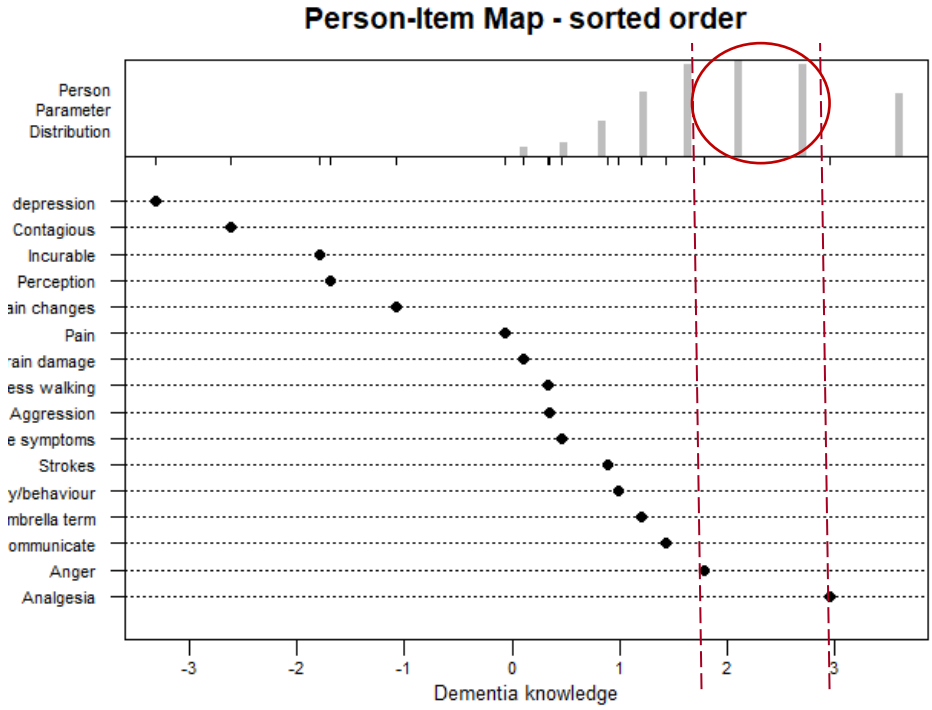


Table 6.2 below shows the raw scores and their associated proficiency estimates based on Rasch model calibration. A useful feature of the Rasch model is demonstrated here: the scores of 0 – 7, despite having zero frequency, have associated ability parameter estimates and standard errors. This table also demonstrates the non-linear transformation from total score to ability estimate, as in, although the distance between all raw scores is 1, the distance between corresponding ability estimates varies due to the differing difficulty parameters across items. As expected, the standard errors at the extreme ends of the scale were much larger than those in the middle of the score range.

Once the item and person parameter estimates had been examined, the next step was to assess whether the KIDE data had a good fit to the Rasch model; this included examination for local dependence between pairs of items, and scrutiny of model and item fit statistics.

Table 6.2: Total scores, frequencies, proficiency estimates and standard errors

Score	**Frequency**	**Ability parameters (CML)**	**Std. error**	**Expected score**
0	0	-7.86	**2.55**	0.40
1	0	-6.47	0.63	1.27
2	0	-5.66	0.35	2.19
3	0	-5.03	0.25	3.15
4	0	-4.49	0.19	4.13
5	0	-4.01	0.16	5.12
6	0	-3.59	0.14	6.10
7	0	-3.20	0.12	7.07
8	3	-2.83	0.11	8.04
9	7	-2.48	0.11	9.00
10	26	2.13	0.11	9.95
11	52	-1.77	0.13	10.91
12	76	-1.39	0.15	11.85
13	80	-0.95	0.20	12.80
14	77	-0.42	0.30	13.75
15	50	0.29	0.56	14.68
16	13	1.63	**2.35**	15.58

6.3.4 Fit statistics

As per *Chapter 5*, the non-parametric testing framework of Ponocny (2001) was used to test for the assumption of local independence. Only six of 105 item pairs tested showed local dependence, and the assumption of local independence held at a global level (p =0.06), indicating that the underlying assumption of local independence for Rasch model fit was upheld.

To assess whether the Rasch model assumption of measurement invariance was upheld, tests for model and item fit included Andersen's (1973) LR-test and Wald tests, as detailed by Andrich and Marais (2019). For the model to fit, item parameters had to be invariant across person subgroups, meaning the fit of the model in subgroups must be approximately the same (Irwing, Booth and Hughes, 2018).

6.3.5 Model fit

Model fit was examined initially by assessing appropriate tests fit using mean and median split criteria. This also enabled comparison to the UWS KIDE results. Fit tests should be run on multiple covariates where possible (Embretson and Reise, 2019). Given the availability of demographic variables in the undergraduate nursing dataset, fit was therefore examined after splitting the sample according to the following variables present in the undergraduate nurse dataset and describing different characteristics:

'*Do you know or have you known someone who lives with dementia*' (variable code: Known);

' *Do you work or have you ever worked with people who live with dementia*' (variable code: Worked).

These variables were binary 'yes/no' so the format for both was suitable for use as subgroup splits in model and item fit tests. See *Table 6.2* below for a reminder of the descriptives for these sample characteristics.

Table 6.3: Undergraduate nurse responses to two external binary covariates for examination of Rasch model fit

	Yes	**No**
Do you know or have you known someone who lives with dementia?	246 (64%)	137 (36%)
Do you work or have you ever worked with people who live with dementia?	199 (52%)	185 (48%)

Application of the LR-test (Andersen, 1973) splitting the sample according to the *mean* (of the sum score) gave a non-significant result ($p = 0.099$) indicating that likelihoods did not differ across these two groups, therefore the assumptions of the Rasch model were upheld. Two items, (Contagious and No depression) were excluded by the test due to inappropriate response patterns. These were the two easiest items in this sample. The LR-test with a *median split* on the sum score returned a significant result ($p= 0.008$) indicating that the measurement invariance assumption of the Rasch model was not upheld using this split criterion. Further, to achieve this result it was necessary to reduce the scale by four items (Incurable, Contagious, Perception, No depression) due to inappropriate response patterns.

Model fit statistics were more favourable using the splits on covariates 'Known' and 'Worked', with Andersen LR-test results of $p =0.535$ and $p = 0.591$, respectively, both non-significant results. Since these variables were independent of the KIDE responses, the issue of items being removed due to inappropriate response patterns was eliminated.

Using four different subgroup splits, the likelihoods differed across the groups only when using the median split criterion; this suggested good fit of the undergraduate nurse data to the Rasch model.

6.3.6 Item fit

To support the model fit results, tests for item fit were performed using the covariates 'Known' and 'Worked' as the split criterion. Given the removal of several items during the LR-tests with mean and median subgroup splits, Wald tests are not reported here for these subgroups, since the covariate subgroup splits provided item fit indices for all 16 items. Results of the Wald tests for individual item fit can be seen in *Table 6.3*. None of the items showed significant misfit, therefore there was no indication to eliminate items from the scale for use in this population based on fit statistics alone.

Table 6.4. Item fit indices for Wald tests with 'Known' and 'Worked' subgroup splits.

Item code	**'Known' covariate z-statistic**	**p-value**	**'Worked' covariate z-statistic**	**p-value**
Brain changes	-0.36	0.72	1.06	0.29
Same symptoms	0.16	0.88	0.51	0.61
Strokes	-0.52	0.61	0.14	0.89
Incurable	1.46	0.14	0.04	0.96
Aimless walking	-0.41	0.68	-0.59	0.56
Communicate	-1.16	0.25	-0.84	0.40
Aggression	-0.52	0.61	-1.50	0.13
Brain damage	1.75	0.08	0.49	0.62
Contagious	0.30	0.77	-0.21	0.83
Perception	-0.87	0.39	-0.98	0.33
No depression	0.21	0.83	-0.15	0.88
Anger	0.25	0.80	-0.07	0.95
Umbrella term	-0.09	0.93	2.12	**0.03**
History/behaviour	-2.20	**0.03**	0.41	0.68
Pain	-0.34	0.73	0.15	0.88
Analgesia	-0.10	0.92	0.99	0.32

Tests for model and item fit strongly suggested that all 16 KIDE items showed sufficient fit to the Rasch model, with all three assumptions of unidimensionality, monotonicity, and local independence being upheld. An additional step that was possible in this study, given the additional covariates in the dataset, was to determine whether any items may be measuring different abilities for members of subgroups. Examination for differential item functioning (DIF) can highlight any unexpected behaviour of test items and determine the presence of bias (Andrich and Marais, 2018).

6.3.7 Differential item functioning

DIF analysis was performed using Lord's (1980) chi-squared method; a common IRT-based method used to detect DIF items and the presence of bias.

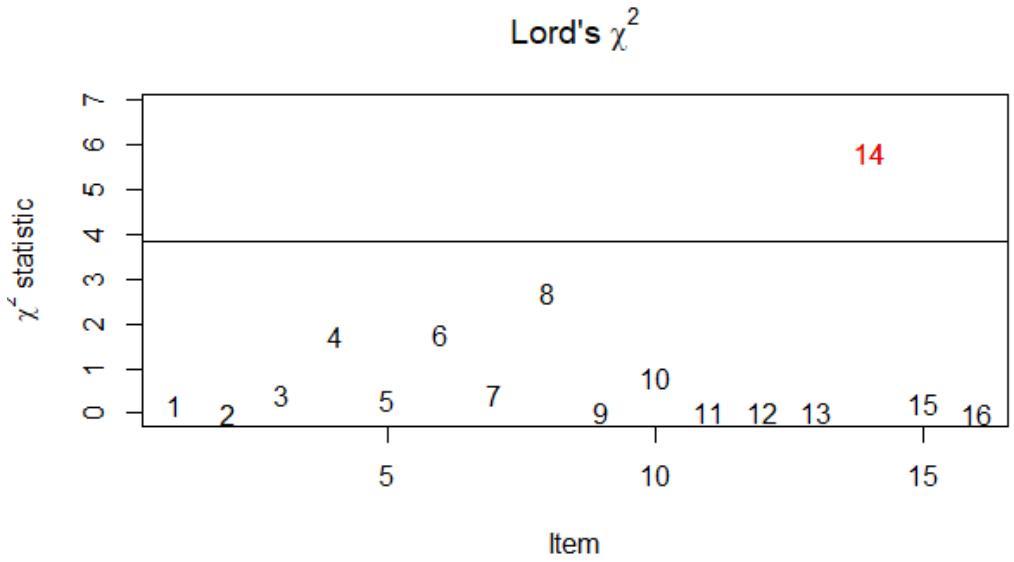
Two subgroups of undergraduate nurses were included in the DIF analysis of the 16-item KIDE: knowing someone with dementia (Yes = 64% vs No = 36%) and experience of working with people with dementia (Yes (52%) vs No (48%)). Gender was not used due to the small sample size for males, who accounted for only 8% of the total sample and as such, it was anticipated this group size was insufficient to have enough statistical power to identify DIF items (DeMars, 2010). *Table 6.5* shows the results of the DIF analysis. Across the two sample splits, only two items (highlighted in bold) showed evidence of potential bias.

Table 6.5: Lord's (1980) chi-square values for sample split using the 'WORKED' and 'KNOWN' covariates.

Item	**Variable 'Worked'**		**Variable 'Known'**	
	Lord's chi-square	p-value	Lord's chi-square	p-value
1	1.13	0.29	0.20	0.66
2	0.42	0.52	0.00	0.97
3	0.09	0.76	0.43	0.51
4	0.01	0.95	1.74	0.19
5	0.24	0.62	0.29	0.59
6	0.49	0.49	1.76	0.18
7	2.05	0.15	0.43	0.51
8	0.35	0.55	2.72	0.10
9	0.03	0.86	0.05	0.82
10	0.83	0.36	0.82	0.37
11	0.02	0.90	0.02	0.88
12	0.03	0.86	0.05	0.83
13	**5.98**	**0.01**	0.04	0.85
14	0.37	0.55	**5.85**	**0.02**
15	0.05	0.82	0.22	0.64
16	2.04	0.15	0.01	0.91

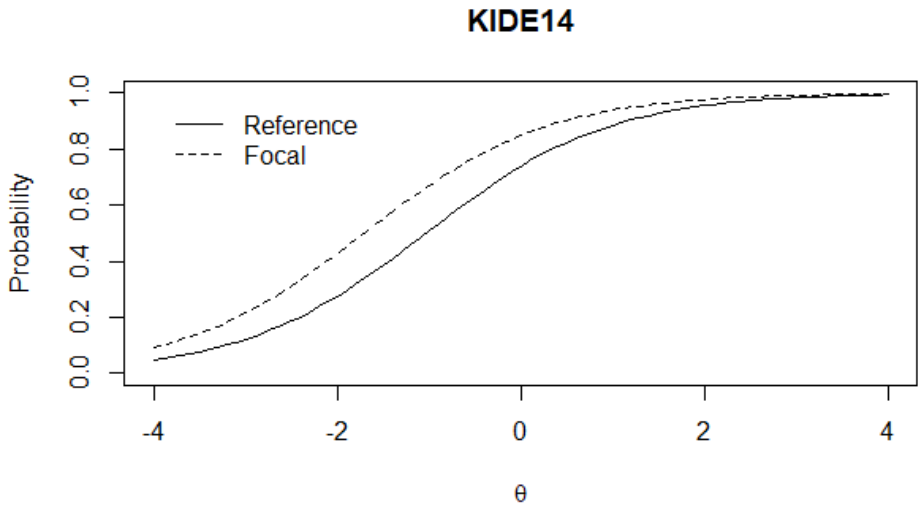
The DIF summary using the 'Known' sub split is displayed graphically in *Figure 6.6* where the y-axis shows the chi-squared statistic, and the horizontal line represents the DIF detection threshold. Using this sample split, item 14 "*A person with dementia's history and background plays a significant part in their behaviour*" (Item code: History/behaviour) was a DIF item.

Figure 6.6: DIF detection using the ‘Known’ split showing item 14 as a DIF item



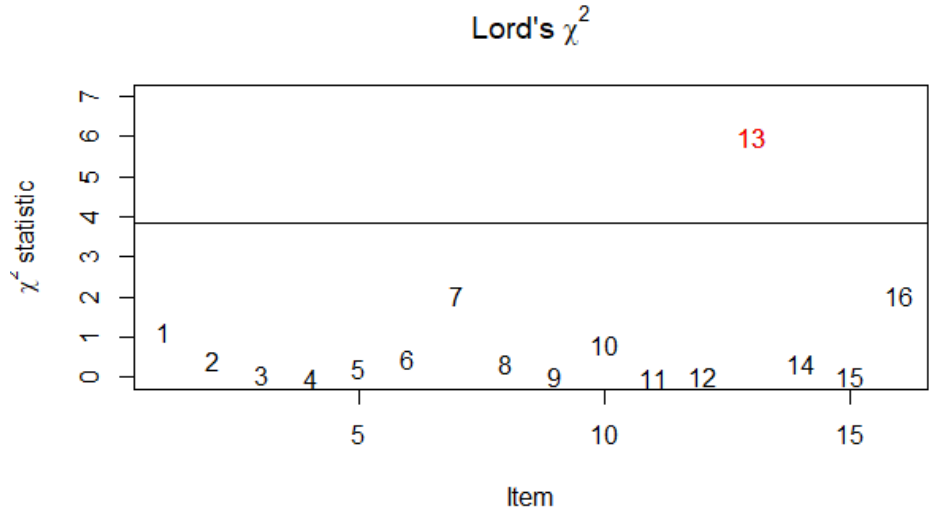
The DIF detected in item 14: ‘A person with dementia's history and background plays a significant part in their behaviour’ (History/background) is displayed visually in *Figure 6.7*, below. Examination of this plot showed that nurse undergraduates who had known someone living with dementia (shown as the Reference curve) scored higher on this item than those who had not known someone living with dementia (Focal curve), given the same latent trait location. Therefore, respondents who had known someone living with dementia had an advantage on this particular item of the scale.

Figure 6.7: Item characteristics curve for DIF item 14, divided by the variable 'Known'.



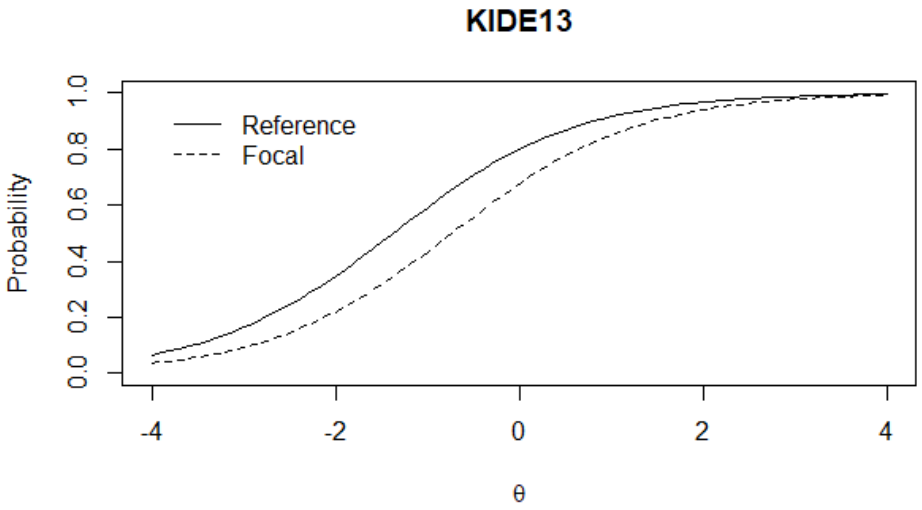
The DIF summary using the 'Worked' sample split is displayed in *Figure 6.8*. Using this sample split, item 13 "*Dementia is a general term which refers to a number of different diseases*" (Item code: Umbrella term) was a DIF item since it falls above the detection threshold line.

Figure 6.8: DIF detection using the ‘Worked’ split showing item 13 to be a DIF item



The DIF detected in item 13 (Umbrella term) is displayed in *Figure 9*, below. As with the previous DIF item, item 13 displayed uniform DIF. Comparison of the curves in this plot show that respondents who had worked someone living with dementia (Reference curve) scored *lower* on item 13 than those who had not worked with people living with dementia (Focal curve). Therefore, respondents who had worked with people living with dementia had a slight disadvantage on this item of the scale, regardless of their position on the latent trait.

Figure 6.9. Item characteristics curve for DIF item 13 using 'Worked' covariate split



The two DIF items: 13 (Umbrella term) and 14 (History/background) also demonstrated the poorest fit based on the Wald tests using the same subgroup splits. Refer back to *Table 6.4* for these item fit indices. Although these items did not demonstrate as strong a fit as the majority of the KIDE items, there was no evidence of misfit and both items' difficulty parameter estimates were above the sample average, therefore there was a strong case to retain both items.

6.4 Summary

The aim of this study was to examine the fit of the Rasch model to dataset two: a set of undergraduate nurse responses to the KIDE scale. Item and person parameters were calibrated on a unidimensional latent trait continuum; item and model fit statistics were examined, and detection for item bias was performed using IRT-based methods. Results from this study strongly suggested that all 16 KIDE items showed sufficient fit to the Rasch model, thus implying that the underlying assumptions of

the model (being unidimensionality, monotonicity, and local independence) were fulfilled.

In this undergraduate sample, the Rasch-calibrated KIDE covered a slightly wider range of the latent trait continuum than the KIDE-11 in the Dementia Champions sample (Chapter 5), however both studies showed that there were number of items with difficulty parameter estimates too low for the study participants (five items from the KIDE-16 and six items from the KIDE-11). Another finding common to both studies was that the most and least difficult items of the KIDE-16 and the KIDE-11 were the same items.

Item difficulty indices were examined in relation to the distribution of Θ in the undergraduate nurse sample and the effect this has on the measurement properties of the test. For dementia knowledge measurement it is desirable to have items that span a reasonably large range of ability (Irwing, booth and Hughes, 2018). In this context, the aim of measurement was to determine which KIDE items provide sufficient information and reliability indices to be capable of precise measurement. In other words, what is the effective measurement range of an instrument and/or sets of items within. In the undergraduate student dataset, although the KIDE covered a wide range theta estimates (-3.32 to 2.95), very few respondents had below-sample-average ability levels, therefore the effective measurement range of the KIDE was much narrower than the calibrated measurement range, realistically containing only the 11 items that had Rasch difficulty parameter estimates of $\Theta > 0$. The section of the latent trait that was unrepresented by item content ($\Theta = 1.79$ to 2.95) further reduced measurement precision in this group of participants. This finding was of particular importance given the large proportion of respondents whose person parameter estimates fell within this range.

The results of this chapter demonstrated that the KIDE did not provide information across the whole ability distribution. Future research would ideally aim to generate and test additional items to address gaps in the latent trait, therefore contributing to measurement precision and improving the effective measurement range of the KIDE.

These findings will be discussed further, and in the wider context of this PhD study throughout *Chapter 8*.

6.5 Conclusions

This chapter has detailed the application and calibration of the Rasch model to a dataset of undergraduate nurse responses to the 16-item KIDE scale. All tests of model and item fit strongly suggested that the data fit the Rasch model sufficiently, therefore giving the KIDE a high seal of approval in this population. However, improvements could be made regarding effective targeting of measurement and overall measurement precision in samples of undergraduate nurses. Generation of new item content may address these issues.

CHAPTER 7

Rasch modelling of dichotomous dementia knowledge data: Undergraduate nursing student responses to the Dementia Knowledge Assessment Scale (DKAS)

7.1 Introduction

The benefits of the Rasch model in the examination of measurement instruments have been demonstrated in *Chapter 5* and *Chapter 6*, which detailed the application and calibration of the Rasch model to responses to the KIDE scale from samples of Dementia Champions participants and undergraduate nurse respondents. A significant finding from *Chapter 5* was that, although the Rasch model showed a good fit to an 11-item set, the KIDE-11 was overall too easy for the Dementia Champions participants, as trained healthcare professionals. This evidence of ineffective measurement range meant that the KIDE-11 was unsuitable to provide precise measurement of KoD in trained healthcare professionals and was therefore not well-targeted in the Dementia Champions educational interventions.

In an attempt to further examine these limitations of the KIDE regarding targeting, *Chapter 6* detailed the fit of the Rasch model to a set of KIDE responses from first year undergraduate nursing students. All 16 items showed sufficient fit to the Rasch model, and the KIDE-16 covered a wider range of the latent trait continuum than the KIDE-11. However, on examination of the relationships between item and person parameters, only 11 of the 16 items contributed to measurement precision in the nursing undergraduates. The results in *Chapter 6* also demonstrated how the Rasch model was able to highlight sections of the trait that were unrepresented, as in, where further item content would be necessary to improve measurement precision of the scale. These results generated the conclusion that the KIDE scale is psychometrically promising, being strong enough to withstand Rasch calibration, but that improvements could be made by generating additional items for future psychometric testing of the scale, given the current items did not provide information across the entire ability distribution.

7.1.1 The Dementia Knowledge Assessment Scale

The primary research question of this PhD study was to determine whether IRT modelling techniques can be used to improve understanding of the performance of dementia knowledge tests and make dementia knowledge testing more informative. This chapter will detail the IRT-based analysis of a second dementia knowledge instrument: the Dementia Knowledge Assessment Scale (DKAS) (Annear et al., 2015).

As described in *Chapter 4, section 4.3*, the DKAS has been reported as a more dimensionally complex instrument than the KIDE, having been subject to exploratory and confirmatory factor analytic (CFA) techniques. Results seem to show the DKAS to be a 4-factor scale (Annear et al., 2017). The DKAS CFA study was conducted in a large sample (n= 3649) of predominantly nurses, professional care workers, and the general public from 97 countries as part of an 'Understanding Dementia' Massive Open Online Course (MOOC); the sample self-selected from the MOOC participants (n= 11,241) and the DKAS was administered in a virtual setting.

As part of the initial analysis of the DKAS for this PhD study, an exploratory factor analysis (EFA) was conducted to examine whether the four-factor model might also be recovered by analysis in the sample of undergraduate nurse. This might be expected given there were key differences between the MOOC respondents and undergraduate nurse respondents.

The three key differences were as follows (*Table 7.1*):

Table 7.1. Key differences between Dataset two and MOOC sample composition and testing conditions

	MOOC	**PhD thesis study**
Sample composition 1	International – 97 countries	Local – from one university course
Sample composition 2	Diverse sample groups, including nurses, professional care workers, general public, family carers, 'other' healthcare workers, health students comprised only 4.7% of the sample.	Homogenous sample – year one student nurses
Method of test administration	Virtual, online testing	In person, pencil and paper survey pack

In Dataset two, EFA of the DKAS in undergraduate student nurses did not suggest a strong 4-factor model, with the scree plot of the eigenvalues suggesting the possibility of either one or three factors present. These (more extensively reported) extended results of the DKAS EFA in nursing undergraduates are presented in *Appendix 7*.

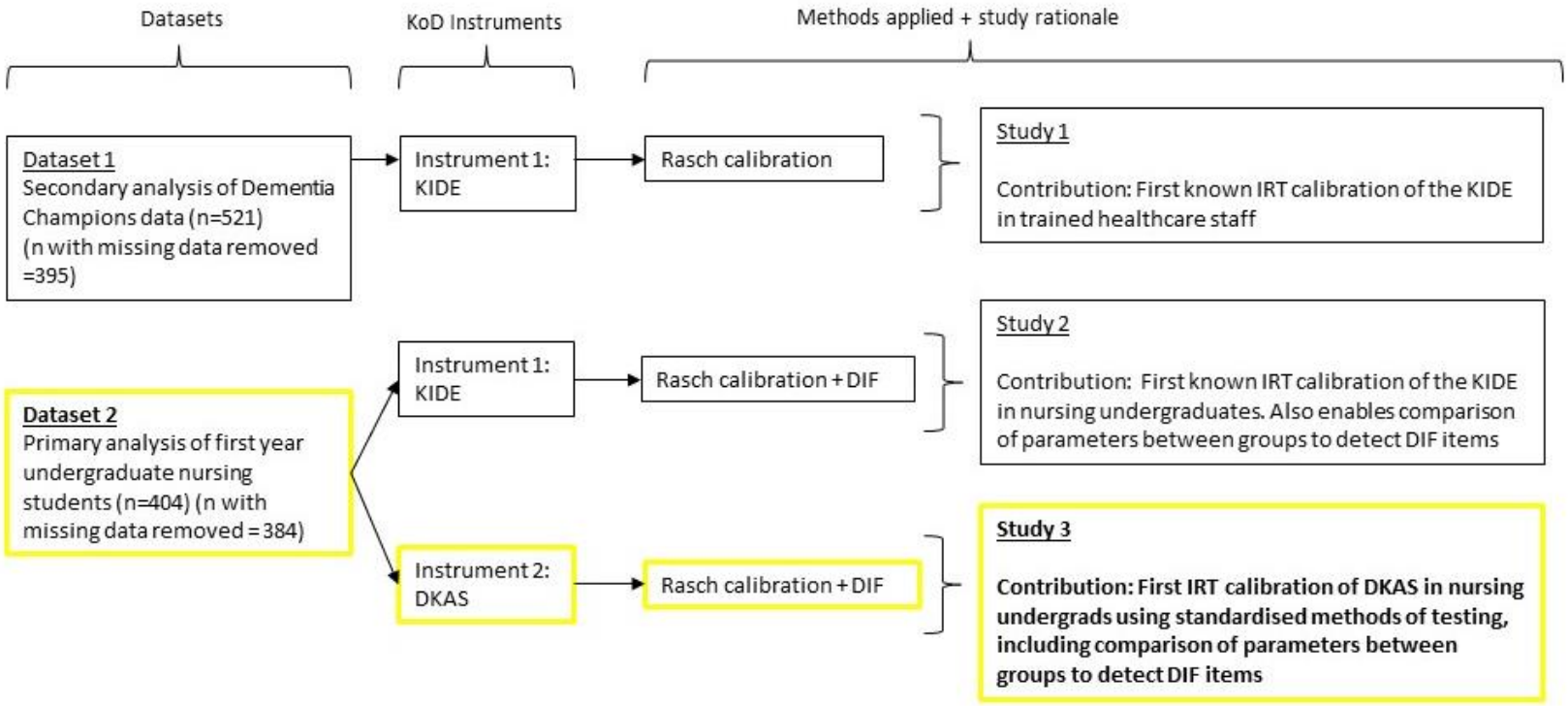
For IRT model selection, the most parsimonious model is generally the model that demonstrates the best fit with the least number of item parameters; if the Rasch model showed adequate fit to the data, then Rasch should be used over, for example, a 2-parameter-logistic (2PL) model (Meijer and Tendeiro, 2018). Given that none of the three a priori factors emerged as particularly strong (were not well defined), Rasch model calibration was conducted to examine; i) whether Rasch calibration could provide additional information about the performance of the DKAS items in this sample, and ii) whether the dataset of DKAS responses demonstrated adequate fit to the Rasch model. A strong subset of Rasch-calibrated items was envisaged as possible alternative scaling model.

7.1.2 Chapter aim

In this chapter, the Rasch model was used to examine the psychometric properties of the DKAS, based on responses from a cohort of first year undergraduate nursing students. In contrast to classical test theory methods, the Rasch model provides a robust method for nonlinear transformation of raw scores relating to items of varying difficulty levels (Andrich, 1988); scores can then be summated without the need to assign weights to the items prior to model calibration. As such, categorising respondents by their total score is justified by the agreement between the Rasch model and the data (Andrich and Marais, 2019).

The results constitute the third and final empirical contribution to this thesis. The objective was to determine to what extent the DKAS items could be located along a unidimensional scale by applying the Rasch model and examining item and person parameters, model and item fit indices, and potentially biased items. As with the results from *Chapters 5* and *6*, this would facilitate in-depth examination of effective measurement range and measurement precision of the DKAS: aspects of the scale that cannot be determined using classical test theory (CTT) methods alone (Embretson and Reise, 2013). The methods employed in this chapter are highlighted in *Figure 7.1*.

Figure 7.1. Methods schematic for Chapter 7 - Rasch calibration of the 25-item DKAS



7.2 Methods – Rasch calibration of dataset two- DKAS responses

The Rasch model was fit to the DKAS data using a conditional maximum likelihood (CML) estimation approach in the eRM package in R (Mair, Hatzinger, and Maier, 2020) and Stata version 16.1. Consistent with *Chapters 5 and 6*, Andersen's LR tests and Wald tests were used to examine for model and item misfit. All graphics were generated using the eRM package. Examination for differential item functioning was done using the difR package in R (Magis et al., 2010).

7.3 Results

The DKAS responses comprised the second part of Dataset two; a sample of first year undergraduate nurses (n=404) who had been on an Adult, Mental Health, or Child nursing degree pathway for four weeks. All analyses in this chapter are based on the complete sample with missing data removed (n=384), for useability across R packages. The output therefore relates to 384 sets of responses across 25 items. See *Chapter 4, section 4.2* for additional information on Dataset two.

7.3.1 Rasch calibration (eRM)

Rasch model calibration was performed using the CML estimation method. Estimation converged after 35 iterations of computation with a conditional log-likelihood value of -3868.52. These results were replicated in the Stata-based analyses.

7.3.2 Calibration of item parameters

Rasch-calibrated item difficulty parameters were examined. The difficulty parameter (*b*) estimates suggested a respectable measurement range; the least difficult item was 'People can recover from the most common forms of dementia' (item code: Recover) with 96% correct and a Θ estimate of -2.21. The most difficult item was 'The sudden onset of cognitive problems is characteristic of common forms of dementia (item

code: Sudden onset), with a Θ estimate of 2.40 and 24% correct. All item parameter estimates, and their associated standard errors are reported in *Table 7.2*. All standard errors were within an acceptable range, demonstrating high precision of estimation across the 25-item set.

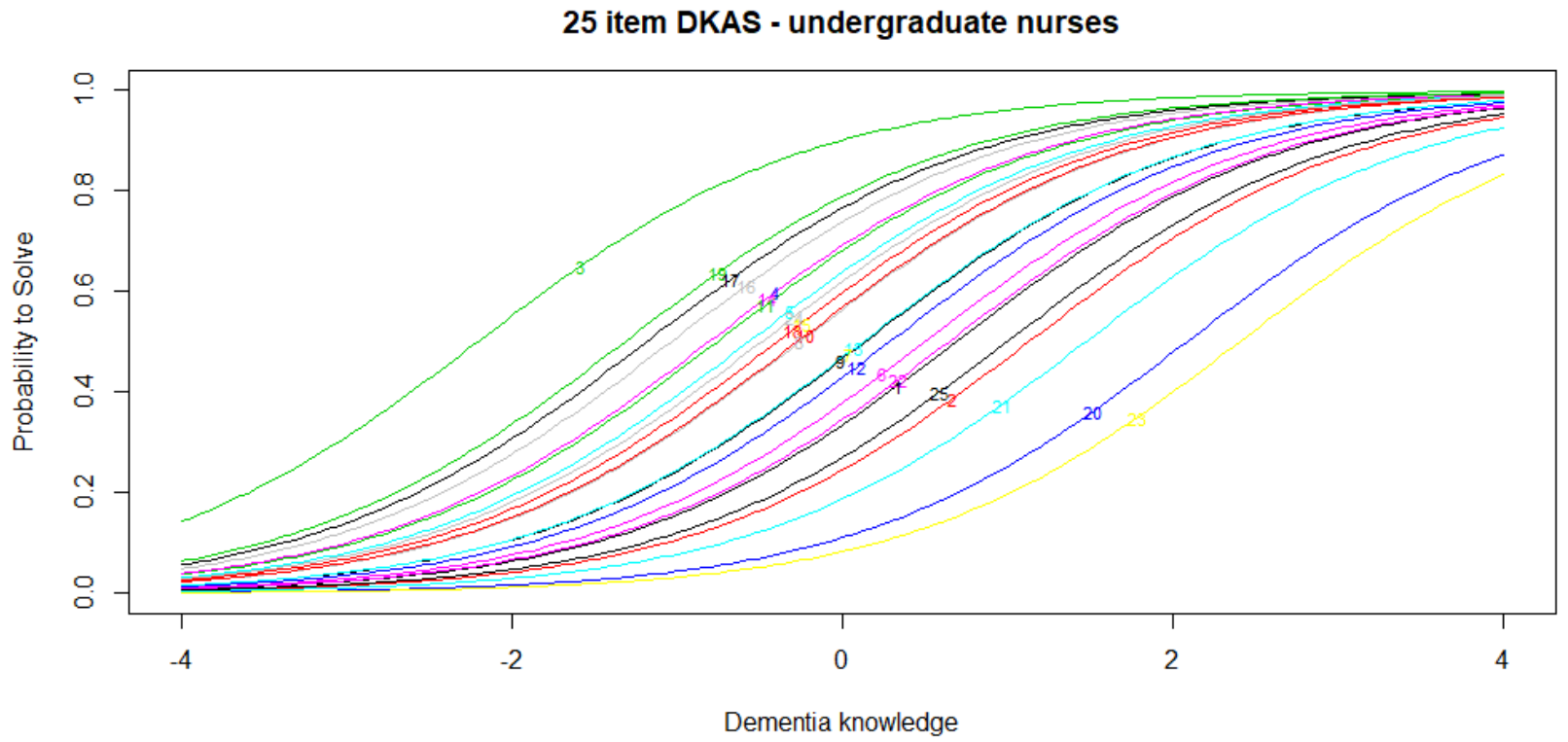
Table 7.2. Rasch CML estimations of difficulty parameters and associated standard errors (correct)

	DKAS items 1-25	***b* parameter (se)**
1	Most forms of dementia do not generally shorten a person's life	0.68 (0.11)
2	Blood vessel disease (vascular dementia) is the most common form of dementia	1.12 (0.11)
3	People can recover from the most common forms of dementia	-2.21 (0.26)
4	Dementia is a normal part of the ageing process	-0.81 (0.15)
5	Dementia does not result from physical changes in the brain	-0.57 (0.14)
6	Planning for end of life care is generally not necessary following a diagnosis of dementia	0.49 (0.11)
7	Alzheimer's Disease is the most common form of dementia	0.13 (0.12)
8	It is impossible to communicate with a person who has advanced dementia	-0.26 (0.13)
9	A person experiencing advanced dementia will not generally react to changes in their physical environment	0.13 (0.12)
10	It is important to correct a person with dementia when they are confused	-0.28 (0.13)
11	People experiencing advanced dementia often communicate through body language	-0.76 (0.15)
12	Uncharacteristic behaviours in a person experiencing dementia are usually a response to unmet needs	0.28 (0.11)
13	Medications are the most effective way of treating behavioural symptoms of dementia	0.12 (0.12)
14	People experiencing dementia do not generally have problems making decisions	-0.81 (0.15)
15	Movement is generally affected in the later stages of dementia	-0.40 (0.13)
16	Difficulty eating and drinking generally occurs in the later stages of dementia	-1.04 (0.16)
17	People with advanced dementia may have difficulty speaking	-1.19 (0.17)
18	People experiencing dementia often have difficulty learning new skills	-0.40 (0.13)
19	Daily care for a person with advanced dementia is most effective when it focuses on providing comfort	-1.32 (0.18)
20	Having high blood pressure increases a person's risk of developing dementia	2.08 (0.11)

	DKAS items 1-25	*b* parameter (se)
21	Maintaining a healthy lifestyle does not reduce the risk of developing the most common forms of dementia	1.47 (0.11)
22	Symptoms of depression can be mistaken for symptoms of dementia	0.64 (0.11)
23	The sudden onset of cognitive problems is characteristic of common forms of dementia	2.40 (0.12)
24	Exercise is generally beneficial for people with dementia	-0.49 (0.14)
25	Early diagnosis of dementia does not generally improve quality of life for people experiencing the condition	0.99 (0.11)

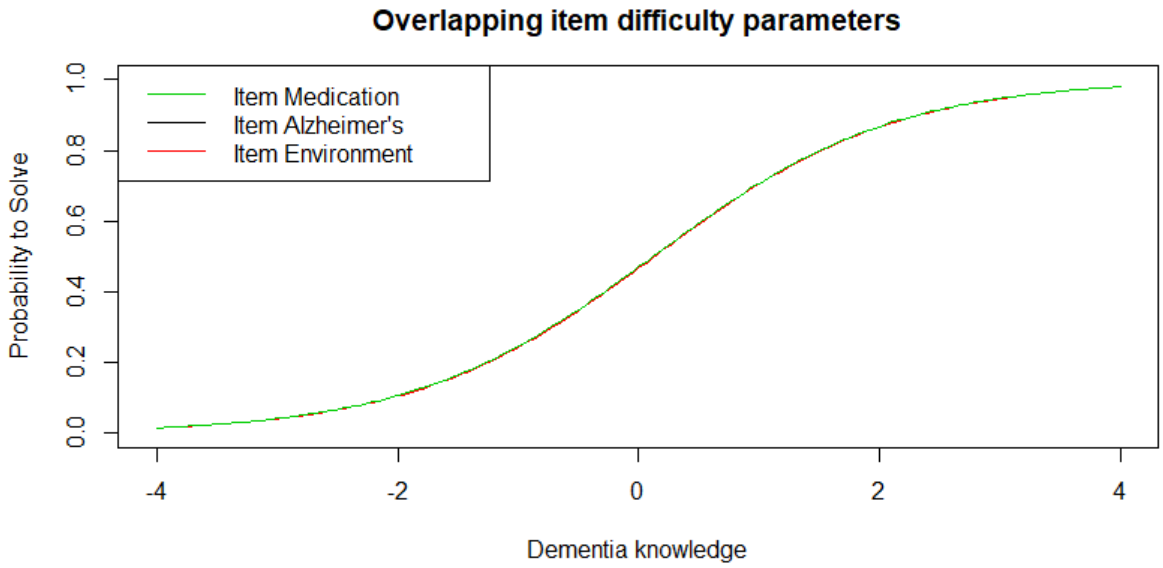
Item characteristics curves (ICCs) for the 25 DKAS items are displayed in *Figure 7.2*. The curves demonstrate graphically that the DKAS items were calibrated across a range of the latent scale, with small intervals between many of the ICCs, indicating a high level of measurement precision within the measurement range of the items. However, the ICCs effective measurement range also highlighted areas of the latent scale where measurement precision could be improved, notably at both extremes of the scale, where larger intervals between the ICCs can be seen. For example, the reasonably large interval between item 3 and item 19, both seen in green on the left-hand side of the plot, demonstrates the lack of DKAS items with difficulty parameter estimates between -2.21 and -1.32, meaning that estimation of dementia knowledge proficiency in respondents at this level would have been limited in precision.

Figure 7.2. Item characteristics curves for the Rasch-calibrated 25-item DKAS



Another factor that can affect precision within a scale is overlap of item difficulty parameter estimates (Mair, 2018). Across the 25 items there was some overlap of item difficulty parameters, for example item 7 (Θ = 0.13): “Alzheimer’s Disease is the most common form of dementia” (item code: Alzheimer’s); item 9 (Θ = 0.13): “A person experiencing advanced dementia will not generally react to changes in their physical environment” (item code: Environment); and item 13 (Θ = 0.12): “Medications are the most effective way of treating behavioural symptoms of dementia” (item code: Medication). The extent of overlap for these curves can be seen in *Figure 7.3*, where the ICCs are indiscernible from one another.

Figure 7.3. Three DKAS items with overlapping item characteristics curves



This item overlap was caused by the items having almost identical difficulty (hence the highly similar parameter estimates). These overlapping items increase participant burden unnecessarily, given the items cannot contribute to the measurement precision of the overall scale, therefore, there would need to be a compelling reason to retain more than one of these three items (Andrich and Marais, 2018).

7.3.3 Person parameters

Rasch-calibrated person parameters were examined to determine distribution and their relationship with the DKAS item parameters. The person-item map *(Figure 7.4)* shows the distribution of persons and items along the KoD latent trait continuum. As per similar plots in previous chapters, item distribution is shown in the lower panel where these can be inspected in relation to the person distribution in the upper panel.

Examination of item locations relative to person distribution showed that six of the DKAS items (see annotation (a) in Figure 7.4) were too easy to suit the ability range of the undergraduate nurse respondents, suggesting poor targeting of measurement, to some extent. Annotation (b) on the person-item map highlights a proportion of respondents with estimated ability levels higher than the difficulty parameter estimates of any DKAS items. However, the 19 DKAS items that span the central area of the map were capable of precise measurement across of the majority of person ability estimates.

Figure 7.4. Person-item map in sorted order

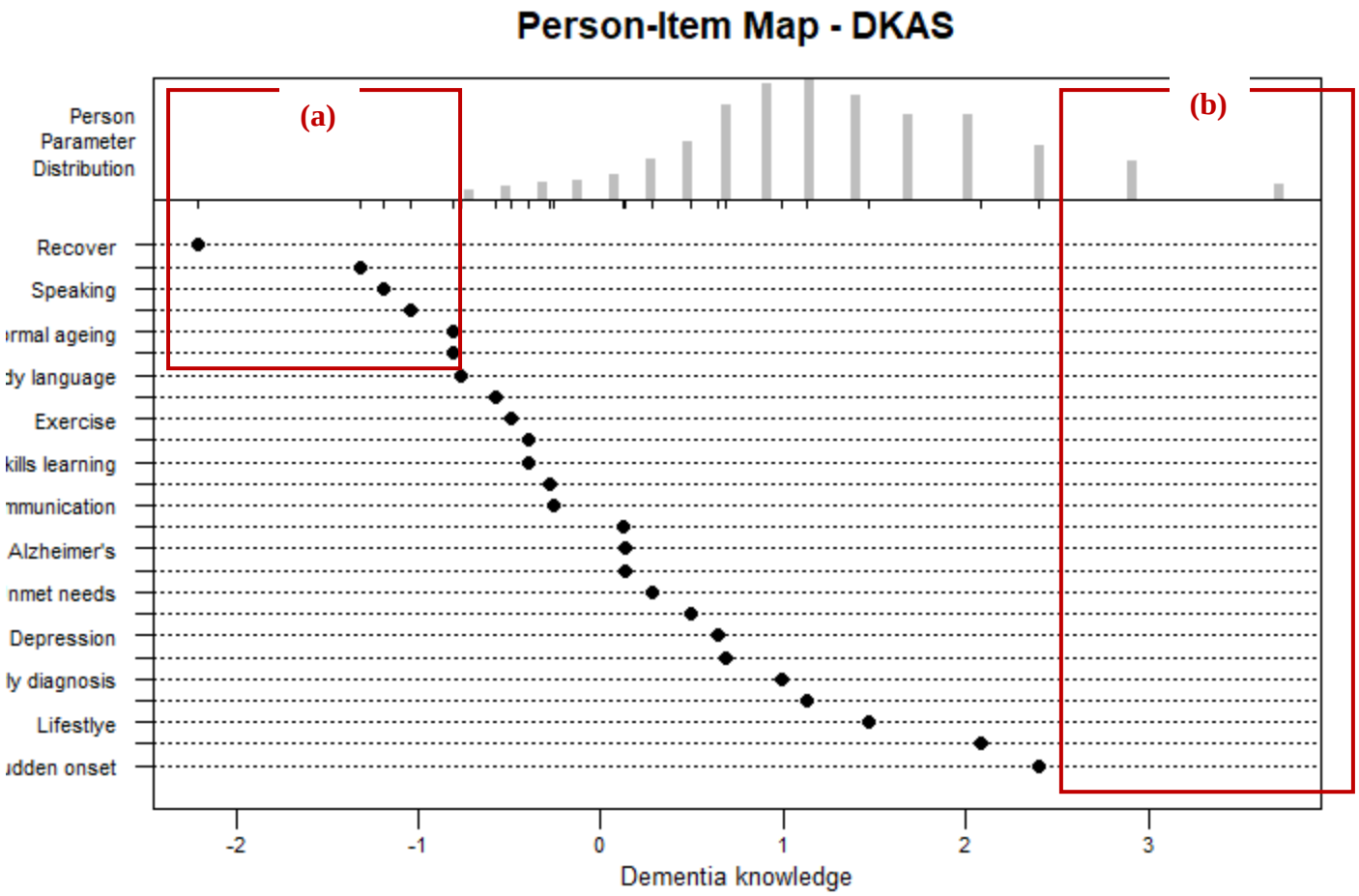


Table 7.3 below shows total scores to the DKAS and their Rasch-calibrated proficiency estimates. This table demonstrates the nonlinear transformation of raw scores to interval measures; a pragmatic benefit of the Rasch model over classical test theory methods (Irwing, Booth and Hughes, 2018).

Table 7.3: Observed scores and their associated Rasch-calibrated ability estimates

Score	**Ability parameters (Std. Err.)**	**Freq.**	**Expected Score**
0	-5.43 (1.98)	0	0.46
1	-4.24 (0.48)	0	1.40
2	-3.64 (0.25)	0	2.35
3	-3.22 (0.16)	0	3.31
4	-2.88 (0.12)	0	4.27
5	-2.60 (0.09)	0	5.23
6	-2.35 (0.07)	0	6.19
7	-2.12 (0.06)	0	7.16
8	-1.90 (0.05)	0	8.12
9	-1.70 (0.05)	1	9.09
10	-1.50 (0.04)	3	10.06
11	-1.31 (0.04)	5	11.03
12	-1.12 (0.04)	6	12.00
13	-0.93 (0.04)	9	12.98
14	-0.73 (0.04)	16	13.94
15	-0.54 (0.05)	25	14.92
16	-0.33 (0.05)	43	15.89
17	-0.12 (0.06)	54	16.86
18	0.11 (0.07)	56	17.84
19	0.35 (0.08)	48	18.81
20	0.62 (0.10)	38	19.78
21	0.93 (0.13)	38	20.75

Score	Ability parameters (Std. Err.)	Freq.	Expected Score
22	1.28 (0.18)	23	21.72
23	1.73 (0.26)	15	22.68
24	2.36 (0.50)	4	23.62
25	3.58 (2.06)	0	24.55

The following sections will report on the fit of the DKAS data to the Rasch model, including model and item fit using a selection of methods for splitting the sample to detect subgroup invariance.

7.3.4 Fit statistics

To test for the assumption of local independence and any violations of unidimensionality, the non-parametric testing framework offered by Ponocny (2001) was used. Local dependence was identified by inspection and identification of increased inter-item correlations, whereas potential multidimensionality was identified through identification of decreased inter-item correlations. Under this framework, a global test for local dependence involved examination of deviations between the observed and expected inter-item correlations (Koller and Hatzinger, 2013).

Of 300 item-pairs tested, 31 showed local dependence, suggesting that some items violated this Rasch assumption by not providing independent information; items 15, 16, and 17 were flagged across multiple item-pairs. Seven item-pairs suggested multidimensionality, with items 2 and 7 being flagged more than once. The global test for local dependence did not hold, returning a significant p-value, indicating that this assumption of the Rasch model was not fulfilled. Given the number of item-pairs flagged for local dependence was more than triple that of the number flagged for multidimensionality, the lack of global fit was likely due to response dependence rather than violations of unidimensionality (Andrich and Marais, 2019).

Another method to assess for response dependence is to inspect the standardised residuals for patterns (Marias and Andrich, 2008). *Table 7.4* reports the correlations between the standardised residuals of DKAS items; all correlations were positive, however a number of considerably large correlations were evident. Violations of item independence, as noted here in the DKAS items, warrant potential cause for concern given that simulation studies have demonstrated that response dependence can potentially lead to increased variance in person parameter distribution, and increased reliability estimates (Andrich and Marais, 2019).

Table 7.4. Correlations between standardised item residuals for the DKAS

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25
1																									
2	0.99																								
3	0.42	0.75																							
4	0.75	0.85	0.15																						
5	0.36	0.99	0.54	0.34																					
6	0.03	0.97	0.74	0.99	0.41																				
7	0.90	0.00	1.00	0.95	0.96	0.33																			
8	0.96	0.61	0.21	0.57	0.89	0.73	0.99																		
9	0.09	0.27	0.47	0.10	0.79	0.88	0.42	0.52																	
10	0.26	0.88	0.75	0.51	0.98	0.82	0.82	0.43	0.20																
11	0.20	0.91	0.68	0.99	0.04	0.24	0.46	0.86	0.05	0.44															
12	0.43	1.00	0.99	0.71	0.91	0.32	0.90	0.58	0.08	0.47	0.06														
13	0.41	0.78	0.97	0.51	0.87	0.47	1.00	0.01	0.20	0.38	0.43	0.18													
14	0.06	0.99	0.39	0.70	0.13	0.12	0.91	0.98	0.87	0.37	0.15	0.05	0.98												
15	0.07	0.93	0.64	0.88	0.55	0.45	0.28	0.97	0.13	0.47	0.06	0.22	0.80	0.00											
16	0.65	0.96	0.58	0.66	0.84	0.94	0.18	0.51	0.25	0.90	0.01	0.03	0.80	0.00	0.00										
17	0.14	1.00	0.80	0.70	0.24	0.06	0.46	0.78	0.47	0.50	0.00	0.02	0.62	0.00	0.00	0.00									
18	0.57	0.79	0.85	0.97	0.17	0.84	0.64	0.80	0.09	0.95	0.09	0.82	1.00	0.15	0.28	0.00	0.00								
19	0.40	0.97	0.46	0.40	0.13	0.77	0.92	0.20	0.86	0.67	0.63	0.27	0.84	0.40	0.86	0.62	0.92	0.22							
20	0.32	0.97	0.53	0.72	0.05	0.94	1.00	0.65	0.22	0.56	0.18	0.24	0.34	0.36	0.84	0.35	0.25	0.57	0.25						
21	0.28	0.98	0.55	0.77	0.29	0.69	1.00	0.52	0.75	0.68	0.15	0.02	0.84	0.86	0.83	0.23	0.12	0.86	0.54	0.00					
22	0.79	0.89	0.81	0.34	0.65	0.99	0.94	0.85	0.98	0.66	0.12	0.04	1.00	0.70	0.90	0.33	0.98	0.95	0.37	0.07	0.06				
23	0.23	0.98	0.67	0.16	0.97	0.75	0.97	0.32	0.37	0.87	1.00	0.81	0.79	0.94	0.65	0.90	1.00	0.86	1.00	0.83	0.98	0.99			

24	0.10	0.99	1.00	0.72	0.03	0.39	0.82	0.89	0.09	0.04	0.02	0.00	0.99	0.73	0.52	0.18	0.04	0.19	0.29	0.02	0.02	0.37	0.52	
25	0.26	0.99	0.94	0.35	0.32	0.37	0.90	0.88	0.50	0.95	0.47	0.22	0.11	0.62	0.76	0.68	0.23	1.00	0.85	0.03	0.00	0.37	1.0	0.01

To further investigate the fit of the Rasch model to the DKAS data, model and item fit indices were examined to determine whether the Rasch assumption of measurement invariance was upheld. Andersen's (1973) LR-tests and Wald tests were performed using multiple sample split criteria in order to determine whether item parameters were invariant across person subgroups.

7.3.5 Model fit

Model fit was examined using mean and median split criteria and the two covariates reported in *Table 7.5*.

Table 7.5: Undergraduate nurse responses to two external binary covariates for examination of Rasch model fit

	Yes	**No**
Do you know or have you known someone who lives with dementia? (Code: Known)	246 (64%)	137 (36%)
Do you work or have you ever worked with people who live with dementia? (Code: Worked)	199 (52%)	185 (48%)

Application of the LR-test splitting the sample according to the *mean* gave a significant result (p = 1.39e-05), as did the median split criterion indicating that likelihoods differed across these two groups, therefore the assumptions of the Rasch model were violated when the sample was split and callibrated at alternative central values (mean and medians). Model fit statitics were more favourable using the split on covariate 'Known', with Andersen LR-test results of (p =0.292), however not so favourable using the 'Worked' covariate split (p = 0.009). Overall, the likelihoods differed across the groups according to three out of four subgroup splits which suggested that the Rasch model did not have a good fit to the DKAS data. Sample size may have contributed here since there may be high power to reject.

7.3.6 Item fit

Following on from the model fit results, tests for item fit were performed using mean and median subgroup splits, and the covariates 'Known' and 'Worked'. Results of the Wald tests for individual item fit can be seen in *Table 7.5*, with misfitting results highlighted in bold. The majority of items showed acceptable fit. Two items showed misfit across more than one sample split: item 2 (Vascular) and item 9 (Environment), hereby suggesting that both items be further evaluated quantitatively and qualitatively for potential elimination.

An additonal method for identification of misfitting items is the Bond-and-Fox (2007) Pathway map which displays item locations against their infit t-statistics. Figure 7.5 suggests two misfitting items (Item 2: Vascular, and item 20: Hypertension).

Figure7.5: Bond-and-Fox pathway map highlighting misfitting DKAS items

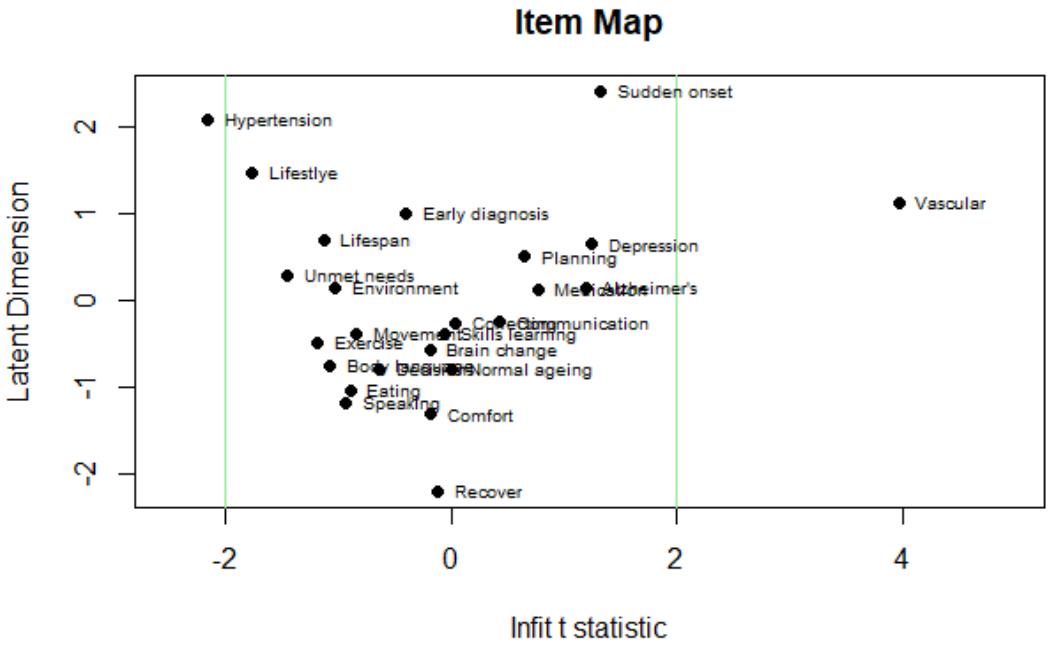


Table 7.5: Item fit statistics using three sample subgroup split criteria. p-values of misfitting items are highlighted in bold.

	DKAS item code	MEDIAN split		KNOWN split		WORKED split	
		z-statistic	p-value	z-statistic	p-value	z-statistic	p-value
1	Lifespan	-1.77	0.08	1.20	0.23	1.81	0.07
2	**Vascular**	**3.46**	**0.00**	0.79	0.43	**2.38**	**0.02**
3	Recover	0.40	0.69	-1.95	0.05	0.23	0.82
4	Normal ageing	0.15	0.88	-0.82	0.41	0.67	0.50
5	Brain changes	-1.06	0.29	1.10	0.27	-0.64	0.52
6	Planning	**2.23**	**0.03**	-0.59	0.55	-1.45	0.15
7	Alzheimer's	**3.59**	**0.00**	0.73	0.47	-0.11	0.92
8	Communication	0.57	0.57	-1.84	0.07	0.64	0.52
9	**Environment**	-0.80	0.43	**-2.69**	**0.01**	**-2.24**	**0.03**
10	Correcting	1.85	0.06	-0.06	0.95	0.78	0.44
11	Body language	-1.14	0.25	-0.24	0.81	**-2.98**	**0.00**
12	Unmet needs	-0.30	0.76	0.26	0.79	-1.83	0.07
13	Medication	0.86	0.39	-0.59	0.55	0.01	0.99
14	Decisions	-0.62	0.54	0.48	0.63	0.04	0.97

	DKAS item code	MEDIAN split		KNOWN split		WORKED split	
		z-statistic	p-value	z-statistic	p-value	z-statistic	p-value
15	Movement	-0.70	0.48	1.72	0.09	**2.55**	**0.01**
16	Eating	-1.54	0.12	1.75	0.08	0.51	0.61
17	Speaking	-1.98	0.05	0.03	0.98	0.69	0.49
18	Skills learning	0.90	0.37	0.60	0.55	0.93	0.35
19	Comfort	0.54	0.59	0.21	0.84	-0.84	0.40
20	Hypertension	-1.20	0.23	-0.53	0.60	-1.38	0.17
21	Lifestyle	-1.08	0.28	0.40	0.69	-1.04	0.30
22	Depression	**2.84**	**0.00**	0.87	0.38	0.30	0.76
23	Sudden onset	**2.61**	**0.01**	0.46	0.65	1.14	0.26
24	Exercise	-1.00	0.32	-0.23	0.82	-0.05	0.96
25	Early diagnosis	0.49	0.62	0.27	0.78	0.11	0.91

Tests for model and item fit suggested that the DKAS data dit not show adequate fit to the Rasch model, with questions arising around violations of unidimensionality and local independence. To further investigate these problems with the Rasch-calibrated DKAS, DIF analyses were conducted to determine whether items were behaving unexpectedly, and therefore contributing to misfit of the data to the model.

7.3.7 Differential item functioning

As per Chapter 6, Lord's (1980) chi-squared method was used to examine the DKAS dataset for the presence of DIF. The two subgroup covariates 'Known' and 'Worked' were included in this DIF analysis.

Table 7.7 shows the results of the DIF analysis. Across the two sample splits, four items (those highlighted in bold) were flagged as DIF items.

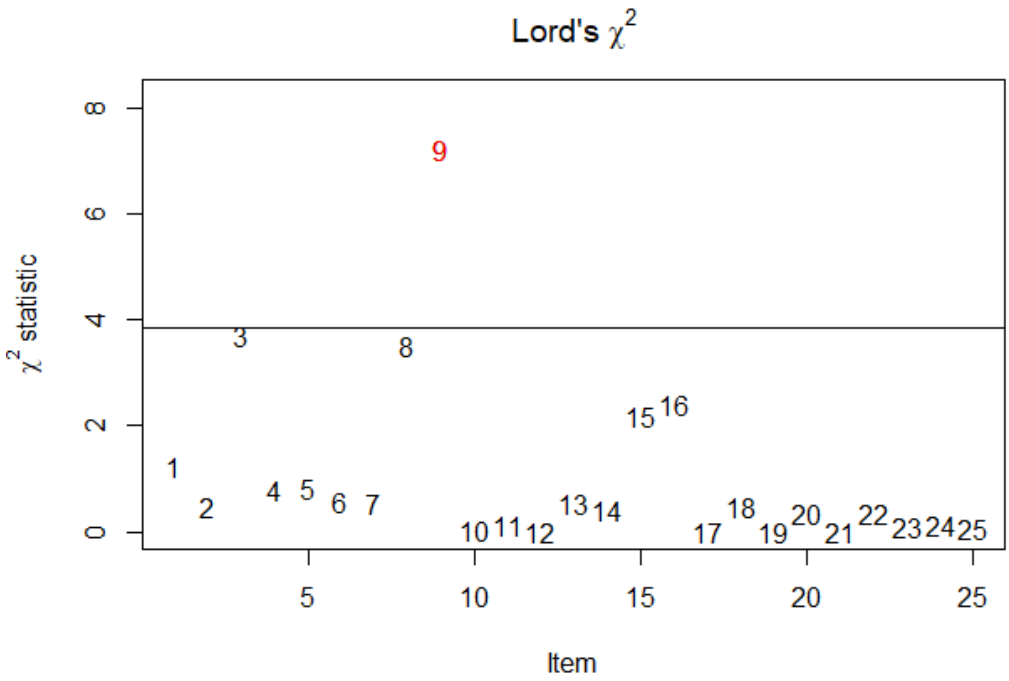
Table 7.7: DKAS - Lord's (1980) chi-square values for sample split using the 'Known' and 'Worked" covariates.

	DKAS item	**'Known' split**		**'Worked' split**	
		Lord's chi-sq.	**p-value**	**Lord's chi-sq.**	**p-value**
1	Lifespan	0.99	0.32	2.64	0.10
2	**Vascular**	0.37	0.54	**4.66**	**0.03**
3	Recover	3.60	0.06	0.04	0.85
4	Normal ageing	0.71	0.40	0.34	0.56
5	Brain changes	0.91	0.34	0.43	0.51
6	Planning	0.44	0.51	1.99	0.16
7	Alzheimer's	0.33	0.56	0.03	0.87
8	Communication	3.30	0.07	0.29	0.59
9	**Environment**	**6.80**	**0.01**	**4.64**	**0.03**
10	Correcting	0.02	0.88	0.44	0.51
11	**Body language**	0.09	0.76	**8.21**	**0.00**
12	Unmet needs	0.02	0.89	3.10	0.08

	DKAS item	'Known' split		'Worked' split	
		Lord's chi-sq.	**p-value**	**Lord's chi-sq.**	**p-value**
13	Medication	0.43	0.51	0.00	0.95
14	Decisions	0.14	0.71	0.00	1.00
15	**Movement**	2.35	0.12	**5.52**	**0.02**
16	Eating	2.54	0.11	0.19	0.66
17	Speaking	0.00	0.97	0.38	0.54
18	Skills learning	0.23	0.64	0.67	0.41
19	Comfort	0.02	0.89	0.69	0.41
20	Hypertension	0.42	0.52	1.75	0.19
21	Lifestyle	0.05	0.82	1.04	0.31
22	Depression	0.48	0.49	0.05	0.83
23	Sudden onset	0.08	0.78	1.13	0.29
24	Exercise	0.09	0.77	0.01	0.92
25	Early diagnosis	0.02	0.90	0.00	0.96

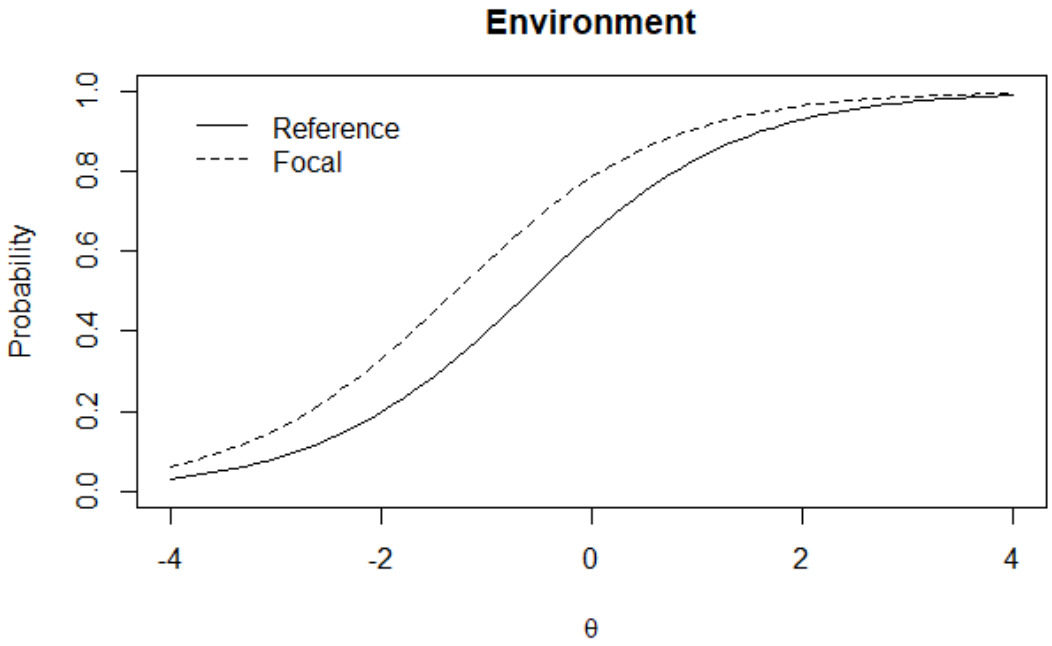
Using the 'Known' sample split, item 9 '*A person experiencing advanced dementia will not generally react to changes in their physical environment'* (item code: Environment) was a DIF item. The DIF summary is displayed in *Figure 7.6*, where it is shown that most of the 25 items had low DIF test statistics, with only one item positioned above the detection threshold line.

Figure 7.6. Plot showing item 9 as a DIF item under the 'known' sample split



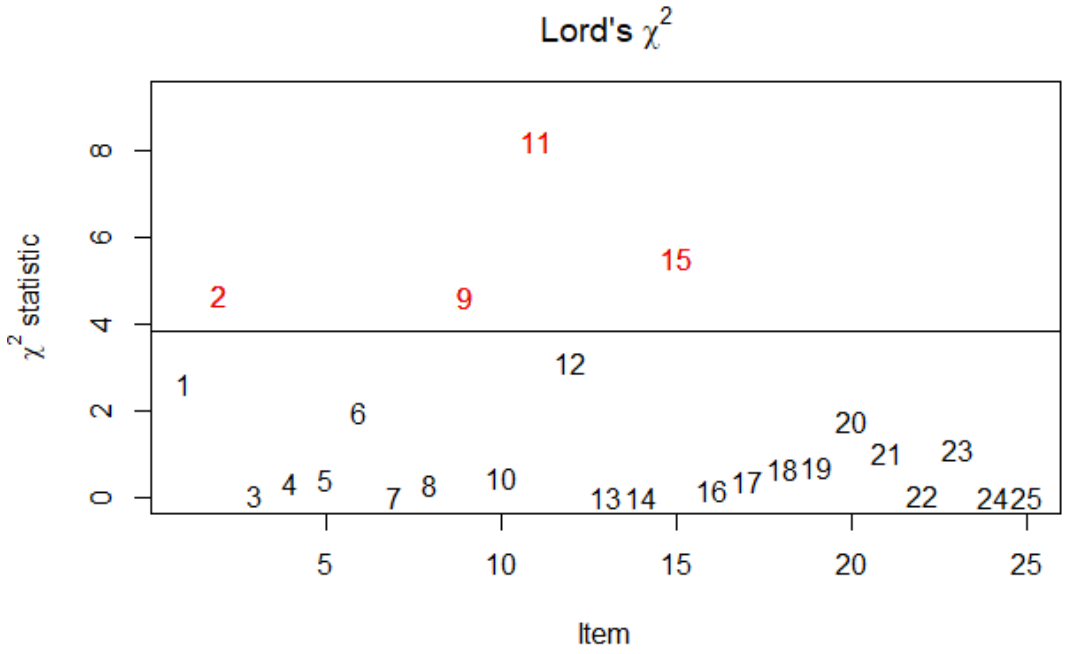
The DIF detected in item 9 (Environment) is displayed visually in *Figure 7.7*, below. Examination of these ICCs showed that this item was biased in favour of respondents who had known someone living with dementia (Reference curve), therefore giving these students an advantage on this specific DKAS item.

Figure 7.7. DIF item 9 under 'Known' split



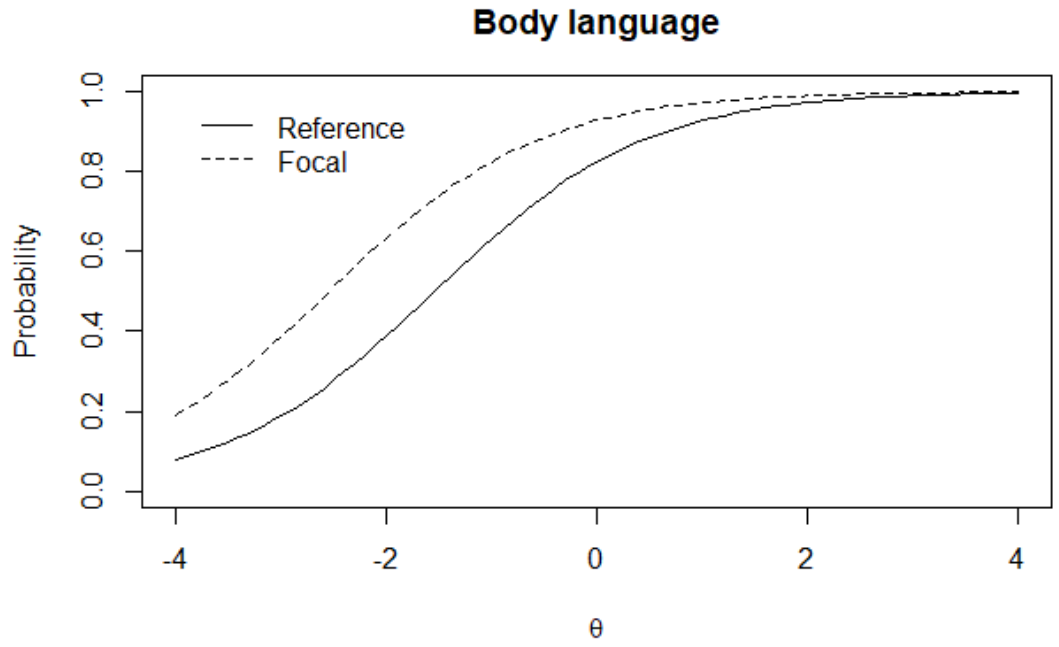
The DIF summary using the 'Worked' sample split is displayed in *Figure 7.8*. Using this sample split, DIF was detected in four items, as follows: Item 2 (Vascular), item 9 (Environment), item 11 (Body language), and item 15 (Movement).

Figure 7.8. DIF summary for 'Worked' split



The ICCs shown in *Figure 7.9* demonstrate the extent of DIF in item 11 (Body language). Comparison of the curves in this plot show that respondents who had worked someone living with dementia (Reference curve) scored higher on item 11 than those who had not worked with people living with dementia (Focal curve). Therefore, respondents who had worked with people living with dementia had an advantage on this item of the scale, regardless of their position on the latent trait.

Figure 7.9. DIF item 11 under the 'Worked' split



Overall, four items were identified as differentially functioning, as follows:

- Item 2: '*Blood vessel disease (vascular dementia) is the most common form of dementia'* (Code: Vascular)
- Item 9: '*A person experiencing advanced dementia will not generally react to changes in their physical environment'* (Code: Environment)
- Item 11: '*People experiencing advanced dementia often communicate through body language'* (Code: Body language)
- Item 15: '*Movement is generally affected in the later stages of dementia'* (Code: Movement)

Item 9 was flagged as a DIF item using both covariate sample splits. These four items also showed misfit based on the item fit statistics, indicating that they should be reviewed for elimination; of the four items, 2 and 11 showed the greatest degree of item misfit.

A key argument in Rasch modelling is that misfit suggests weaknesses in the data that cannot be resolved by the inclusion of additional parameters, such as the item discrimination parameter in 2-parameter logistic models (Hagquist, Bruce and Gustavsson, 2008), but that the test in question should be revised, re-examined, and

ideally improved based upon new or updated data (Andrich and Marias, 2019). This notion is discussed further in the following section.

7.4 Summary

The aim of the study in this chapter was to examine the properties of the Dementia Knowledge Assessment Scale (DKAS) through application and calibration of the Rasch model. Given that the DKAS had been subject to prior psychometric evaluation under the classical test theory framework, the key aim was to determine whether item response theory-based methods could provide additional information about this set of items when administered in a sample of undergraduate student nurses, and to what extent the DKAS data showed sufficient fit to the Rasch model.

Comparably to the KIDE results detailed in *Chapters 5 and 6*, the DKAS items demonstrated an element of mistargeting in this sample of undergraduate nurses, with six of the 25 items being too easy for these undergraduate nurse respondents. Further, there were a small proportion of respondents whose ability levels were greater than the DKAS was able to measure with any precision. This finding highlights a key benefit of IRT scoring methods over CTT scoring methods: none of the respondents scored 100% on the DKAS, which, under the CCT framework, would have suggested that no respondents had sufficient ability to match the maximum difficulty of the test (Embretson and Reise, 2013). By applying the Rasch model, items and persons were positioned along the same latent scale, transforming raw scores to person estimates which varied across the latent scale (Andrich and Marais, 2019). Importantly, these features of the Rasch model lead to the sum score being a sufficient statistic for the latent trait estimate, and therefore grouping by sum score is grouping by latent trait (Boone, Staver and Yale, 2013).

Given the fit indices that suggested misfit of the data to the Rasch model, as well as violations of local independence, it can be concluded that misfit was a result of weak data, rather than weakness related to the formal Rasch model (Andrich, 1988).

The findings from this chapter will be discussed in depth and in the context of the PhD study throughout Chapter 8.

7.5 Conclusions

This chapter has reported on the Rasch-model-calibration of a set of undergraduate nurse respondents to the 25-item Dementia Knowledge Assessment Scale. Tests for model and item fit suggested that the data did not show adequate fit to the Rasch model, with nine items being highlighted for further evaluation and potential elimination due to misfit and bias detection. Despite this element of misfit, application and calibration of the Rasch model to the DKAS dataset was justified, given that this was the first known attempt at item response theory-model calibration of the DKAS, over classical test theory methods.

CHAPTER 8

DISCUSSION AND CONCLUSIONS

The purpose of this chapter is to collate the results from this PhD thesis and discuss the findings in the context of the wider literature; this discussion frames the contribution to knowledge that this thesis provides. Implications for research and practice within dementia education will be discussed, and statements will be made about the strengths and limitations of the work contained in this thesis. Following this is discussion around the direction of future work.

8.1 Overview and context of the study

The global, national, and individual burden caused by dementia is well documented (Nichols *et al*., 2019). Due to the increasing prevalence of dementia and national incentives to facilitate independent living for as long as possible, there is a need, now more than ever, for some degree of health literacy across multiple population groups (Scerri, Innes and Scerri, 2020). In lay population groups, a basic level of knowledge of dementia (KoD) is required in order to understand modifiable risk factors, recognise early signs and symptoms, and provide support to others living with dementia in their communities. A recent systematic review reported that, despite national and global initiatives to increase dementia awareness, public knowledge of modifiable risk factors remains low (Parial *et al*., 2021) In population groups who provide formal care, support and treatment, more in-depth levels of dementia knowledge are required; developments in healthcare practices are advancing rapidly and in these groups it is important to be equipped with the latest evidence-based research and practice.

Measurement of dementia knowledge across populations is necessary to inform public health campaigns and to direct educational and clinical competencies; shifts or trends in dementia knowledge cannot be measured without baseline estimates from which to reliably benchmark change. To measure latent concepts such as dementia knowledge, tests or instruments must be not only available, but also have evidence of

development and evaluation using robust psychometric methods and guidelines. Such guidelines are widely available and cited heavily in educational and health-related research, for example, in Thissen and Wainer (2001), Embretson and Reise (2013), and Streiner and Norman (2015).

The following sections will outline the key objectives of this PhD thesis, and how these objectives were addressed by the research. The primary research question that informed these objectives was: Can item response theory (IRT) applications be used to make dementia knowledge tests and testing more informative?

8.1.1 Research objective One

The first objective of this PhD was as follows:

1. Explore the dementia knowledge testing landscape through identification and psychometric appraisal of currently available measurement instruments.

The systematic review in *Chapter 3* demonstrated that, although growing, the field of measurement of dementia knowledge is small. A prior review of dementia knowledge instruments by Spector *et al*. in 2012 identified five relevant measures; the review in this thesis identified 14 instruments that were developed to measure dementia knowledge across various population groups, including laypeople, informal carers, and all grades of healthcare staff. Critical psychometric appraisal of these instruments determined that reporting of psychometric strength and further evaluation was very limited. Further, the review found outdated item content and content that was not appropriate for use in UK-based samples. Key findings from the review in relation to this thesis were as follows:

- Many instruments were author-developed and have not been used widely since their development
- These is an overall lack of dimensionality assessment
- Instruments have been administered in improper sample populations
- There is widespread improper use of Cronbach's coefficient alpha
- The field of dementia knowledge measurement is deeply anchored by classical test theory (CTT) methods

- There is an overall lack of IRT-based methods, nor are IRT frameworks cited in recommendations for further development of the instruments

These findings mirror results reported in the wider literature, with reviews on measurement instruments in health-related outcomes and education also describing significant limitations in psychometric evaluation and reporting (Trevena & Waters, 2014; Yang *et al*., 2015; Davies, Waters & Marshall, 2016; Clari, *et al.*, 2016; Lui, Kim and Alessio, 2020). This further highlights the importance of encouraging standardisation of psychometric evaluation methods and reporting procedures. Useful guidelines have been developed by Mokkink *et al*., (2010) and are widely cited in health-status questionnaire evaluations. These guidelines have also been discussed by Rosenkoetter and Tate (2018) in the context of other available frameworks for psychometric evaluation of measurement instruments more generally. However, as this review has emphasised, such guidelines are not widely used or cited in dementia knowledge measurement.

The findings from this systematic review highlighted a gap in the field of dementia knowledge assessment and emphasised an opportunity to examine the properties and performance of KoD instruments under a more advanced framework of measurement theory. As such, the remaining research objectives for this thesis were as follows:

2. Acquire and/or generate appropriate dementia knowledge datasets of responses to currently available instruments for further examination using IRT-based modelling.
3. Application and calibration of the Rasch model to evaluate measurement range, measurement precision, fairness to test takers (presence of item bias), and areas for improvement in methods to help understand KoD instruments.

8.1.2 Research objective Two

A new academic collaboration with nurse researchers at a university in Scotland resulted in a dataset of responses (Dataset one) to one of the KoD instruments from the systematic review phase being shared for secondary analysis. Dataset one comprised (n=521) participant responses to 16 dichotomous KoD items from the Knowledge in Dementia Scale (KIDE) (Elvish *et al*., 2014).

On completion of the secondary analysis phase, a second dataset (Dataset two) was generated by collecting responses from undergraduate student nurses to two KoD instruments; the KIDE (Elvish *et al*., 2014) and the Dementia Knowledge Assessment Scale (DKAS) (Annear *et a*l., 2017). Dataset two comprised (n=404) participant responses to 41 dichotomous KoD items. Given the sample size and structure of these datasets, both were appropriate candidates for application of IRT-based methods.

8.1.3 Research objective Three

Rasch analysis techniques were used to explore the measurement properties of the KIDE and DKAS responses, including targeting, local independence, model and item fit statistics, and differential item functioning. The results of these studies are discussed in detail in the following sections, 8.2 and 8.3.

8.2 Rasch calibration of the KIDE in two population samples

Across *Chapter 5* and *Chapter 6*, two sets participant responses were independently calibrated along a unidimensional latent trait by fitting the KIDE data to the Rasch model. In dataset one, 11 of the 16 items showed sufficient fit to the Rasch model, whereas in dataset two, all 16 KIDE items demonstrated sufficient fit. In dataset two, fit statistics and tests for local dependence garnered poorer results when using mean and median sample sub splits as opposed to independent variables as split criteria. In this respect, the misfit of five items demonstrated in Dataset one may have been a disadvantage caused by the lack of independent covariates to act as split criteria.

In both datasets, the three fundamental Rasch model assumptions of unidimensionality, monotonicity, and local independence were upheld to some degree, albeit in a reduced set of items for the healthcare professional sample.

8.2.1 Targeting and reliability

In measurement, it is desirable to have items that span a large range of ability, this range should ideally be wider than the ability range of the target population, if all

abilities are to be captured with any precision (Irwing, Booth and Hughes, 2018). Measurement cannot be effective if sets of items display poor targeting and as such, the severity of items must correspond directly to the ability levels in the population of interest (Andrich and Marias, 2019). A key aim of examining the measurement properties of dementia knowledge instruments was to determine which items provided adequate information and reliability indices to be capable of precise measurement; thus, determining the 'effective measurement range' of a test. The KIDE was poorly targeted to the *Dementia Champions* in *Dataset one*, therefore the effective measurement range was extremely limited. There was an imbalance between item and person locations when calibrated along the same latent scale, with the items being too easy for the average ability level displayed by the Dementia Champions respondents. In short, poor targeting resulted in a reduced measurement range and in turn, poorer reliability of the scale and the items it contains.

Given the ceiling effects in Dataset one, with almost a fifth of the Dementia Champions respondents scoring 100% correct on the KIDE, measurement reliability in this sample was already significantly reduced. This in turn diminished the functionality of the Rasch model since extreme scores cannot be used to estimate person parameters (Hagquist, Bruce and Gustavvson, 2009); this was also reflected in the high standard errors associated with the least difficult KIDE items (those for which all or most respondents answered correctly).

For *Dataset two*, the *undergraduate students* were anticipated to have less knowledge of dementia than the Dementia Champions participants, therefore potentially addressing the problems with targeting and reliability that emerged in *Chapter 5*. The Dataset two KIDE did indeed capture a wider range of ability parameters than the Dataset one KIDE, at (θ = -3.32 to 2.95), versus (θ = -2.92 - 2.05), respectively. However, the proportion of undergraduate respondents with below-average ability was small, resulting in five of the 16 items being flagged for further evaluation, despite showing no statistical misfit. Importantly, this finding highlights the difference between the *calibrated* measurement range and the *effective* measurement range in relation to scale reliability; the difference between the two was significant in the undergraduate nurse sample, even without the influence of the ceiling effects that compromised the reliability of the Dementia Champions data.

The following section discusses unrepresented trait locations in the Rasch-calibrated KIDE, and how these relate to scale reliability.

Trait locations unrepresented by KIDE item content

On calibration of item and person locations on the latent trait, there were notable portions of the latent scale that were not represented by KIDE item content. For example, in Dataset one, there were no items located between θ = 2 to θ = 3.5, meaning that KIDE was unable to capture the true ability of the Dementia Champions participants at the positive end of the scale. Similarly in Dataset two, there were no KIDE items representing θ estimates of -1.07 to -0.06, and 1.79 to 2.95. As discussed above, poor targeting does not support the aims of measurement (Andrich, 1988), and in this case led to lower measurement precision and reliability.

Item redundancy

Although not directly responsible for compromising measurement precision, redundant items increase burden for the test taker and do not contribute any information to the test, therefore such items should be further investigated and potentially eliminated (Andrich and Marias, 2019). On examination of KIDE item parameters, redundant items emerged in both Dataset one and Dataset two. In instances of item redundancy, the decision of whether to retain an item or not is reliant on qualitative evaluation of the item and researcher judgement. For example, in the Dementia Champions data (Dataset one), two pairs of items exhibited an element of redundancy, as follows:

Item-pair a)

- **Item 1**: Permanent changes to the brain occur in most types of dementia (θ = -0.54)
- **Item 6**: Brain damage is the only factor that is responsible for the way people with dementia behave (θ = -0.61)

Item-pair b)

- **Item 2**: Dementia can be caused by a number of small strokes (θ = 0.98)

- **Item 10**: A person with dementia's history and background plays a significant part in their behaviour ($\theta = 1.05$)

Given that the KIDE had already been reduced and re-profiled under these analyses down to a strong 11-item set, further removal of items would not necessarily be desirable, and any further elimination would warrant serious consideration, since reliability is affected by scale length and might be reduced further when the number of items in a scale is reduced - even one at a time (Embretson and Reise, 2013).

However, some consideration might be given to eliminate item 1 (Brain changes) and retain items 6, 2, and 10 (Brain damage, Strokes, History/Behaviour), based on overlap of item content between items 1 and 6, and usefulness of the items 2 and 10 due to their positioning on the latent scale, being within the *effective* measurement range.

Two pairs of overlapping items also emerged in the Dataset two KIDE, indicating item redundancy. Were these analyses to be continued (for example, to publish results), item evaluation would continue using the methods mentioned above. However, it should be made clear in any form of IRT modelling, item elimination is an iterative process; items must be eliminated one at a time and the model recalibrated for evaluation of person and item parameters, and model fit statistics (Andrich, 1988).

8.2.2 Model and item fit

When data demonstrate acceptable fit to the Rasch model, this is an indication of true measurement of a latent trait and constitutes a high seal of approval for a scale (Andrich and Marias, 2019). In Dataset one, 11 of 16 items showed adequate fit, with the 5 items with the highest *b* parameters eliminated based on fit statistics. Dataset two demonstrated statistically sufficient fit across all 16 items. In future analyses, model and item fit would be iteratively re-evaluated on attempting to address the above mentioned problems with targeting.

8.2.3 Differential item functioning

Model fit was also examined using analyses to detect differential item functioning (DIF) where a priori specified sample groups show a lack of invariance (Bond and Fox, 2013). Examining DIF between groups can determine whether measurement differs as a function of sample characteristics; in biased items, any difference in respondent score may be due the item functioning differently between groups, rather than differences in dementia knowledge levels (Andrich and Marias, 2019). Investigation for DIF was conducted in Dataset two only, given the lack of demographic variables in Dataset one.

DIF analyses were carried out in the undergraduate nurse dataset using the covariates 'Do you know, or have you known someone who lives with dementia?' (Code: Known), and 'Do you work, or have you ever worked with people who live with dementia?' (Code: Worked). Given the current study has a focus on dementia knowledge measurement in healthcare staff and students, it was pertinent to understand whether prior experience with people with dementia had any effect on the way the KIDE items functioned in this sample.

In Dataset two, item 13: 'Dementia is a general term which refers to a number of different diseases', and item 14: 'A person with dementia's history and background plays a significant part in their behaviour' were identified as DIF items. Respondents who had worked with people with dementia were disadvantaged on item 13, whereas respondents who had known people with dementia had the advantage on item 14. DIF is generally not a sufficient reason alone to eliminate an item (Bond and Fox, 2013), but instead suggests that further investigations are warranted, for example, into how the item contributes to the construct of interest. Further, the size of these demographic groups may limit the extent to which interpretations about measurement consequences can be made. Rasch analysis facilitates investigation of the extent to which measurement instruments can appropriately distinguish between groups (Irwing, Booth and Hughes, 2018), however, the sample proportions in Dataset two were insufficient regarding gender, with only 8% of respondents being male. Tests must be capable of targeting heterogenous populations; heterogeneity was addressed in this study with regard to prior experience with people living with dementia, but unfortunately not with regard to gender.

8.2.4 Overall Rasch contribution to the KIDE

The above discussion of Rasch calibration of the KIDE in two datasets showcases the additional information about measurement properties of a test that can be gleaned by application of a simple but powerful IRT model. Current CTT-based evaluation of the KIDE comprises of face validity: reported as adequate based on qualitative consensus in the developers, and a strong internal consistency reliability of 0.72 in a sample of healthcare staff (Elvish *et al*., 2014).

The following sections discuss the results from the Rasch analysis of the DKAS instrument in Dataset two.

8.3 Rasch calibration of the DKAS in nursing undergraduates

In *Chapter 7*, a set of responses to the 25-item DKAS were fit to the Rasch model; item difficulty indices were examined in relation to the distribution of Θ in the undergraduate nurse respondents and the effect this has on the measurement properties of the test. Overall, the data did not show good fit to the Rasch model, with questions arising around response dependence and violations of unidimensionality. However, the question driving this PhD research asked whether Rasch model calibration of dementia knowledge tests could provide information about their usefulness and measurement properties that were unable to be captured by CTT methods alone. Sections 8.3.1 to 8.3.4 discuss these details in further depth.

8.3.1 Targeting and reliability

Appropriately matching tests to populations requires an assortment of mid-range items; the reliability of test scores is inherently reduced where the test contains items that are too easy or too difficult for the respondents (DeMars, 2010). The Rasch-calibrated item parameters captured a wide range of the latent scale of dementia knowledge, with items located at reasonably consistent intervals across the central portion of the scale. Person parameters were also associated with predominantly acceptable reliability estimates. This initially indicated a high level of measurement

precision of the DKAS in undergraduate nursing students. These reliability statistics increased confidence in the results, implying that patterns would be comparable if the methods were repeated in an additional sample (Andrich and Mair, 2019).

There were however some problems identified with regard to targeting. Although all 25 item locations had acceptable reliability estimates, six items contributed very little information to the scale, since they were located at the low end of the trait continuum with few corresponding person parameters. The remaining 19 items however showed promise of an item-set capable of precise measurement given the reliability estimates and locations in relation to person parameters.

There were a further two pairs of items that demonstrated redundancy, seen as overlapping *b* parameters. Thus, almost a third of the DKAS items did not contribute to precise measurement of knowledge within the effective measurement range. This is an important finding given CTT estimates of reliability have been reported as good, with Cronbach's alpha coefficients of $\alpha = 0.86$ (Annear et al., 2016) and $\alpha = 0.85$ (Annear et al., 2017) in participants of a 'Dementia Knowledge' MOOC. Further, these studies reported that there was no evidence of item redundancy. Such comparisons were not possible with the KIDE data, given the absence of further applications of the KIDE in the literature. This highlights the importance of multiple administrations of instruments where reliability estimates are evaluated and reported, to build their psychometric strength (Streiner and Norman, 2008).

Regarding comprehensive coverage of the latent scale by DKAS items, there were no items with *b* parameter estimates of between $\Theta = -2.21$ and $\Theta = -1.32$, however this had minimal impact on the precision of the scale since these locations had very few associated person parameters in the sample of undergraduate nurses. There was however a lack of item parameters between $\Theta = 1.47$ and $\Theta = 2.08$; a high proportion of students were parameterised between these locations, and as such, measurement precision was limited in students with proficiencies between these levels. The only way to address this would be to generate additional item content and recalibrate the Rasch model (Mair, 2018).

8.3.2 Response dependence

Violations of local independence are an indicator of problems or weakness in the dataset, potentially indicating multidimensionality or response dependence (Andrich and Marias, 2018). In the DKAS data, 31 out of 300 item-pairs showed local dependence, suggesting that a small group of items did not provide independent information. Although violations of item independence can lead to wider variance in person parameters and in turn, artificially improve reliability estimates (Marais and Andrich, 2008), the number of item-pairs flagged was relatively small and therefore only a minor effect was anticipated. Items that were frequently flagged for local dependence across the item-pairs were:

Item 15: *Movement is generally affected in the later stages of dementia*

Item 16: *Difficulty eating and drinking generally occurs in the later stages of dementia*

Item 17: *People with advanced dementia may have difficulty speaking*

These items are similar in that they all pertain to impaired skills as a result of advanced dementia. As such, it is possible that these items form a small additional construct within the DKAS, hereby contributing to response dependence and violation of unidimensionality (Marias and Andrich, 2008).

A confirmatory factor analysis conducted by the DKAS development team suggested a strong four-factor structure in a diverse sample of healthcare staff, healthcare students, and public groups (Annear *et al*., 2017), however a maximum likelihood (ML) factor analysis conducted as part of this PhD study (see Appendix 7) suggested the presence of either one or three factors, neither of which was strong enough to capture all 25 items. The first four eigenvalues of the ML solution were: 3.12; 1.41; 1.0; and 0.87, suggesting that the first factor was responsible for the majority percentage of variance, and therefore analysis continued by examining the DKAS as capturing a potentially unidimensional construct. The reasons for this discrepancy of factor solutions between the development study and the current PhD study were likely to be the differences in sample composition (heterogeneous vs. homogenous) and methods of test administration (digital, unsupervised vs. pencil and paper in a lecture theatre).

8.3.3 Model and item fit

As discussed in previous sections, misfit to the Rasch model is attributed to poor data rather than a weakness of the formal model (Wright and Mok, 2000). In simple stochastic models such as Rasch there will generally always be misfit to some extent, given the principles underlying the model are strict, and therefore difficult to conform to (Andrich, 2004). A key step is to examine the extent of any misfit, and to investigate possible reasons (Meijer and Tenderio, 2018).

In the Rasch calibrated DKAS, the majority of items showed acceptable fit. Two items showed misfit across more than one sample split, and a further seven items showed misfit according to a one of the three sample splits. Global model fit statistics were poor, indicating that the data did not sufficiently fit the model. In some cases, misfit might be explained by external factors, such as uninterested participants, or environmental factors that affect participant responses (Meijer and Tenderio, 2018), however in the case of the DKAS, misfit was likely due to an element of multidimensionality and response dependence.

8.3.4 Differential item functioning

Four of the nine misfitting items were also flagged in the DIF analyses as being biased items when previous experience of dementia was considered. As discussed above, DIF alone is not a sufficient reason to eliminate items from a test, however, empirically establishing the quality of an item is essential where DIF has been detected (Bond and Fox, 2013). Importantly, DIF was not a feature of the 19 DKAS items that showed promise with regard to scaling properties and reliability.

8.3.5 Overall Rasch contribution to the DKAS

Previous analyses of the DKAS have utilised CTT-based methodology only. This discussion around the first application and calibration of DKAS data to the Rasch model has highlighted the value of using respondents' raw test scores to express their performance on a linear scale continuum that takes varying item difficulty

parameters into account. This is a powerful aspect of the Rasch model that facilitates investigation of the measurement properties of a test (Boone *et al*., 2016). Through Rasch modelling, new insights about dimensionality, scale reliability and targeting of items have emerged.

The following section will confirm that the primary research question of this thesis was addressed, and state how the findings contribute to current knowledge and practice in the field of dementia knowledge measurement.

8.4 Contribution to knowledge and practise

This thesis details the first in-depth examination of dementia knowledge tests under the measurement framework of item response theory. A systematic review of measurement instruments was timely, given the increase in KoD instrument development studies published over the past decade. This review supports and advances the field of dementia knowledge assessment instruments and highlights the importance of standardised psychometric evaluation and reporting procedures.

With regard to the empirical contribution of this thesis; the underlying structure of the KIDE has not been examined in any published study, and the measurement properties of the DKAS had not been examined using IRT modelling. The current study contributes to existing knowledge and advances the understanding of dementia knowledge assessment in healthcare staff and students through item-level analysis of measurement instruments in current use. Through application and calibration of the Rasch model, additional measurement properties of the KIDE and the DKAS have emerged, including findings around targeting, scale reliability, and effective measurement range within the calibrated range of items. These findings, reported transparently alongside the methods used to obtain them, have improved understanding of how the KIDE and the DKAS work in populations of healthcare staff and students. As such, the research question that informed this PhD study has been addressed.

The following section will highlight some directions for future research.

8.5 Directions for future research

Future studies might examine currently available dementia knowledge scales in additional populations of healthcare students and staff, as well as in lay populations. Concurrent application of instruments or sets of items would facilitate a more robust evaluation of the measurement properties of item sets. As recommended in the systematic review chapter of this study, an important aspect of further data collection would be to systematically evaluate psychometric properties and make use of available guidelines on procedures for reporting the measurement properties of instruments, such as those detailed by Mokkink *et al*., (2010) and Rosenkoetter and Tate (2018). Vitoratou and Pickles (2017) argue that scale development and evaluation studies are not widely welcomed by journals, however, to support robust measurement and reliable benchmarking of change, the systematic reporting of such studies in the published literature must be normalised and encouraged.

In addition to further administration of available instruments, additional item content might be developed and evaluated to address the issues of item redundancy and improper targeting that emerged in the KIDE and DKAS scales. As an example, an international Delphi study was conducted by Annear *et al*., (2015b) to inform the development of the DKAS. The methods in this study are transparent, making it achievable to repeat the study and provide an updated bank of dementia knowledge instruments.

Regarding methods used in the development and evaluation of instruments, there are a range of approaches that might be usefully applied to determine additional measurement properties of scales. For example, application of probabilistic non-parametric approaches such as Mokken scale analysis might be used to evaluate measurement properties under less strict theoretical assumptions than those of parametric IRT models.

8.6 Strengths of the current study

The systematic review methodology is considered to be a strength of the current study; monthly citation tracking alerts provide a simple yet effective method for

capturing new KoD instruments and studies that cite existing instruments, therefore might report on addition psychometric evaluation.

Similarly, this study contributes to the psychometric reporting landscape, which is unfortunately limited, as evidenced by the systematic review phase of this study. As noted in the above section, if measurement properties of instruments are reported more frequently in the published literature, this aspect of dementia knowledge measurement studies may gain traction and become normalised.

The methods applied in this study are transparent and can be applied across different domains of health literacy in health-related education and professional development interventions. As such, a strength of this thesis is that the contents are more widely applicable than to dementia knowledge alone.

8.7 Limitations of the study

This study would have benefitted from additional demographic variables in both datasets. Had the same demographic information been available in both datasets, further exploration of how the KIDE performs as a scale, and how measurement properties might differ between groups of healthcare professionals and undergraduate student nurses, would have been facilitated.

Another limitation if this study is that the samples were predominantly female. Although this is reflective of current demographics in undergraduate nurse programmes, the sample was not necessarily reflective of current demographics in healthcare professions more broadly.

Software constraints may also have limited the current study. Despite a proliferation in software capable of data analysis using IRT methods and the Rasch model, the work of this thesis was restricted by software availability. R Studio is a very powerful statistical analysis platform, however advances in the available methods rely heavily on updates from individuals as opposed to being under purely corporate management. Additional software packages such as Winsteps and RUMM2030 might have generated alternative statistics and additional graphics.

8.8 Conclusions

The need for reliable instruments with which to measure knowledge of dementia across population groups is apparent, given the increasing prevalence of dementia worldwide. Robust measurement of dementia knowledge is a realistic and achievable goal. Currently, test development and evaluation procedures in relation to dementia knowledge are anchored by classical test theory methods, with reports of item response theory modelling being absent. The work contained in this thesis details the first in-depth examination of dementia knowledge tests under the measurement framework of item response theory. Discussion of the results has accentuated the value of IRT methods by emphasising the additional measurement properties that can be extracted through Rasch model calibration of item and person parameters along a single latent scale. This exploration of IRT modelling in currently available instruments offers insight into how improvements in instrumentation may be advanced on a practical level.

References

Alacreu, M., Pardo, J., Azorín, M., Climent, M.T., Gasull, V. and Moreno, L. (2019). Importance of increasing modifiable risk factors knowledge on alzheimer's disease among community pharmacists and general practitioners in Spain. *Frontiers in pharmacology*, *10*, p.860.

Alzheimer Scotland. (2020) *Scottish Dementia Research Consortium Annual Report 2019/20.* Available at: Scottish Dementia Research Consortium annual report 2019/20 | Alzheimer Scotland (alzscot.org) [Accessed August 2021].

Alzheimer Scotland. (2021). Online. *About dementia*. Available at: https://www.alzscot.org/what-is-dementia/about-dementia [Accessed July 2021].

Alzheimer's Disease International. (2019a). *World Alzheimer Report 2019: Attitudes to dementia*. London: Alzheimer's Disease International.

Alzheimer's Disease International. (2019b). Online. *About Dementia*. Available at: https://www.alz.co.uk/about-dementia [Accessed July 2021].

Alzheimer's Disease International. (2020). Online. *Types of Dementia*. Available at: https://www.alzint.org/about/dementia-facts-figures/types-of-dementia/ [Accessed August 2021].

Alzheimer's Research UK. (2018). Online. *Types of dementia*. Available at: https://www.alzheimersresearchuk.org/about-dementia/types-of-dementia/vascular-dementia/about/ [Accessed July 2021].

Alzheimer's Research UK. (2021). Online. *Dementia statistics hub: Diagnoses in the UK*. Available at: https://www.dementiastatistics.org/statistics/diagnoses-in-the-uk/ [Accessed July 2021].

Alzheimer's Society. (2016). *Fix Dementia Care: Hospitals*. London: Alzheimer's Society.

Alzheimer's Society. (2021a). Online. *Facts for the media*. Available at: https://www.alzheimers.org.uk/about-us/news-and-media/facts-media [Accessed July 2021].

Alzheimer's Society. (2021b). Online. How dementia progresses. Available at: https://www.alzheimers.org.uk/about-dementia/symptoms-and-diagnosis/how-dementia-progresses [Accessed July 2021].

Alzheimer's Society. (2021c). Online. *Dementia tax*. Available at: https://www.alzheimers.org.uk/about-us/policy-and-influencing/what-we-think/dementia-tax [Accessed July 2021].

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38(1), 123-140. doi: 10.1007/bf02291180.

Andrich, D., (1988). *Rasch models for measurement* (Vol. 68). Sage.

Andrich, D., (2004a). Controversy and the Rasch model: a characteristic of incompatible paradigms?. *Medical care*, pp.I7-I16.

Andrich, D. (2004b) Understanding resistance to the data-model relationship in Rasch's paradigm: a reflection for the next generation. In E. Smith & R. Smith (Eds.) Introduction to Rasch measurement. Theory, models and applications. (pp. 25-47) Maple Grove, MN: JAM Press

Andrich, D. and Marais, I., (2019). *A course in Rasch measurement theory*. D. Andrich y I. Marais (Coords.), Measuring in the Educational, Social and Health Sciences. Springer Nature Singapore Pte Ltd.

Annear, M. J., Toye, C. M., Eccleston, C. E., McInerney, F. J., Elliott, K.-E. J., Tranter, B. K., Robinson, A. L. (2015a). Dementia Knowledge Assessment Scale: Development and Preliminary Psychometric Properties. *Journal Of The American Geriatrics Society*, *63*(11), pp.2375–2381. https://doi.org/10.1111/jgs.13707.

Annear, M. J., Toye, C., McInerney, F., Eccleston, C., Tranter, B., Elliott, K.-E., & Robinson, A. (2015b). What should we know about dementia in the 21st Century? A Delphi consensus study. *BMC Geriatrics*, *15*(1), pp.1.

Annear, M.J., Eccleston, C.E., McInerney, F.J., Elliott, K.E.J., Toye, C.M., Tranter, B.K. and Robinson, A.L., (2016). A new standard in dementia knowledge measurement: Comparative validation of the dementia knowledge assessment scale and the alzheimer's disease knowledge scale. *Journal of the American Geriatrics Society*, *64*(6), pp.1329-1334.

Annear, M.J., Toye, C., Elliott, K.E.J., McInerney, F., Eccleston, C. and Robinson, A., (2017). Dementia knowledge assessment scale (DKAS): confirmatory factor analysis and comparative subscale scores among an international cohort. *BMC geriatrics*, *17*(1), p.168.

Bannigan, K. and Watson, R., 2009. Reliability and validity in a nutshell. *Journal of clinical nursing*, *18*(23), pp.3237-3243.

Barrett, J. J., Haley, W. E., Harrell, L. E., & Powers, R. E. (1997). Knowledge about Alzheimer disease among primary care physicians, psychologists, nurses, and social workers. *Alzheimer Disease & Associated Disorders*, *11*(2), pp.99–106.

Bartholomew, D.J., Knott, M. and Moustaki, I., (2011). *Latent variable models and factor analysis: A unified approach* (Vol. 904). John Wiley & Sons.

Bettens, G. F., Ownsworth, T., Hohaus, L., & McKendry, Y. (2014). Assessing accuracy of knowledge of cognitive effects of normal ageing and mild stage of Alzheimer's disease. *Aging & Mental Health*, *18*(3), pp.296–303. https://doi.org/10.1080/13607863.2013.827629.

Bond, T.G. and Fox, C.M., (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.

Boone, W.J., (2016). Rasch analysis for instrument development: why, when, and how?. *CBE—Life Sciences Education*, *15*(4), p.rm4.

Boone, W.J., Staver, J.R. and Yale, M.S., (2013). *Rasch analysis in the human sciences*. Springer Science & Business Media.

Brown M, Waugh A, Sharp B, Duffy RF, Macrae R. (2018) What are dementia champions and why do we need them? *Dementia*, 17(4):397-400. doi:10.1177/1471301217743413

Brown, K.F., Rumgay, H., Dunlop, C., Ryan, M., Quartly, F., Cox, A., Deas, A., Elliss-Brookes, L., Gavin, A., Hounsome, L. and Huws, D., (2018). The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. *British journal of cancer*, *118*(8), pp.1130-1141.

Cahill, S., Pierce, M., Werner, P., Darley, A., & Bobersky, A. (2015). A systematic review of the public's knowledge and understanding of Alzheimer's disease and dementia. *Alzheimer Disease & Associated Disorders*, *29*(3), 255–275 21p. https://doi.org/10.1097/WAD.0000000000000102

Carpenter, B. D., Balsis, S., Otilingam, P. G., Hanson, P. K., & Gatz, M. (2009). The Alzheimer's Disease Knowledge Scale: Development and Psychometric Properties. *Gerontologist*, *49*(2), pp.236–247. https://doi.org/10.1093/geront/gnp023.

Cherry,K.E., Robin L. West, Celinda M. Reese, Michael P. Santa Maria, Monica Yassuda, K., (2000). The knowledge of memory aging questionnaire. *Educational Gerontology*, *26*(3), pp.195-219.

Clari, M., Matarese, M., Alvaro, R., Piredda, M. and De Marinis, M.G., (2016). Measurement properties of instruments evaluating self-care and related concepts in people with chronic obstructive pulmonary disease: a systematic review. *Heart & Lung: The Journal of Acute and Critical Care*, *45*(5), pp.441-448.

Curyto, K. J., & Vriesman, D. K. (2016). Development of the Knowledge of Dementia Competencies Self-Assessment Tool. *American Journal of Alzheimer's Disease and Other Dementias*, *31*(1), pp.18–26. https://doi.org/10.1177/1533317515581703.

Davies, C.J., Waters, D. and Marshall, A., (2017). A systematic review of the psychometric properties of bronchiolitis assessment tools. *Journal of advanced nursing*, *73*(2), pp.286-301.

De Champlain, A.F., (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical education*, *44*(1), pp.109-117.

DeMars, C., (2010). *Item response theory*. Oxford University Press.

DeMars, C.E., (2018). Classical Test Theory and Item Response Theory. *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development*, pp.49-73.

Dementia UK. (2021). Online. *Alcohol related brain damage.* Available at: Alcohol related brain damage - Dementia UK [Accessed July 2021].

Department of Health. (2015). *Prime Ministers challenge on dementia*. London: Department of Health.

Desjardins, C.D. and Bulut, O., (2018). *Handbook of educational measurement and psychometrics using R*. CRC Press.

DeVellis, R.F., (2016). *Scale development: Theory and applications* (Vol. 26). London: Sage publications.

Devine, P. (2016). *Public Attitudes and Knowledge of Dementia: Northern Ireland, Republic of Ireland & Scotland*. Belfast: ARK Ageing Programme.

Dewing, J. and Dijk, S., (2016). What is the current state of care for older people with dementia in general hospitals? A literature review. *Dementia*, *15*(1), pp.106-124.

Dieckmann, L., Zarit, S. H., Zarit, J. M., & Gatz, M. (1988). The Alzheimer's disease knowledge test. *The Gerontologist*, *28*(3), pp.402–408.

Dunn, T.J., Baguley, T. and Brunsden, V., (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British journal of psychology*, *105*(3), pp.399-412.

Eccleston, C., Doherty, K., Bindoff, A., Robinson, A., Vickers, J. and McInerney, F., (2019). Building dementia knowledge globally through the understanding dementia Massive Open Online Course (MOOC). *npj Science of Learning*, *4*(1), pp.1-6.

Eccleston, CE, Courtney-Pratt, H, McInerney, F, Johnstone, A, Doherty, K. (2021). Predictors of dementia knowledge in a rural general public sample. *Aust J Rural Health* 29: 530– 537. https://doi.org/10.1111/ajr.12777

Elvish, R., Burrow, S., Cawley, R., Harney, K., Graham, P., Pilling, M., Keady, J. (2014). "Getting to Know Me": the development and evaluation of a training programme for enhancing skills in the care of people with dementia in general hospital settings. *Aging & Mental Health*, *18*(4), pp.481–488. https://doi.org/10.1080/13607863.2013.856860.

Embretson, S.E. and Reise, S.P., (2013). *Item response theory*. Psychology Press.

Finch, W.H. and French, B.F., (2015). *Latent variable modeling with R*. Routledge.

Gilleard, C., & Groom, F. (1994). A study of 2 dementia quizzes. *British Journal of Clinicial Psychology*, *33*(4), pp.529–534.

Gomes, D.E., Guedes dos Santos, J.L., Borges, P., Wicto, J., Pedroso Alves, M., de Andrade, D.F. and Erdmann, A.L., (2018). Theory of the response to the item in research in public health. *Journal of Nursing UFPE/Revista de Enfermagem UFPE*, *12*(6).

Griffith, L., van den Heuvel, E., Fortier, I., Hofer, S., Raina, P., Sohel, N., Payette, H., Wolfson, C. and Belleville, S., (2013). *Harmonization of cognitive measures in individual participant data and aggregate data meta-analysis.* In: Harmonization of Cognitive Measures in Individual Participant Data and Aggregate Data Meta-Analysis. Agency for Healthcare Research and Quality (US), Rockville (MD); 2013. Available from https://www.ncbi.nlm.nih.gov/books/NBK132553 PMID: 23617017.

Guyatt, G.H., Feeny, D.H. and Patrick, D.L., (1993). Measuring health-related quality of life. *Annals of internal medicine*, *118*(8), pp.622-629.

Hulin, C.L., Drasgow, F. and Parsons, C.K., (1983). *Item response theory: Application to psychological measurement*. Dorsey Press.

Irwing, P., Booth, T. and Hughes, D.J. eds., (2018). *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development.* John Wiley & Sons.

Jack-Waugh, A., Macrae, R. and Ritchie, L., (2017), November. Scotland's national dementia champions programme: six years on. In *UK Dementia Congress 2017: Caring Times*.

Jack-Waugh, A., Ritchie, L. and Macrae, R., (2018). Assessing the educational impact of the dementia champions programme in Scotland: Implications for evaluating professional dementia education. *Nurse education today*, *71*, pp.205-210.

Janssens, A. C. J. W., & Gwinn, M. (2015). Novel citation-based search method for scientific literature: application to meta-analyses. *BMC Medical Research Methodology*, *15*(1), pp.1–11. https://doi.org/10.1186/s12874-015-0077-z.

Koller, I., & Hatzinger, R. (2013). Nonparametric tests for the Rasch model: Explanation, development, and application of quasi-exact tests for small

samples. *Interstat, 11*, 1-16.http://interstat.statjournals.net/YEAR/2013/abstracts/1311002.php

Kuhn, D., King, S. P., & Fulton, B. R. (2005). Development of the Knowledge about Memory Loss and Care (KAML-C) test. *American Journal of Alzheimer's Disease and Other Dementias*, *20*(1), pp.41–49.

Kuper, H., Nicholson, A., & Hemingway, H. (2006). Searching for observational studies: what does citation tracking add to PubMed? A case study in depression and coronary heart disease. *BMC Medical Research Methodology*, *6*(1), pp.1–4. https://doi.org/10.1186/1471-2288-6-4.

Leung, K., Trevena, L., & Waters, D. (2014). Systematic review of instruments for measuring nurses' knowledge, skills and attitudes for evidence-based practice. *Journal of Advanced Nursing*, *70*(10), pp.2181–2195.

Lewis, F., Karlsberg Schaffer, S., Sussex, J., O'Neill, P., and Cockcroft, L. (2014). Online. *The trajectory of dementia in the UK – making a difference. Report for Alzheimer's Research UK by Office of Health Economics Consulting*. Available at: The Trajectory of Dementia in the UK - Making a Difference (ohe.org) [Accessed August 2021].

Liu, W., Kim, S. and Alessio, H., (2020). Mealtime Caregiving Knowledge, Attitudes, and Behaviors for Persons Living with Dementia: A Systematic Review of Psychometric Properties of Instruments: Instruments of dementia mealtime caregiving attributes. *International Journal of Nursing Studies*, p.103824.

Long, C. O., Sowell, E. J., Hess, R. K., & Alonzo, T. R. (2012). Development of the Questionnaire on Palliative Care for Advanced Dementia (qPAD). *American Journal Of Alzheimers Disease And Other Dementias*, *27*(7), pp.537–543. https://doi.org/10.1177/1533317512459793.

Lord, F.M. and Novick, M.R., (2008). *Statistical theories of mental test scores*. IAP.

Lorio, A.K., Gore, J.B., Warthen, L., Housley, S.N. and Burgess, E.O., (2017). Teaching dementia care to physical therapy doctoral students: A multimodal experiential learning approach. *Gerontology & geriatrics education*, *38*(3), pp.313-324.

Lovie, P. and Lovie, A.D., (1996). Charles Edward Spearman, FRS (1863-1945). *Notes Rec. R. Soc. Lond*. 50 (1), pp.75-88

Macrae, R., Jack-Waugh, A., Mellon, K., and Gorrie, L. (2019). *'Scotland's National Dementia Champions evidence, actions and outcomes*', Second National NHS Education for Scotland Nursing, Midwifery and Allied Health Professions (NMAHP) Education Conference, Edinburgh, 9/05/19.

Magis, D., Yan, D. and Von Davier, A.A., (2017). *Computerized adaptive and multistage testing with R: Using packages catr and mstr*. Springer.

Mair, P., (2018). *Modern psychometrics with R*. Cham: Springer International Publishing.

Marais, I. and Andrich, D., (2008). Effects of Varying Magnitude and Patterns of Response Dependence. *Journal of applied measurement*, *9*(2), pp.105-124.

Marcinkiewicz, A., Montagu, I., Waterton, J. and Reid, S., (2016). Online. *Scottish Social Attitudes 2015: attitudes to government, the National Health Service, the economy and standard of living*. Available at: Scottish Social Attitudes| ScotCen Social Research (natcen.ac.uk) [Accessed June 2021].

McDonald, R.P., (2013). *Test theory: A unified treatment*. Psychology Press.

McDonald, R.P., (2014). *Factor analysis and related methods*. Psychology Press.

McGrory, S., Doherty, J.M., Austin, E.J., Starr, J.M. and Shenkin, S.D., (2014). Item response theory analysis of cognitive tests in people with dementia: a systematic review. *BMC psychiatry*, *14*(1), pp.1-15.

Mid Staffordshire NHS Foundation Trust Public Inquiry. (2013). Online. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: Executive summary.* Available at: Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry - GOV.UK (www.gov.uk) [Accessed May 2021].

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M. and De Vet, H.C., (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research*, *19*(4), pp.539-549.

National Records for Scotland. (2021). Online. *Leading causes of death in Scotland*. Available at: https://www.nrscotland.gov.uk/statistics-and-data/statistics/scotlands-facts/leading-causes-of-death-in-scotland [Accessed July 2021].

Nichols, E., Szoeke, C.E., Vollset, S.E., Abbasi, N., Abd-Allah, F., Abdela, J., Aichour, M.T.E., Akinyemi, R.O., Alahdab, F., Asgedom, S.W. and Awasthi, A., (2019). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, *18*(1), pp.88-106.

Oxford University Press. (2021 Online. *Oxford living dictionaries: definition of 'knowledge'*. Available at: https://en.oxforddictionaries.com/definition/knowledge [Accessed June 2021].

Parial, L.L., Lam, S.C., Ho, J.Y.S., Suen, L.K. and Leung, A.Y.M., (2021). Public knowledge of the influence of modifiable cardiovascular risk factors on dementia: a systematic literature review and meta-analysis. *Aging & Mental Health*, *25*(8), pp.1395-1409.

Pentzek, M., Abholz, H.-H., Ostapczuk, M., Altiner, A., Wollny, A., & Fuchs, A. (2009). Dementia knowledge among general practitioners: first results and psychometric properties of a new instrument. *International Psychogeriatrics*, *21*(6), pp.1105–1115. https://doi.org/10.1017/S1041610209990500.

Polit, D.F. and Yang, F., (2016). *Measurement and the measurement of change: a primer for the health professions*. Philadelphia, PA: Wolters Kluwer.

Ponocny, I., (2001). Nonparametric goodness-of-fit tests for the Rasch model. *Psychometrika*, *66*(3), pp.437-459.

Prince, M., Comas-Herrera, A., Knapp, M., Guerchet, M., and Karagiannidou, M. (2016). World Alzheimer Report 2016. *Improving healthcare for people living with dementia. Coverage, quality and costs now and in the future*. Alzheimer's Disease International: London.

Reise, S.P. and Revicki, D.A. eds., (2014). *Handbook of item response theory modeling: Applications to typical performance assessment*. Routledge.

Revelle, W. and Zinbarg, R.E., (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), p.145.

Rosenkoetter, U. and Tate, R.L., (2018). Assessing features of psychometric assessment instruments: A comparison of the COSMIN checklist with other critical appraisal tools. *Brain Impairment*, *19*(1), pp.103-118.

Rust, J. and Golombok, S., (2009). *Modern psychometrics: The science of psychological assessment*. Routledge.

Scerri, A., Innes, A. and Scerri, C., (2020). Person-centered dementia care in acute hospital wards—The influence of staff knowledge and attitudes. *Geriatric Nursing*, *41*(3), pp.215-221.

Schmitt, N., (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, *8*(4), p.350.

Scottish Government (2017). *National Dementia Strategy: 2017-2020*. Edinburgh: Scottish Government

Scottish Government. (2010). *Scotland's National Dementia Strategy: 2010-2013*. Edinburgh: Scottish Government.

Shanahan, N., Orrell, M., Schepers, A. K., & Spector, A. (2013). The development and evaluation of the DK-20: a knowledge of dementia measure. *International Psychogeriatrics*, *25*(11), 1899–1907. https://doi.org/10.1017/S1041610213001142

Spearman, C., (1987). The proof and measurement of association between two things. *The American journal of psychology*, *100*(3/4), pp.441-471.

Spector, A., Orrell, M., Schepers, A., & Shanahan, N. (2012). A systematic review of "knowledge of dementia" outcome measures. *Ageing Research Reviews*, *11*(1), pp.67–77. https://doi.org/http://dx.doi.org/10.1016/j.arr.2011.09.002.

Streiner, D.L., Norman, G.R. and Cairney, J., (2015). *Health measurement scales: a practical guide to their development and use*. Oxford University Press, USA.

Sullivan, K. A., & Mullan, M. A. (2016). Comparison of the psychometric properties of four dementia knowledge measures: Which test should be used with dementia care staff? *Australasian Journal on Ageing*, n/a–n/a. https://doi.org/10.1111/ajag.12299

Sunderland, M., Batterham, P., Calear, A., Carragher, N., Baillie, A. and Slade, T., (2018). High agreement was obtained across scores from multiple equated scales for social anxiety disorder using item response theory. *Journal of clinical epidemiology*, *99*, pp.132-143.

Surr, C., Baillie, L., Jack-Waugh, A. and Brown, M., (2017). Position paper: The importance of including dementia in pre and post-qualifying curricula for health and social care professionals.

Terwee, C.B., Bot, S.D., de Boer, M.R., van der Windt, D.A., Knol, D.L., Dekker, J., Bouter, L.M. and de Vet, H.C., (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*, *60*(1), pp.34-42.

Thissen, D.E. and Wainer, H.E., (2001). *Test scoring*. Lawrence Erlbaum Associates Publishers.

Traub, R.E., (1997). Classical test theory in historical perspective. *Educational Measurement, 16*, pp.8-13.

Vitoratou, S. and Pickles, A. (2017). A note on contemporary psychometrics, Journal of Mental Health, 26:6, 486-488, DOI: 10.1080/09638237.2017.1392008

Wiese, L.K., (2017). A Novel Measure to Assess Basic Knowledge of Alzheimer's Disease in Underserved Populations. *Journal of the American Geriatrics Society*, 65, pp.S297-S298.

World Health Organisation. (2017). *Global action plan on the public health response to dementia 2017–2025*. Geneva: World Health Organization.

World Health Organisation. (2020). Online. *Dementia: Key facts*. Available at: https://www.who.int/news-room/fact-sheets/detail/dementia [Accessed September 2020].

Wright, B.D. and Mok, M., (2000). Understanding Rasch measurement: Rasch models overview. *Journal of applied measurement*.

Wright, K., Golder, S., & Rodriguez-Lopez, R. (2014). Citation searching: a systematic review case study of multiple risk behaviour interventions. *BMC Med Res Methodol, 14*. https://doi.org/10.1186/1471-2288-14-73.

Yang, L., Liao, L., Lam, F., He, C., & Pang, M. (2015). Psychometric properties of dual-task balance assessments for older adults: A systematic review. *Maturitas*, *80*(4), pp.359–369.

Zinbarg, R.E., Yovel, I., Revelle, W. and McDonald, R.P., (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ωh. *Applied Psychological Measurement*, *30*(2), pp.121-144.

Appendix 1: *CONFIDENTIAL* Memorandum of Understanding

Memorandum of Understanding

between

University of the West of Scotland

and

University of Dundee

THIS AGREEMENT is dated 14 February 2019 ("Effective Date")

PARTIES

(1) **UNIVERSITY OF WEST OF SCOTLAND** whose registered office is at Paisley PA1 2BE (**UWS**); and

(2) **University of Dundee** established by Royal Charter dated 20 July 1967 and a registered Scottish Charity (no. SC015096) having its principal office at 149 Nethergate, Dundee DD1 4HN (" **Dundee"**).

Background

(A) This Memorandum of Understanding confirms the intention to establish a co-operative relationship between UWS and Dundee which will be to the benefit of both organisations. This working relationship reflects the interests of both organisations to engage in collaborative research on approaches measurement/assessment of Knowledge of Dementia (KoD) from the perspective of contemporary applied psychometrics, using mixed methods combining expert judgement, systematic and psychometric review, application of Test Theory and item-set development.

(B) The parties agree and acknowledge that this Memorandum of Understanding is intended to demonstrate the intention of both partners to strengthen and develop links between UWS and Dundee and, except for the provisions of clause 6-10 , shall have no legal effect. It is the intention of the parties that following and subject to further discussion and mutual agreement to proceed these points of principle be recorded in a legally binding agreement.

Memorandum of Understanding ("MoU")

1 The specific objectives and activities of the co-operative relationship to be established by this MoU between the UWS and Dundee will be as follows:

f) Create an agreement for sharing previous Dementia Champions data collected using the Knowledge in Dementia Scale (Elvish et al, 2016) with the Dundee team to be utilised in analysis of empirical data sets and updating of items (questions).
g) Subject to an agreement, to share results of the above activities with UWS to inform the evaluation methods for ASCPP education activities, starting with Champions Cohort 10 in March 2019.
h) For Dundee to conduct further analysis on Cohort 10 data in line with validation.
i) Generate shared narrative and plan future activities to be coordinated in relation to Dundee thesis submission and REF 2021.
j) Plan for joint dissemination activities including, local seminars/training for UWS staff on item level analyses, academic publications and conference presentations.

2 The UWS link person will be Dr Rhoda Macrae and the link person for Dundee will be Clair Gamble. They will be responsible for ensuring communication links are effective and that activities are progressing to the satisfaction of both parties.

3 This Memorandum of Understanding comes into effect from the Effective Date and will remain in place for a period of 36 months (unless otherwise stated) or until superseded by a formal partnership agreement. As noted in the Background above, no formal partnership other than as set out in the MoU should commence without a full partnership agreement being concluded.

4 Implementation of the provisions of this MoU shall be the subject of further communication between both parties.

5 This MoU does not preclude either party entering in to similar joint arrangements with other parties in the UK or overseas.

6 All announcements regarding this MoU, the relationship established herein, and/or release of any information pertaining hereto shall require the mutual consent of both parties. Neither party shall use the logo or branding of the other party on any documents, press releases or promotional materials without the prior consent of the other party.

7. Under this MoU the Parties may share anonymised statistical data with each other to plan and target activities appropriately. At times this information will be information that has not yet been made public and/or is confidential. The Parties must only use such information (whether confidential or not) for the purposes of fulfilling their obligations under this MoU. The Parties must not disclose confidential information to any third parties without prior consent nor use this for commercial advantage or to disadvantage or discredit the other party to the MoU or anyone else. Both Parties agree that under this MoU they will not share any data which would allow any individuals to be identified.

8. Neither party has the right, power or authority to bind the other in any manner unless authorised in writing by the other party in a specific instance.

9. If either of the Parties wishes to terminate this MoU, a minimum of three months' notice must be given in writing to the other Party.

10. This MoU and its terms shall be governed by and construed in accordance with the law of Scotland.

Subscribed on behalf of University of West of Scotland	Subscribed on behalf of University of Dundee
	
Signature	Signature 

Name: Julie Edgar
Date: 30/01/2019

Name: Clair Gamble
Date: 14/02/19

Appendix 2: Dementia Champions pre-programme questionnaire

DEMENTIA CHAMPIONS

Pre-Programme Questionnaire

In order to evaluate the Dementia Champions programme and ensure that it is providing you with the opportunity to develop as change agents, we would be grateful if you would complete this questionnaire prior to Study Day 1. Please bring the completed form with you.

If you are unable to complete the form prior to the first day there will be copies available on day 1 and we will ask for them to be completed before you attend any of the sessions.

The information derived from this questionnaire will provide us with a baseline measure. You will be asked to answer the same questions again on the Study Day 5.

Your response will only been seen by the research team and will be anonymized for reporting purposes.

In this questionnaire we are asking for the last three digits of your home postcode your initials and NHS Board/SSSC Area so that we can link your responses to those you give on the final study day. Any additional surveys or questionnaires will be anonymous.

Last three digits of your home postcode:

.....
.....

NHS Board/SSSC Area:

.....
.....

Your Initials

……
……

A. Please indicate to what extent you agree or disagree with each of the following statements: *Please tick:*

ADQ item content	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
1. It is important to have a very strict routine when working with people with dementia					
2. People with dementia are very much like children					
3. There is no hope for people with dementia					
4. People with dementia are unable to make decisions for themselves					
5. It is important for people with dementia to have stimulating and enjoyable activities to occupy their time					
6. People with dementia are sick and need to be looked after					
7. It is important for people with dementia to be given as much choice as possible in their daily lives					
8. Nothing can be done for people with dementia, except for keeping them clean and comfortable					
9. People with dementia are more likely to be contented when treated with understanding and reassurance.					
10. Once dementia develops in a person, it is inevitable that they will go down hill					
11. People with dementia need to feel respected, just like anybody else					
12. Good dementia care involves caring for a person's psychological needs as well as their physical needs					
13. It is important not to become too attached to people with dementia					

14. It doesn't matter what you say to people with dementia because they forget it anyway					
15. People with dementia often have good reasons for behaving as they do					
16. Spending time with people with dementia can be very enjoyable					
17. It is important to respond to people with dementia with empathy and understanding					
18. There are a lot of things that people with dementia can do					
19. People with dementia are just ordinary people who need special understanding to fulfil their needs					

B. Please indicate whether you agree, disagree or don't know. *Please tick*

KIDE		**Agree**	**Disagree**	**Don't know**
1	Permanent changes to the brain occur in most types of dementia			
2	People who have dementia will usually show the same symptoms			
3	Dementia can be caused by a number of small strokes			
4	Currently, most types of dementia cannot be cured			
5	When people with dementia walk around it is usually aimless			
6	People with dementia will eventually lose all their ability to communicate			

7	People with dementia who are verbally aggressive nearly always become physically aggressive			
8	Brain damage is the only factor that is responsible for the way people with dementia behave			
9	It is possible to catch dementia from other people			
10	My perception of reality may be different from that of a person with dementia			
11	People with dementia never get depressed			
12	Anger and hostility occur in dementia mostly because the “aggression” part of the brain has been affected			
13	Dementia is a general term which refers to a number of different diseases			
14	A person with dementia’s history and background play a significant part in their behaviour			
15	Physical pain may result in a person with dementia becoming aggressive or withdrawn			
16	A person with dementia is less likely to receive pain relief than a person without dementia when they are in hospital			

C. Please indicate whether you think these statements are true or false.

Please tick

DKAS	Item Content	True	False
1	Most forms of dementia do not generally shorten a person's life		
2	Blood vessel disease (vascular dementia) is the most common form of dementia		
3	People can recover from the most common forms of dementia		
4	Dementia is a normal part of the ageing process		
5	Dementia does not result from physical changes in the brain		
6	Planning for end of life care is generally not necessary following a diagnosis of dementia		
7	AD is the most common form of dementia		
8	It is impossible to communicate with a person who has advanced dementia		
9	A person experiencing advanced dementia will not generally react to changes in their physical environment		
10	It is important to correct a person with dementia when they are confused		
11	People experiencing advanced dementia often communicate through body language		
12	Uncharacteristic behaviours in a person experiencing dementia are usually a response to unmet needs		
13	Medications are the most effective way of treating behavioural symptoms of dementia		
14	People experiencing dementia do not generally have problems making decisions		
15	Movement is generally affected in the later stages of dementia		
16	Difficulty eating and drinking generally occurs in the later stages of dementia		

17	People with advanced dementia may have difficulty speaking		
18	People experiencing dementia often have difficulty learning new skills		
19	Daily care for a person with advanced dementia is most effective when it focuses on providing comfort		
20	Having high blood pressure increases a person's risk of developing dementia		
21	Maintaining a healthy lifestyle does not reduce the risk of developing the most common forms of dementia		
22	Symptoms of depression can be mistaken for symptoms of dementia		
23	The sudden onset of cognitive problems is characteristic of common forms of dementia		
24	Exercise is generally beneficial for people with dementia		
25	Early diagnosis of dementia does not generally improve quality of life for people experiencing the condition		

D. The following statements list a number of criteria that a Dementia Champion will be expected to fulfil. Please rate your degree of confidence just now by recording a number from 0 to 100 using the scale given below:

0 10 20 30 40 50 60 70 80 90 100
Cannot do at all — Moderately certain can do — Highly certain can do

Please put the number responding to your level of confidence to the right of each statement:	**Confidence (0-100)**
a) Recognise and respond to the impact of the physical, emotional, social, cultural and spiritual environment on the maintenance of rights, choice, identity, dignity and equity for the person with dementia, in an acute hospital setting, intermediate, anticipatory and other community teams.	

b) Respond with evidence based best practice, to the physical and mental health issues that may affect the individual course of a person's journey before, during and after receiving care in the acute hospital environment.

c) Recognise and deal with the complexities associated with dementia in the acute hospital setting, and other physical health care and community settings that may have legal and ethical implications and act to safeguard the best interests of people with dementia, families and carers.

d) Apply and evaluate a range of interventions to reduce stress and distress and promote functional capacity and promote ability, strengths and quality of life for the person with dementia, paying particular attention to demonstrating kindness, empathy, enablement, partnership working and compassion.

e) Implement leadership and change agent skills and knowledge to enhance and improve the care of the person with dementia in every area of their influence, utilising existing and developing quality improvement systems, sharing good practice forums and knowledge networks

Appendix 3: UoD ethical approval letter

University of Dundee

University of Dundee Schools of Nursing & Health Sciences and Dentistry Research Ethics Committee (SREC)

University of Dundee

Dundee

DD1 4HN

23 September 2019

Dear Clair,

Application Number: UoD\SNHS\RPG\2019023

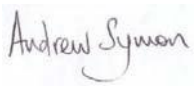
Title of Project: Assessing what we know about dementia: how can we design more useful tests?

I am writing to advise you that your ethics application has been reviewed and approved on behalf of the University of Dundee Schools of Nursing & Health Sciences and Dentistry Research Ethics Committee (SREC).

Any changes to the approved documentation (e.g., study protocol, information sheet, consent form) must be approved by this SREC before the changes are implemented. Requests for amendments should be requested using the Post-Approval Request for an Amendment form.

Approval is valid for the duration of the project, as stated in the original application. Should you wish your study to continue beyond the stated project end date, you must request an extension to this approval using the Post-Approval Request for an Extension form. The extension request must be lodged during your period of study and the period requested must not extend beyond the deadline for submission of your research project.

Yours sincerely



Dr Andrew Symon

Convener, Schools of Nursing & Health Sciences
and Dentistry Research Ethics Committee

Appendix 4: Participant information sheet

Participant information sheet

Version 2.0, 23th September 2019

Research project

Assessing what we know about dementia: how can we design more useful tests?

You have been given this information sheet because you are being invited to take part in a research study. This information sheet describes the study and explains what will be involved if you decide to take part.

What is the purpose of this study?

In this study you will be asked to complete a survey containing questions about dementia.
I will use your results to examine how educational tests are developed and how they can be made more efficient. I will achieve this by fitting statistical models to the data I receive from this survey and examining the results.
Your answers will be anonymised, and the results will not be shared with any of your teaching or admin staff.

Who is conducting the study?

My name is Clair Gamble. I am a Registered General Nurse and a PhD researcher in the School of Nursing and Health Sciences (SNHS) at the University of Dundee.

What will participating in this project involve?

If you agree to participate in the project, you will be invited to complete a survey relating to dementia. The survey should take 10- 20 minutes to complete, but you can take as much time as you need. There are no 'trick' questions, and I do not expect you to be able to answer all questions correctly. There will be no follow up to this study, so I will not contact you again to ask for further participation.

As a thank you for participating, you can be entered into a draw to win one of four £25 Amazon vouchers. If you would like to enter the draw, please provide your email address in the relevant section on the consent form. These will be stored separately from the surveys, so your answers will remain anonymous.

Do I have to take part?

No, it's completely up to you whether or not you take part in the study. If you agree to take part, you are free to change your mind at any time without giving any reason.
Choosing to either take part or not take part in this study will have no impact on your grades, assessments or future studies.

Are there any risks?

This survey asks questions about dementia which may cause emotional discomfort, for example, if you have experienced a loss caused by dementia. I remind you that participation is entirely voluntary.

What are the possible benefits of taking part?

There will be no immediate benefits for you, but by taking in part in this study you can help us better understand how we can improve methods to assess knowledge of dementia. For example, this research may inform educational testing in health-related undergraduate degrees, or social attitudes surveys that examine any shifting trends in dementia knowledge in the UK.

Will the results of this study be confidential?

All information collected in this study will be kept strictly confidential. The consent form with your initials on it will be immediately detached and kept separately from the rest of the survey.
All electronic data will be stored on a password protected computer. All paper copies will be kept in a locked filing cabinet within a locked office on campus. All data collected for this study will be retained, in a secure location, for a period of 10 years, as per data protection laws.

What will happen to the results of the research study?

The (anonymous) results of this study will be used to inform my PhD thesis. They will also be used in academic papers for publication, presentations, and in future teaching of research methods. I would be happy to send you a summary of the results if you wish.

Who has reviewed the study?

This research has been approved by the School Research Ethics Committee, SNHS, University of Dundee.

Contact for Further Information

I am the principal investigator and main contact for the study. Please don't hesitate to get in touch if you have any questions about the project or your participation.
My contact details are: Clair M. Gamble
Email: c.z.gamble@dundee.ac.uk
School of Nursing and Health Sciences (SNHS), 11 Airlie Place, University of Dundee (UoD), Dundee, DD1 4HJ.

If you have any questions, concerns, or complaints that you wish to address to someone other than the investigator, you may contact Tim Croudace, Professor of Applied Health Research, SNHS (t.j.croudace@dundee.ac.uk). For advice or quesries on data management please contact dataprotection@dundee.ac.uk.

Project team

Clair Gamble, SNHS, UoD., Prof. Tim Croudace, SNHS, UoD., Prof. Judith Sixsmith, SNHS, UoD., Prof. Wendy Moncur, SNHS, UoD.

Thank you for considering taking part in this study and taking the time to read this information. If you are willing to complete the survey for this research project, please complete the consent form on the next page.

Appendix 5: Consent form

CONSENT FORM

Project: Assessing what we know about dementia: how can we design more useful tests?

Researcher: Clair M Gamble, School of Nursing and Health Sciences, University of Dundee. Email: czgamble@dundee.ac.uk

Please initial box

1. I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving reason. ☐
3. I agree to take part in the above study. ☐
4. I agree that an anonymised data set, gathered for this study may be stored in a specialist data centre/repository relevant to this subject area for future research. ☐

Name of Participant	Date	Signature

Name of Researcher	Date	Signature

I wish to be entered into the prize draw for one of 4 Amazon vouchers. ☐

If yes, please enter your email address below (these will be kept separately from the survey answers)

Appendix 6: 41-item Dementia knowledge survey

University of Dundee

Dementia Knowledge Survey

Project:

Assessing what we know about dementia: how can we design more useful tests?

Researcher:

Clair M Gamble,

School of Nursing and Health Sciences, University of Dundee. Email: czgamble@dundee.ac.uk

Thank you for agreeing to complete this survey.

It should only take around 10 minutes to complete.

You will be shown 41 statements. You will be asked if you think each is True or False

(Please circle one option T for True, F for False, then move on to the next statement)

T F

When you turn over to the back page, you will be asked about your experience with dementia and also 3 questions about your demographic characteristics.

Please be sure to respond to all questions, even if some seem quite similar.

Dementia knowledge survey	True	False
Permanent changes to the brain occur in most types of dementia	T	F
People who have dementia will usually show the same symptoms	T	F
Dementia can be caused by a number of small strokes	T	F
Currently, most types of dementia cannot be cured	T	F
When people with dementia walk around it is usually aimless	T	F
People with dementia will eventually lose all their ability to communicate	T	F
People with dementia who are verbally aggressive nearly always become physically aggressive	T	F
Brain damage is the only factor that is responsible for the way people with dementia behave	T	F
It is possible to catch dementia from other people	T	F
My perception of reality may be different from that of a person with dementia	T	F
People with dementia never get depressed	T	F
Anger and hostility occur in dementia mostly because the "aggression" part of the brain has been affected	T	F
Dementia is a general term which refers to a number of different diseases	T	F
A person with dementia's history and background play a significant part in their behaviour	T	F
Physical pain may result in a person with dementia becoming aggressive or withdrawn	T	F
A person with dementia is less likely to receive pain relief than a person without dementia when they are in hospital	T	F

Dementia knowledge survey	True	False
Most forms of dementia do not generally shorten a person's life	T	F
Blood vessel disease (vascular dementia) is the most common form of dementia	T	F
People can recover from the most common forms of dementia	T	F
Dementia is a normal part of the ageing process	T	F
Dementia does not result from physical changes in the brain	T	F
Planning for end of life care is generally not necessary following a diagnosis of dementia	T	F
Alzheimer's Disease is the most common form of dementia	T	F
It is impossible to communicate with a person who has advanced dementia	T	F
A person experiencing advanced dementia will not generally react to changes in their physical environment.	T	F
It is important to correct a person with dementia when they are confused	T	F
People experiencing advanced dementia often communicate through body language	T	F
Uncharacteristic behaviours in a person experiencing dementia are usually a response to unmet needs	T	F
Medications are the most effective way of treating behavioural symptoms of dementia	T	F
People experiencing dementia do not generally have problems making decisions	T	F
Movement is generally affected in the later stages of dementia	T	F
Difficulty eating and drinking generally occurs in the later stages of dementia	T	F
People with advanced dementia may have difficulty speaking	T	F

Dementia knowledge survey	True	False
People experiencing dementia often have difficulty learning new skills	T	F
Daily care for a person with advanced dementia is most effective when it focuses on providing comfort	T	F
Having high blood pressure increases a person's risk of developing dementia	T	F
Maintaining a healthy lifestyle does not reduce the risk of developing the most common forms of dementia	T	F
Symptoms of depression can be mistaken for symptoms of dementia	T	F
The sudden onset of cognitive problems is characteristic of common forms of dementia	T	F
Exercise is generally beneficial for people with dementia	T	F
Early diagnosis of dementia does not generally improve quality of life for people experiencing the condition	T	F

Do you know (or have you known) someone who lives with dementia?

(Yes or No)

(Please circle one option)

Yes	No	Prefer not to say

Have you ever worked with people who live with dementia?

(Yes or No)

(Please circle one option)

Yes	No	Prefer not to say

Demographic section *(Experience, Age and Gender)*

What is your highest qualification?
(Please circle one option)

High school or equivalent qualifications	College-level qualifications	Bachelor's degree	Master's degree or beyond	Prefer not to say

What is your age?
(Please circle one option)

Under 18	18-24	25-34	35-44	45-54	Above 54	Prefer not to say

Regarding gender, do you…
(Please circle one option)

Identify as female	Identify as male	Identify in another way	Prefer not to say

Appendix 7: Exploratory factor analysis of the DKAS instrument in a cohort of first-year student nurses.

This appendix contains a preliminary solution for an exploratory factor analysis of the DKAS in undergraduate nursing students. This is not a complete record of the results; however, it has been included as it may be useful to the reader.

Introduction

Factor analysis techniques are widely used to explore potential underlying structures in a set of inter-related variables (Flora & Flake, 2017). In order to be useful in establishing baseline knowledge and changes or trends in knowledge levels, measurement instruments must be established as valid and reliable. Construct validity is an aspect of test validity in which latent variables (or factors) are identified through examination of the correlations between variables in a dataset; such factors (if present) generally form subscales of the instrument (Streiner & Norman, 2008). Hypotheses about the underlying structure of a measurement instrument can be generated using exploratory factor analysis (EFA) techniques, confirmatory factor analysis (CFA) can then be used to test and confirm hypothesised models (Mair, 2018). As reported in the literature review in *Chapter 3*, EFA and CFA techniques have not been widely implemented in the field of dementia knowledge measurement scales.

The Dementia Knowledge Assessment Scale (DKAS) (Annear *et al*., 2015b) is one of the more well-established knowledge of dementia (KoD) instruments in current use. This appendix reports an abridged set of factor analysis results to determine structural validity of the DKAS in a sample of undergraduate nursing students.

Background to the dataset

The DKAS was administered alongside the Knowledge In Dementia scale (KIDE) (Elvish *et al*., 2014), to explore, using IRT modelling techniques, how the instruments (and the items they contain) performed when administered in samples of

healthcare students, compared to registered health and social care staff. The application of appropriate IRT models is dependent on knowing whether the items in question form a unidimensional scale (as was the case for the KIDE) or a multi-dimensional scale (as the DKAS developers reported).

The initial stage of the analysis for the DKAS data was to explore the underlying structure of the scale, to establish whether the four-factor model reported by the instrument developers (Annear *et al*., 2017) was mirrored in this sample.

Methods

Sample and setting

First-year nursing undergraduates were recruited, using a purposive sampling approach, from a nursing programme in one university in Scotland. Upon ethical approval, permission to access students was sought via the head of undergraduate studies as the gatekeeper. The sample consisted of 479 students who were enrolled in their first year of the 2019/2020 adult nursing, mental health nursing, or child nursing pathway, across two campuses.

Dementia knowledge instrument

The DKAS (Annear *et al.,* 2015b) was developed and validated for use in healthcare staff and students, with items generated on the basis of an international Delphi survey and a literature review. It contains 25 statements that cover four domains of dementia knowledge (causes and characteristics, communication and engagement, care needs, risks and health promotion).

The DKAS was originally developed with a modified Likert scale with five options for response: false, probably false, probably true, true, don't know. In order to reduce burden and for the response format to be consistent with the KIDE, response options were reduced to: True, False, Don't know.

The DKAS has been administered in predominantly Australian samples of healthcare staff as part of a massive open online course (MOOC) and has been subject to more rigorous psychometric testing and evaluation than the KIDE. In a comparative

validation study (Annear et al., 2016) between the DKAS and the Alzheimer Disease Knowledge Scale (ADKS) (Carpenter *et al*., 2009), the DKAS was reported as superior with regards to parameters of response, minimisation of ceiling effects, and ability to discriminate between pre-and post-intervention scores.

The DKAS is the only KoD instrument in current use to have tested and confirmed scale structure using EFA and CFA. A strong four-factor model (GFI = .974: RMSEA = .040) was reported in a large, diverse sample (n=3649) of health and care staff, healthcare students, family carers, and general population. The item content of the DKAS is reported in *Table a,* and the factor structure as reported by the developers is displayed in *Figure a*.

Table a) Item content and codes for the DKAS instrument

Item	Item wording	Item code
1	Dementia is a normal part of the ageing process	Normal ageing
2	Alzheimer's Disease is the most common form of dementia	Alzheimer's
3	People can recover from the most common forms of dementia	Recover
4	Dementia does not result from physical changes in the brain	Brain physiology
5	Planning for end of life care is generally not necessary following a diagnosis of dementia	Planning
6	Blood vessel disease (vascular dementia) is the most common form of dementia	Vascular
7	Most forms of dementia do not generally shorten a person's life	Lifespan
8	Having high blood pressure increases a person's risk of developing dementia	Blood pressure
9	Maintaining a healthy lifestyle does not reduce the risk of developing the most common forms of dementia	Healthy lifestyle
10	Symptoms of depression can be mistaken for symptoms of dementia	Depression and dementia
11	Exercise is generally beneficial for people with dementia	Exercise
12	Early diagnosis of dementia does not generally improve quality of life for people experiencing the condition	Early diagnosis
13	The sudden onset of cognitive problems is characteristic of common forms of dementia	Sudden onset
14	It is impossible to communicate with a person who has advanced dementia	Communication
15	A person experiencing advanced dementia will not generally react to changes in their physical environment	Environmental changes

Item	Item wording	Item code
16	It is important to correct a person with dementia when they are confused	Correcting dementia
17	People experiencing advanced dementia often communicate through body language	Body language
18	Uncharacteristic behaviours in a person experiencing dementia are usually a response to unmet needs	Behaviour and needs
19	Medications are the most effective way of treating behavioural symptoms of dementia	Medication and behaviour
20	People experiencing dementia do not generally have problems making decisions	Decision making
21	Movement is generally affected in the later stages of dementia	Movement affect
22	People with advanced dementia may have difficulty speaking	Difficulty speaking
23	People experiencing dementia often have difficulty learning new skills	Learning skills
24	Difficulty eating and drinking generally occurs in the later stages of dementia	Eating and drinking
25	Daily care for a person with advanced dementia is most effective when it focuses on providing comfort	Daily care

Figure a) Reported four factor structure of the DKAS

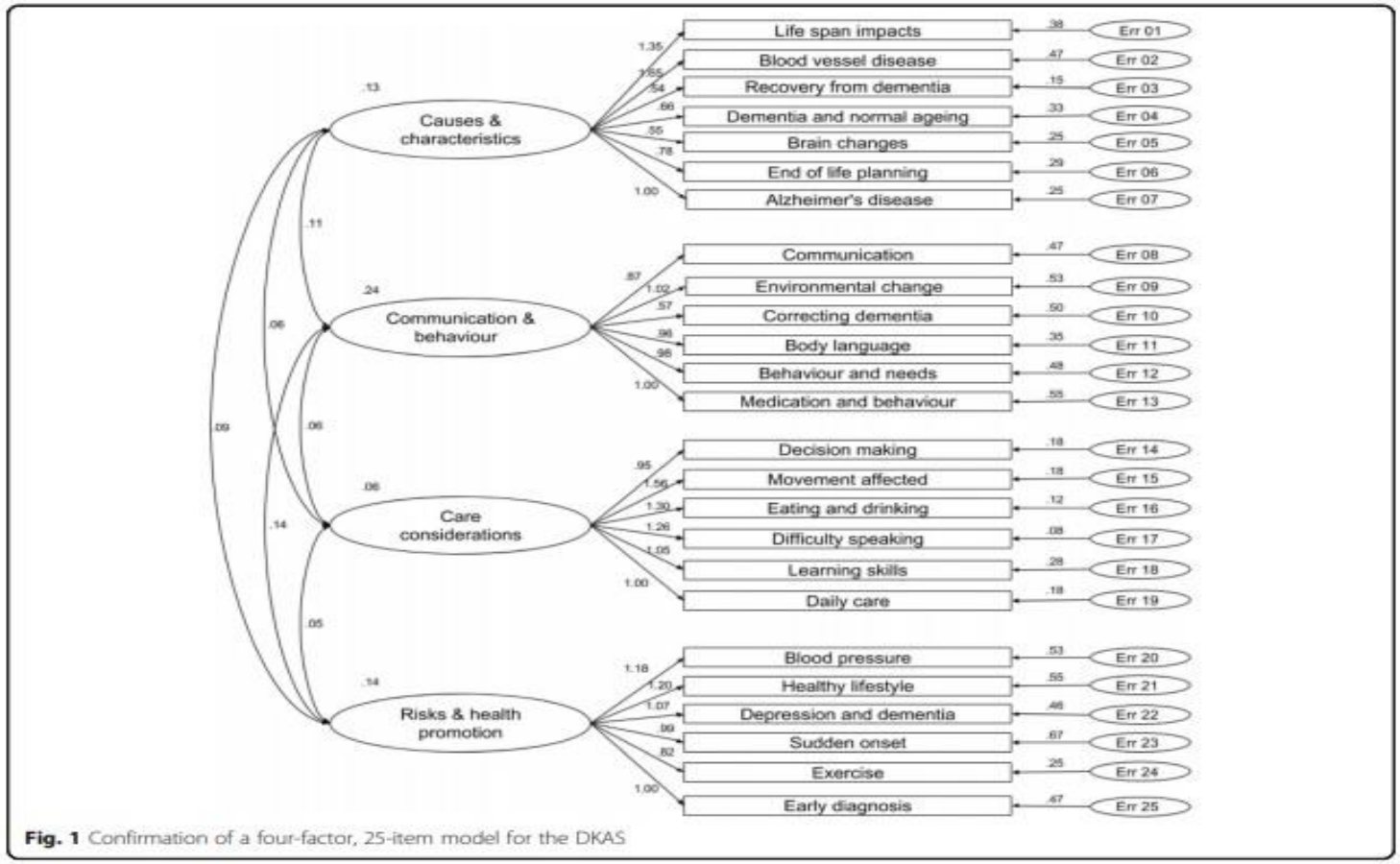


Fig. 1 Confirmation of a four-factor, 25-item model for the DKAS

Statistical analysis methods

The underlying structure of the DKAS was examined using the following methods:

1. Examination of tetrachoric correlations appropriate to binary data (for patterns of very low or very high correlations)
2. Examination of scree plot of eigenvalues generated from tetrachoric correlations
3. EFA - maximum likelihood factor analysis using oblimin rotation

All analyses were conducted using RStudio (version 3.6.1)

Results

Sample

Data were collected from 404 undergraduate students. The participants had been on the course for one month, and had undertaken nursing theory only, no clinical placement. Across campuses there were a total of 479 students, 363 at campus 1 and 116 at campus 2. At campus 1, 342 students attended the lecture and 318 agreed to participate in the study. Four survey packs contained incomplete consent forms and were therefore not useable, leaving 314 completed survey packs. At campus 2, 98 students attended the lecture and 93 agreed to participate. There were 9 survey packs with either no consent form or more than half of the questions unanswered; these were removed, leaving 84 completed survey packs from this campus.

Overall, the response rate was very strong, at 92% (93% for campus 1, and 95% for campus 2).The participants were predominantly female (92%), aged between 18-25 (60%), and did not have degree-level education prior to their current course (83%). Further, 64% of participants reporting knowing someone with dementia, and 52% reported having previously worked with people living with dementia.

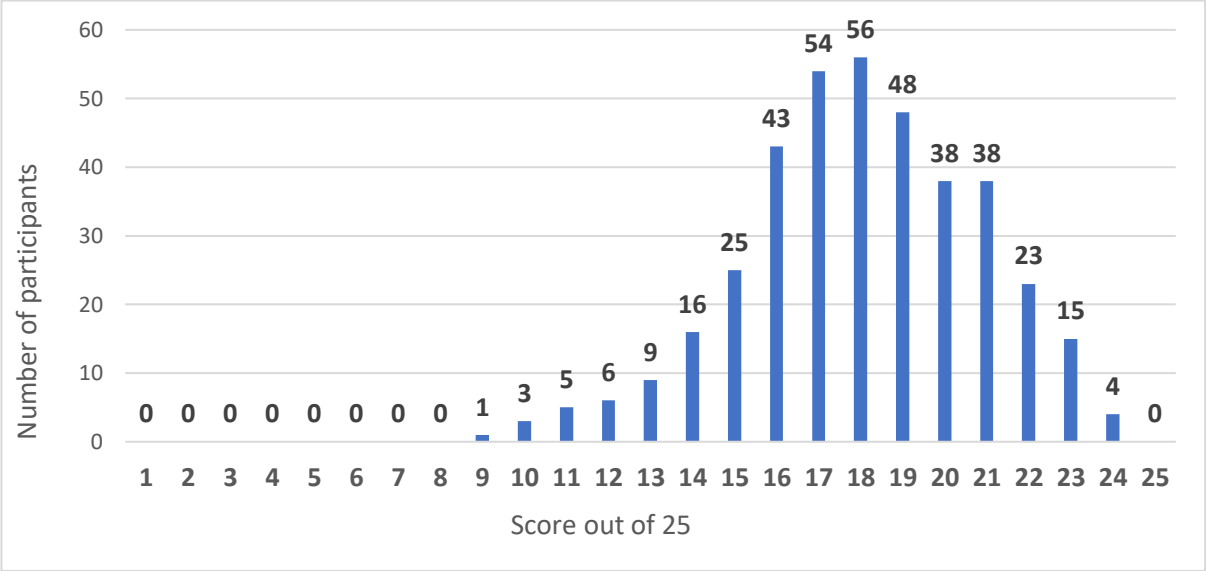
Cases with missing data were removed since a number of the R packages used throughout the analysis do not accept missing data. Therefore the final sample was (n=384).

Descriptive statistics

The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.61; the recommended value for suitability for factor analysis is 0.6 or above.

Figure b shows the distribution of "number correct scores" from participant responses to the DKAS. The distribution of scores was reasonably even, though predominantly towards the higher end of the scale. Most scores were in the range of 15 – 22 out of 25. The lowest score achieved was 9 out of the possible 25, and no participants achieved 100% correct responses.

Figure b) Distribution of DKAS scores for undergraduate nursing students



Tetrachoric correlations between item pairs are shown below in *Table b*. Associations between items were predominantly very low, confirming that the DKAS is not a unidimensional scale. No groupings or patterns of higher correlating items that may indicate the presence of factors were evident.

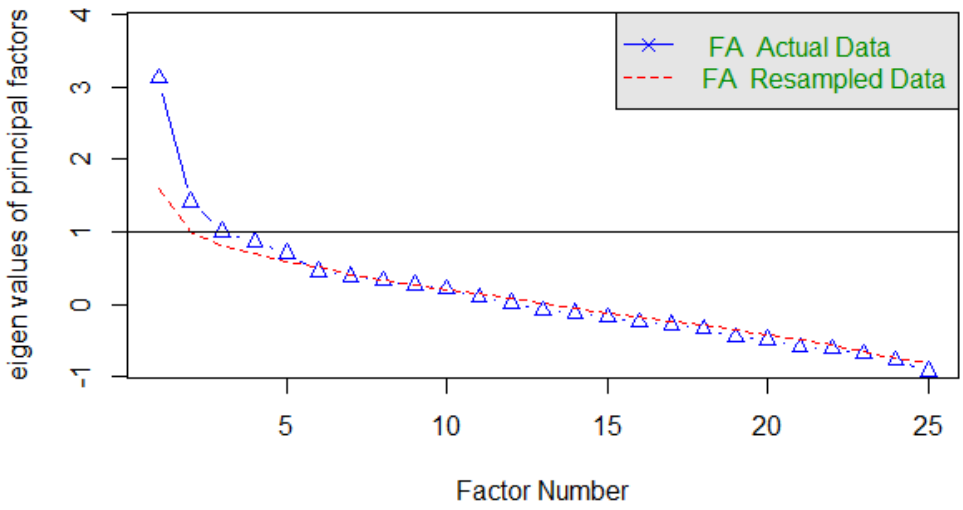
Table b) Matrix of tetrachoric correlations of pairs of DKAS items

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	-0.102																								
3	0.129	-0.025																							
4	0.018	0.009	0.225																						
5	0.099	-0.137	0.041	0.120																					
6	0.229	-0.043	0.035	-0.155	0.087																				
7	-0.020	0.563	-0.366	-0.056	-0.107	0.119																			
8	-0.083	0.082	0.236	0.100	-0.057	-0.017	-0.111																		
9	0.193	0.157	0.134	0.240	0.007	-0.032	0.131	0.141																	
10	0.162	-0.030	-0.018	0.081	-0.073	-0.004	-0.005	0.110	0.171																
11	0.153	-0.046	-0.035	-0.195	0.226	0.142	0.060	-0.022	0.206	0.070															
12	0.126	-0.100	-0.270	0.025	-0.052	0.126	-0.002	0.051	0.194	0.106	0.263														
13	0.121	0.005	-0.209	0.118	-0.054	0.079	-0.121	0.281	0.162	0.126	0.090	0.188													
14	0.245	-0.138	0.113	0.018	0.259	0.173	-0.063	-0.174	-0.048	0.140	0.165	0.253	-0.114												
15	0.219	-0.060	0.015	-0.057	0.109	0.068	0.139	-0.105	0.193	0.116	0.248	0.159	0.016	0.424											
16	0.040	-0.094	0.049	0.051	-0.032	-0.078	0.191	0.072	0.158	-0.053	0.313	0.281	-0.007	0.380	0.617										
17	0.180	-0.224	-0.096	0.022	0.219	0.240	0.104	-0.034	0.089	0.095	0.400	0.277	0.049	0.473	0.521	0.524									
18	0.057	0.001	0.000	-0.126	0.244	-0.040	0.056	0.036	0.235	-0.074	0.155	-0.027	-0.271	0.229	0.174	0.416	0.430								
19	0.093	-0.083	0.112	0.129	0.265	-0.019	-0.084	0.166	-0.047	0.029	0.028	0.142	-0.042	0.177	0.003	0.057	-0.003	0.201							
20	0.166	-0.037	0.068	0.046	0.208	-0.022	-0.109	0.038	0.149	0.070	0.197	0.196	0.154	0.136	-0.028	0.138	0.143	0.029	0.148						
21	0.142	-0.023	0.074	0.007	0.095	0.080	-0.141	0.068	0.027	0.049	0.223	0.303	0.027	-0.026	-0.006	0.171	0.201	-0.054	0.057	0.530					
22	0.013	0.009	-0.016	0.128	0.061	-0.120	-0.048	0.020	-0.086	0.027	0.218	0.218	-0.109	0.072	-0.031	0.108	-0.134	-0.022	0.137	0.242	0.219				
23	0.144	-0.061	-0.004	0.190	-0.139	0.043	-0.064	0.133	0.123	-0.036	-0.279	-0.022	-0.011	-0.101	0.004	-0.033	-0.212	0.002	-0.270	-0.002	-0.114	-0.107			
24	0.201	-0.098	-0.369	0.009	0.260	0.107	0.019	-0.049	0.167	0.257	0.333	0.379	-0.138	0.040	0.086	0.181	0.279	0.128	0.136	0.327	0.341	0.143	0.041		
25	0.110	-0.110	-0.022	0.142	0.150	0.117	-0.012	0.007	0.101	-0.049	0.110	0.141	0.170	0.054	0.039	0.050	0.193	-0.118	-0.004	0.203	0.300	0.105	-0.085	0.241	

Scree plot of eigenvalues

The first four eigenvalues from the observed data tetrachoric correlation matrix were 3.12, 1.41, 1.0, and 0.87 *(% of variance*). Parallel analysis was conducted to plot and compare the eigenvalues of the tetrachoric matrix for the observed data alongside those for simulated data with the same difficulties as the observed data (*Figure c*). Factors were retained based on the 'number of eigenvalues before the elbow of the plot' rule (DeMars, 2010). The scree plot suggested the presence of one or two dominant factors. Another commonly used factor retention method is to retain factors with eigenvalues greater than one (Mair, 2019), which would suggest two factors should be retained.

Figure c) Scree plot of eigenvalues. The blue (triangles) line shows eigenvalues of the observed data, the red (dashed) line depicts eigenvalues in random matrices of the same size (and difficulty parameters) as the original data matrix



Exploratory Factor Analysis results

Maximum likelihood factor analysis using oblimin rotation was conducted for four-factor, three-factor, two-factor, and one-factor solutions. Resulting factor loadings suggested that a

two-factor solution was the most appropriate model for this data. Loadings for the three-factor exploratory model are shown in *Table 3*; factor loadings of less than 0.2 have been omitted to demonstrate how items potentially group to form subscales similar to the original factor 2 (communication and behaviour) and factor 3 (care considerations).

Table c) Factor loadings for 3-factor solution. In column 1, items have been grouped according to their original factors from the DKAS development studies.

	EFA factor 1	**EFA factor 2**	**EFA factor 3**
Factor 1: Normal ageing			**0.225**
Factor 1: Alzheimer's Disease			
Factor 1: Recovery	0.998		
Factor 1: Brain physiology	0.267		
Factor 1: Planning			**0.340**
Factor 1: Vascular			
Factor 1: Lifespan		**0.230**	
Factor 2: Communication		0.549	
Factor 2: Environment		0.758	
Factor 2: Correcting		0.742	
Factor 2: Body language		0.732	
Factor 2: Unmet needs		0.449	
Factor 2: Medication			**0.271**
Factor 3: Decisions			0.648
Factor 3: Movements			0.640
Factor 3: Speaking			0.327
Factor 3: Skills learning			
Factor 3: Eating			0.539
Factor 3: Comfort			0.441
Factor 4: Hypertension	0.22		
Factor 4: Lifestyle			
Factor 4: Depression			

Factor 4: Exercise		**0.412**	0.274
Factor 4: Early diagnosis		0.229	**0.351**
Factor 4: Sudden onset			