Quality of life in childhood

(Volume 1)

Joanne Cremeens

Thesis submitted to the University of Sheffield, Department of Psychology for the degree of Doctor of Philosophy.

May, 2004.

Abstract

The aim of this thesis was to develop a child self-report quality of life (QOL) measure for children below eight years. Two questions were central to the development of our instrument. First, can children below eight years self-report on their thoughts, feelings, and lives? Second, if so what are the best ways to gain self-reports from children? In answering these questions, we produced a set of guidelines that can be applied by researchers developing self-report measures for children.

Studies 1 and 2 report the initial validation of our child self-report QOL measure (the teddy bear QOL measure, TedQL.1 & 2). In Study 1, children's TedQL.1 scores were positively correlated to their scores on an established measure (the PedsQLTM4.0). In Study 2, the response scale used to complete TedQL.2 items impacted on the psychometric properties of our measure. Study 3 reported further development of the content of our measure, using interview data from children about their lives. Based on the results of Study 3, a new version of our measure was developed (due to deletion, alteration, and addition of items).

Study 4 established the most appropriate response scale for the TedQL.4, by comparing the psychometric properties of children's responses to TedQL.3 items across three response scales. Study 4 showed that children used concrete examples of specific situations to answer the TedQL.3 items, which may explain why young children's self-reports are less stable over time compared to older children. The analysis in Study 4 revealed eight items that could be removed from the TedQL.3. Study 5 reported further validation of the child and parent versions of the TedQL.4. Both children's and parent's TedQL.4 scores were correlated to their PedsQLTM4.0 scores. No relations between child and parent rated child QOL were found for the PedsQLTM4.0 scores, however children's and parent's TedQL.4 scores were correlated across some of their scores.

This thesis has shown the importance of gaining self-reports from children themselves, and highlighted the best methods to use for such instruments. The applications of our TedQL measure have been discussed in the concluding section.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank both the institution that funded my research and the people who have helped me during the past four years. First, the psychology department and Sheffield University for providing me with a studentship since I started my research in 2000. Thank you for your financial support which enabled me to conduct my research while I was in Sheffield. Second, Professor Christine Eiser for giving me the opportunity to come and conduct my research with the CRUK Group in Sheffield. I want to thank her and Dr Mark Blades for all their guidance, knowledge, support, and patience while supervising me throughout my research. I would also like to thank Dr David Saarnio (University of Arkansas) who helped me during my final writing-up year after I moved to Jonesboro, Arkansas.

I would also like to thank the schools, children, and parents who took part in my research studies, from both South Yorkshire and Kent. I would like to make a special mention to Bridge and Patrixbourne Primary School and St. Mary's Primary School for allowing me to return many times to work with their children. I really enjoyed my time working in these schools, and without your help my research would not have been possible.

I also want to thank my friends who have helped and supported me throughout my research, both in Sheffield and further a field. They have all been wonderful, thank you especially Kate, Yvonne, Sarah, Lou, and Claire.

Finally, I would like to say a big thank you to my family – my parents, my brother and his wife, and of course my husband, Keith. My parents have constantly been there for me, and have followed my progress throughout my post-graduate studies with interest and excitement. You are the best, mum and dad, thank you. And my husband Keith, you have supported me and helped during my move to Arkansas, and you have always been there to listen and help me through it all. Thank you, I couldn't have done it without you.

Declaration

The author was funded by a studentship from the Psychology Department, University of Sheffield. The author was married during completion of this thesis, and consequently had a name change from Lawford to Cremeens.

The data for Study 1 (reported in Chapter 4) was collected by an undergraduate, Miss Nicole Volvaka, studying in the Department of Psychology, University of Sheffield. All the data analysis for Study 1 was conducted by the author. Two papers were published as a result of the first two studies in this thesis (please see the reference section under Lawford). The data for all the other studies reported in this thesis were collected by the author.

Section 1: Obtaining self-reports from young children

Chapter 1: Gaining information on young children's views and experiences.

Summary

Information about young children's lives has been collected by proxy report, using parents, guardians, teachers, or health professionals. This has been due to the assumption that children below eight years can not accurately and reliably report on their thoughts and feelings about their lives. However the assumption that children below eight years can not self-report has been challenged, partly due to evidence from the developmental literature for the emergence of cognitive skills from a young age. Researchers have also found evidence of lack of agreement between proxy and child reports, which has led to calls for child self-report instruments to be developed. The changing legal status of children and children's rights has provided additional impetus for developing child self-report instruments.

In this thesis we focused on the concept of QOL, and the measurement of QOL in children below eight years. The aim of the empirical studies in this thesis was to develop a child-centred generic QOL self-report measure for children aged below eight years.

QOL has become an important concept in the medical and psychological literature for a variety of reasons. The changing epidemiology of childhood disease and improvements in medical technology have meant that survival is longer an appropriate endpoint for choosing treatment options. Measuring QOL has provided one solution to the need for patient-centred outcome instruments. QOL has been defined in a variety of ways, and this has led to wide variation in the types of domains and items that have been included in QOL instruments. We drew on the WHO definition of QOL, and also the work of Calman (1987) and Bergner (1989) in defining QOL in this thesis.

1.1 Introduction

This chapter discussed the necessity for researchers to gain self-reports from young children on their lives, thoughts and feelings. First, we discussed how information on young children's lives has been collected (i.e., by proxy report, usually parents). Second, we considered the evidence and arguments for gaining *self-reports* from young children. Third, we discussed the concept of QOL. The concept of QOL provided the focus for the empirical work conducted in this thesis. The aim of our work was to develop a generic child self-report QOL measure for children below eight years. (When we refer to 'young children' in this thesis we were referring to children aged below eight years, unless otherwise stated).

1.2 Gaining information about young children's lives

Questionnaire-based measures have been the preferred method over interviews or observational methods for gaining information about young children's lives, as questionnaires can be designed to be quick, easy, and simple to administer (Ravens-Sieberer & Bullinger, 2002).

For many subjective psychological concepts such as pain and QOL, we must rely on self-reports (Beyer & Knapp, 1986). Self-reports are by very nature subjective, constructed within an individual's memory and hence prone to many biases and misrepresentations, even those provided by adults. Traditionally self-report measures of concepts such as QOL, self-esteem, and mental health have been aimed at children above eight years, due to the assumption that children under this age cannot accurately self-report their thoughts and feelings (Priestley & Pipe, 1997, Tyler & Krane, 1990). There has been concern that young children could not accurately recall events, their reports were highly susceptible to suggestion and fabrication, and that they have difficulty in distinguishing reality from fantasy (Priestley & Pipe, 1997). Due to these concerns, until recent years information about young children's lives below eight years has been gained using proxy reports, such as parents, guardians, teachers, and health professionals (Landgraf, 1996, O'Donoghue & Archbold, 2002). For example, Bullinger and Ravens-Sieberer (1995) conducted a review of the literature relating to QOL measures. Bullinger and Ravens-Sieberer (1995) reported that in over 50% of the child studies they identified parents were used to rate the child's QOL, and another 40% of the studies used clinic staff or health professionals as proxy ratings.

However interest in gaining information about young children's lives directly from children themselves has increased in recent years, and some researchers have developed *self-report* measures for children below eight years (e.g., Chapman & Tumner, 1995, Christie, French, Sowden, & West, 1993, Collier, Mackinlay, & Phillips, 2000, Harter & Pike, 1984, Riley, Forrest, Rebok, Starfield, Green, Robertson, & Friello, 2004, Valla, 2000, Varni, Katz, Quiggins, & Friedman-Bender, 1998). Researchers have begun to acknowledge that children should be involved more in decisions about their own lives, families, and health care (Alderson & Montgomery, 1996, Hart & Chesson, 1998).

1.3 Importance of gaining *self-reports* from young children

There are three reasons that can be cited as arguments for the necessity to develop child *self-report* measures: 1) increasing evidence for the emergence of cognitive skills earlier than previously thought; 2) lack of concordance between child and proxy reports; and 3) the changing legal status of young children. These reasons have been discussed below.

Evidence for the emergence of cognitive skills and understanding from an early age

There is evidence that young children can understand much more than has been previously thought (e.g., Choy & Mahoy, 1998, Eder, 1990, Youngstrom & Goodman, 2001), and that their cognitive skills may be more advanced than some developmental theories have advocated (e.g., Piaget, 1929). The Piagetian stage model of development maintained that children under eight years have no clear conceptual understanding and have not yet mastered the skills necessary to think abstractly about themselves or others (Flavell, Miller & Miller, 2002). However evidence for the existence of more advanced cognitive skills in children as young as preschool as been found by some researchers (e.g., Flavell, 1999, Siegal, 1997). Such work has challenged the Piagetian model of development, and provided impetus for a more optimistic view of young children's cognitive skills and abilities.

There has been a shift in emphasis in developmental research from focusing on what children cannot do, towards attempting to highlight their competencies from a young age (Kalish, 1999). Researchers have begun to develop more sensitive methodologies that minimise extra task demands on children, which in turn have revealed evidence of advanced cognitive skills from a young age (Flavell et al., 2002). We have discussed these issues in more detail in Chapter 2 (p. 16-19).

Lack of concordance between child and proxy (parent) reports

Researchers have advocated that a necessary requirement for the validation of new child measures is agreement between proxy and child reports (e.g., Graham, Stevenson, & Flynn, 1997, Langeveld, Koot, Loonen, Hazebroek-Kampschreur, & Passchier, 1996). However a number of researchers have shown that the level of parent-child agreement may be quite poor (e.g., Ennett, Devellis, Earp, Kredich, Warren, & Wilhelm, 1991, Glaser, Davies, Walker & Brazier, 1997, Langeveld, Koot, & Passchier, 1997, Le Coq, Colland, Boeke, Bezemer, & Van Ejik, 2000, Vance, Morse, Jenney & Eiser, 2001, Vogels, Verrips, Verloove-Vanhorick, Fekkes, Kamphuis, Koopman, & Theunissen, 1998). For example Ennett et al. (1991) found that parents reported more negative consequences of illness on their children's QOL than the children themselves did.

Eiser and Morse (2001) argued that if concordance between child and parent report is poor this may not mean that a measure is inadequate, but may be a result of differing perspectives between parents and children. There may be a variety of reasons why children's and their parent's views about children's lives do not match each other. Vance et al. (2001) argued that parents report their child's lives from their own perspective as both an adult and a parent, rather than considering how their child feels about their lives. Eiser, Vance, Horne, Glaser, and Galvin (2003) make the point that parents may not actually know much about certain aspects of their children's lives, for example about their friendships or school lives. Children may also be adept at hiding their feelings from their parents (Eiser et al., 2003), especially as with increasing age they search for their own sense of independence. Guyatt, Juniper, Griffith, Feeny, and Ferrie, (1997) reported that parents of young children were likely to use information gained from the more easily observable aspects of their children's lives (i.e., from their child's actual overt external behaviour), and therefore parents may have little insight into subjective aspects (i.e., their children's thoughts and feelings).

The lack of concordance found between parent and child reports has led researchers to obtain both proxy and child ratings, and to consider the relative value of these differing perspectives, rather than trying to resolve any lack of agreement (Koot & Wallander, 2001). Parsons, Barlow, Levy, Supran, and Kaplan (1999) argued that the issue is not "who is right" but "what each rater contributes to understanding children's lives". The need to obtain the child's point of view whenever possible has been increasingly recognised, which has made the development of appropriate child *self-report* measures important.

Changing legal status of young children in medical decision-making

In the past parents were provided with the right and responsibility to provide consent for their children (i.e., under 18 years) as parents were seen as the most competent to make decisions for their children, and it was assumed that they would also be motivated to act in their child's best interests (McCabe, 1996). For example, the Children's Act of 1948 emphasised the importance of considering children's best interests. The focus of this piece of legislation was to protect children within a framework of adult decision-making (Sinclair Taylor, 2000). The UN declaration on the Rights of Children in 1959 highlighted the need for specific care and protection of children in terms of nutrition, medical care, and education. However decisions for children below 18 years of age were still made primarily by parents or other responsible adults, although this was modified to 16 years of age under the Family Law Reform Act in 1969 (Peterson & Siegal, 1999).

Children's legal status has changed over the last two decades (Peterson & Siegal, 1999). Several child abuse cases in the 1970's highlighted the need for children's own views to be recognised in legal situations. One case played an important part in helping the recognition of children's views – the Gillick versus West Norfolk and Wisbech Area Health Authority case in 1986. As a result of this case the House of Lords ruled that any child can make a case for competency for making a medical decision (known as the "Gillick competency", Peterson & Siegal, 1999). To qualify for this a child needs to show enough intelligence and knowledge to fully understand treatments and their consequences, and the level of understanding required to make a decision is directly related to the decision to be taken (Masson, 2000).

There is evidence that children can be competent in legal contexts. Researchers have provided evidence that preschoolers can accurately recall events over long time periods

(Fivush, Haden & Adam, 1995, Fivush, Haden & Reese, 1996, Hudson & Fivush, 1991), and can resist misleading suggestions (Newcombe & Siegal, 1997, Saywitz, Goodman, Nicholas & Moan, 1991). Peterson (1991) showed that three-year old children could understand the distinction between truthfulness and lying and could also appreciate the importance of truth in court. McCabe (1996) argued that young children are able to understand simple information and to ask questions in situations of uncertainty. Weithorn and Campbell (1982) showed that children as young as nine were able to participate meaningfully in making decisions on their personal health-care.

The UN Convention of the Rights of the Child in 1989 gave children various political and civil rights – such as the right to freedom of speech (Sinclair Taylor, 2000). Following this, other statutes have recognised children's capacity to make specific decisions at certain ages (Masson, 2000). Young children are now within the protection of the European Convention on Human Rights (linked to the Human Rights Act of 1998) and as such they should be afforded the same rights as adults whenever possible (Lowden, 2002). Gaining information from children themselves for treatment decisions is becoming necessary from a legal standpoint (Department of Health, 2000). Therefore it is important to be able to gain information from children directly about their own views, thoughts, and feelings.

1.4 Introducing the concept of QOL

In this thesis we focused on the concept of QOL and its measurement in children below eight years. The aim of thesis was to develop a generic child self-report QOL measure for children below eight years. We aimed to use information provided directly by children themselves to inform the content, response format, and presentation style of our measure, and also use the lessons learned from the existing literature in the development and validation of our instrument. In the following sections we considered the importance of QOL, and discussed how QOL has been defined, operationalised, and measured by researchers.

Importance of QOL

The concept of QOL has become an important concept within medical and psychological research due to a paradigm shift in medical thinking and research (Joyce, McGee & O'Boyle, 1999). Researchers have begun to evaluate outcomes using patient-centred outcome measures as opposed to relying on mortality rates and reductions in symptomatology (Eiser, 2002).

The changing epidemiology of childhood disease (from acute to chronic, and from mere survival to management over time) was the major impetus for researchers for developing child QOL instruments (Eiser & Morse, 2001). During the 1980s advances in medical technology and treatment resulted in increased survival rates for children with life-threatening conditions. Improvements were seen in areas such as paediatric oncology, with survival rates increasing from less than 50% in the 1960s to over 70% in the 1990s (Stiller, 1994). Survival rates also increased for cystic fibrosis, heart disease and many other childhood conditions (Gortmaker & Stappenfield, 1984). In addition medical decision-making became complex with the development of an array of treatment options, each with relative advantages and disadvantages, making decisions as to the 'best' treatment difficult (McCabe, 1996). Therefore outcome measures need to reflect the fact that survival alone is no longer an appropriate way to choose between treatment options (i.e., treatments should not only increase life expectancy but also improve life quality, Eiser, 2002). Measuring QOL provides a solution to the need for more comprehensive patient-centred outcome measures.

Conceptualising QOL

The concept of QOL can have different meanings for every individual, and these meanings are dependent on cultural, social, and economic circumstances. Different approaches can be taken when conceptualising QOL. The philosophical approach is concerned with relating QOL to rising to challenges and coping with adversity, while the economic approach focuses on the allocation of resources and the accumulation of wealth. In this thesis we have taken both a psychological and a medical approach to the concept of QOL. From a clinical perspective QOL we focused on young children's functional health status and their observable behaviour. From a psychological perspective we focused on the subjective aspects of QOL (i.e., the meaning and impact of functioning for young children).

Defining and operationalising QOL

The concept of QOL has been characterised by diversity in definition and a multiplicity of approaches (Hyland, 1999). There is no universally accepted definition of QOL even within the adult literature, which has resulted in QOL becoming "a kind of umbrella term under which are placed many different indexes dealing with whatever the user wants to focus on" (Feinstein, 1987, p. 636).

One of the most widely cited definitions of QOL has been provided by the World Health Organisation (WHO) who defined QOL as an individual's physical health, psychological states, level of independence, social relationships, and their relationship to salient features of the environment (WHO, 1994, World Health Group, 1995). The concept of health-related QOL (HRQOL) refers to the impact of health and illness on an individual's QOL, and includes the physical, psychological, and social domains outlined by the WHO as well as disease-specific and treatment-related symptoms (Seid, Varni, & Jacobs, 2000).

Gill and Feinstein (1994) distinguished three ways in which QOL can be operationalised for medical and research purposes. First, QOL can be measured using clinical parameters such as blood sugar level or blood pressure. Second, QOL can be assessed in terms of what an individual can actually do, i.e., walking one block or climbing a flight of stairs. Third, QOL can incorporate more subjective aspects from an individual's own perspective in their life and abilities. What is important here is the meaning and importance that individuals place on specific abilities or levels of functioning, for example not being able to do sports may not be a problem for someone who was not very active, although it may difficult for another person to accept if they are used to doing such activities. Linked to this third definition researchers such as Calman (1987) have offered alternative definitions of QOL, arguing that QOL is related to "the perceived gap between an individual's hopes and expectations and their present experience" (p. 7). Bergner (1989) extended these ideas by arguing that QOL can be enhanced by narrowing the distance between an individual's attained and desired goals.

In defining QOL we drew on the WHO definition (WHO, 1994), and also on the work of researchers such as Calman (1987) and Bergner (1989). We defined QOL as a multidimensional concept encompassing physical, psychological, social, and cognitive aspects of functioning, but we also maintained that QOL can be subjective and individual (i.e., the meaning and importance individuals place on aspects of their lives has also been taken into account).

Measuring QOL

QOL measures fall into two separate types - either generic or disease-specific, both of which have relative advantages and disadvantages. Generic measures allow assessment of QOL across many different populations and groups of children (Spieth, 2001), however such measures may not be sensitive enough to specific symptom- or treatment-related problems in ill children (Eiser & Morse, 2001). Disease-specific measures allow a more detailed examination of the impact of specific illnesses and conditions, however it may prove difficult to separate the effects of disease from other aspects of life (Eiser, 2002). We focused on developing a generic measure of QOL to provide a tool with applications for many areas of paediatric research.

There is a lack of well-validated self-report QOL measures for children below eight years of age (Feeny, Furlong, Mulhern, Barr, & Hudson, 1999). Riley et al. (2004) argued that this lack of instruments is not surprising given the obstacles to obtaining self-reports from young children. In addition the accuracy and value of the information gained from children under eight years has been debated in the literature (Riley et al., 2004, Rebok et al., 2001). Riley et al. (2004) recommend that the content, presentation, and response format of child self-report measures need to be designed to take advantage of young children's cognitive strengths.

In recognition of the lack of instruments, the aim of our empirical work was to develop a QOL self-report measure for children under eight years, using a child-centred approach. We aimed to produce an instrument that would:

i. use information from children themselves to develop the content of items
ii. provide an alternative presentation style to written measures, to increase the attractiveness of the task, and maintain children's attention
iii. establish the most appropriate response format for the targeted age group, by comparison of different response scales.

These aims have been explored in more detail in Chapters 2, 3, 6, and 7.

A theory of QOL

Many researchers have become disillusioned with the concept of QOL and it's measurement (Eiser & Morse, 2001). The lack of agreement for definition has been an impediment to QOL research and for developing QOL measures (Lindstrom, 1992, Koot, 2001), and has led to an array of different domains and items being included in instruments. Hunt (1997) makes the point that there has been a "general acceptance" that there is little agreement on the nature of QOL, and as a result a variety of conflicting definitions have been used by researchers. The lack of agreement in definition and diversity of measurement primarily stem from the absence of a theoretical understanding of what factors underlie an individual's QOL. Without a theory there is no means of linking what is actually being measured with what is supposedly being measured (Hunt, 1997). Researchers should avoid defining QOL in terms of what is being measured by instruments named QOL (Wallander, 2001).

Providing a theoretical model for child QOL would help in developing a operational definition of QOL, and to rectify the wide variation in domains and items included within existing QOL measures. A theory would also distinguish QOL from other related concepts, such as health status, functional status, well-being, optimism, and self-esteem (Jenney, Kane, & Lurie, 1995). Despite the overlap between a number of established concepts related to QOL and QOL itself, few researchers have examined the relationships between all these concepts. Wallander (1992, 2001) has argued that in the absence of a theoretical framework there is no way of distinguishing which factors are relevant to the measurement of QOL.

Therefore, we also wanted to incorporate a theoretical model in our measure of QOL. However there has been little empirical work focused on developing a theory of QOL relevant to children below eight years, or examining the appropriateness of adult or adolescent theories of QOL for younger children (Wallander, 2001). We argued that one model that could be applied to young children's QOL is based on the idea of an individual's QOL being equal to discrepancy between their 'ideal' and their 'actual' self (Bergner, 1989, Calman, 1987, see p. 9). This model involves judgements of how much an individual's current situation, abilities and functioning (actual self) differ from how they would like them to be (ideal self). Ideal selves, or standards for achievement or skills, can be formed using social comparisons to other people that individuals come into contact with (Guay, Boivin, & Hodges, 1999, Huguet, Dumas, Monteil, & Genestoux, 2001, Keil, McClintock, Kramer, & Platow, 1990). In the following chapter (see Chapter 2, see p. 26-8) we have considered the evidence for whether young children below eight years are cognitively capable of making social comparisons necessary for forming ideal selves. This provided support for using a discrepancy-based QOL measure with children below eight years.

1.5 Conclusions

This chapter considered the need to obtain information from young children themselves about their lives, thoughts, and feelings. In the past researchers have assumed that children below eight years were incapable of accurate and reliable self-reports on their lives due to their cognitive immaturity. Due to this assumption, information on children below eight years was collected by proxy report (e.g., parents). However the value of children's own perspectives on their experiences is being recognised within both the psychological and medical worlds.

The imperative to gain information from children directly has come from three main areas. First, research from the developmental literature has provided evidence that children's cognitive abilities develop at a much earlier age than had been thought (Flavell et al., 2002). This has resulted in a shift in the focus of developmental research, from focusing on young children's limitations to emphasising what they can actually do. We discussed these issues and the research evidence in more detail in Chapter 2 (see p. 16-19). Second, researchers have found evidence for a lack of concordance between proxy and child reports (e.g., Vogels et al., 1998). This lack of concordance has led researchers to recognise that information from children themselves should be collected wherever possible. Third, children's legal status has changed over the last two decades, and therefore making it necessary to include children's own views in decisions that are being made about their lives.

We introduced the concept of QOL in this chapter, a concept which provides the focus for our empirical studies. QOL is a concept that has been used more frequently since the 1980's in medical and psychological research. We have defined QOL as both a multidimensional concept (i.e., broadly assessing physical, psychological, social, and cognitive functioning) and also as a subjective concept (i.e., meaning and value placed on given aspects of functioning). As we have discussed in this chapter the aim of this thesis was to develop a child-centred generic self-report QOL measure for children below eight years.

The following chapter (Chapter 2) reviews a number of theories on children's cognitive development, and discusses young children's understanding of four concepts implicit behind items in many self-report measures.

Chapter 2: Young children's capacity to self-report - a developmental perspective.

Summary

A number of theories of children's cognitive development were reviewed. Piagetian theorists maintain that children below eight years have no clear conceptual understanding and cannot think abstractly. Researchers have challenged Piaget's stage theory. Carey (1985) argued that children are capable of complex cognitive operations but they are limited by their lack of experience. Flavell et al. (2002) argued that preschoolers do have some cognitive skills (e.g., the ability to categorise into basic concepts). There has been a shift in the developmental literature from emphasising young children's limitations to documenting their competencies, which has been influenced in part by the development of more sensitive child-centred methodologies.

Young children's understanding of five concepts implicit in child QOL measures was reviewed. Children can understand emotion states from two years, although some aspects of emotion understanding do not emerge until four years. Therefore it was viable to include simple emotion items such as whether they get cross or sad in measures aimed below eight years. At two years children have a limited representation of other people's minds, and by three years children can appreciate that mental representations are linked to but separate from the physical environment. A change occurs in children's mental representations around four years, and their mental abilities increase significantly after this age. QOL items asking about children's psychological and cognitive functioning could be justifiably used with children as young as four years. Children below eight years are capable of holding a sense of self that includes positive and negative aspects. QOL measures for this age could include items asking children about what they think they are like. By three or four years children are able to understand the physical and biological aspects of illness, although there is still debate as to the extent and coherency of children's health concepts. Items asking children about their everyday health would be appropriate for children below eight years.

2.1 Introduction

Young children's ability to self-report accurately and reliably on their health, and other subjective states (such as QOL), is dependent on their understanding and interpretation of the content of items in self-report measures. Children differ from adults in their cognitive abilities – for example young children may not understand concepts (such as emotions or self) in the same ways as adults.

This chapter reviewed the developmental literature and discussed relevant issues for gaining self-reports from young children (specifically in relation to QOL). First, we discussed contrasting theories of children's cognitive development. There are a number of theories concerning the development of cognition, which have led researchers to contrasting hypotheses as to the extent of young children's cognitive abilities. Second, we considered the evidence and arguments for how much understanding young children have in relation to: emotions, mental representations, self and others, and biology (in relation to health and illness). Understanding of such concepts is necessary for answering many items in self-report measures. The age at which children acquire such knowledge and capabilities has implications for the lower age limit for self-report. Third, we discussed the implications that our review of young children's understanding has for the development of the items included in our child self-report QOL measure.

2.2 Children's cognitive abilities – contrasting developmental theories

The accuracy, stability, and validity of young children's self-reports have been topics of debate in the literature (Shahinfar, Fox, & Leavitt, 2000). Children's ability to self-report is dependent on whether they can understand the content of items, i.e., if they are capable of understanding the concepts implicit in measures. There are a number of theories concerning the development of cognition in children. There has been much written on these developmental theories, and this work has only been briefly reviewed and discussed here.

Piagetian perspectives

Piagetian theorists argue that children move through four stages of cognitive development, and children cannot move to the next stage until they have mastered the skills from the stage before. Flavell et al. (2002) summarise these stages: from knowing the world by their overt actions on it (sensorimotor), to a symbolic representation of the world (pre-operational), to forming simple notions of causality and physical reality (concrete operational), to the ability to form and hold abstract thoughts and representations of the world, self, and others (formal operational). Piaget viewed human cognition as an active process, where development was a gradual process (Flavell et al., 2002) achieved through two other related processes: assimilation (using already acquired knowledge) and accommodation (altering existing knowledge due to new experiences). Piaget argued that children below eight years have no clear conceptual understanding and have not yet developed the ability for abstract thinking. While Piaget's theory is very comprehensive, Piaget may have under-estimated the capabilities of the preschool and early school-age child. For example, Flavell et al. (2002) argued that performance at various experimental tasks that Piaget used as evidence for his stage theory (such as conservation and perspective-taking tasks) may not actually reflect competence. Children may be able to understand what is involved in a given task before they can show correct performance at it.

Theories contrasting to traditional viewpoints

Researchers have challenged Piaget's stage theory (e.g., Flavell et al., 2002, Siegal, 1997), and Piaget himself placed less emphasis on his cognitive stages in later work. Flavell et al. (2002) discuss evidence that children's cognitive development is less stage-like than advocated by Piaget's theory, and also that infants and pre-school children are more competent than Piaget had suggested. Based on these findings subsequent theories of cognitive development have been proposed, many of which have taken a more optimistic view of young children's abilities.

Some researchers have suggested that although children below eight years may have not developed the ability to perform complex mental operations and manipulate complex abstract thought, they may have cognitive skills needed to perform the cognitive tasks involved in processing, interpreting, and answering the types of items in self-report measures (e.g., Carey, 1985, Siegal, 1997). Carey (1985) argued that the main difference between adults' and children's knowledge is 'domain-specificity', i.e., that children are 'novices' and adults are 'experts' in specific knowledge domains. Flavell et al. (2002) argued from a similar position: that younger children have 'content specific' knowledge (they have limited knowledge in specific domains which restricts the level of their concepts and mental reasoning). Flavell et al. (2002) have also argued that preschoolers possess more knowledge and potential for cognitive operations than previously thought. For example, preschoolers can use some external representations, can form basic concepts to organise the world into categories, and have an intuitive grasp of number (Flavell et al., 2002). Flavell et al. (2002) argued that preschoolers' knowledge is largely acquired informally, and that they form an intuitive understanding of objects, events and people.

Siegal (1997) maintained that preschoolers have some understanding of the physical and mental world, even if their understanding is restricted to a specific set of contexts. Young children may be thinking in the same ways as older children, but they may simply have less knowledge on given subjects when compared to adults or older children (Kalish & Gelman, 1992). However this does mean they are not capable of holding their own theories about the world, just that their theories are different to those held by adults (Siegal, 1997). Flavell et al. (2002) argued that the young children's cognitive systems may not be as qualitatively different from older children's systems as has been advocated previously.

Reasons for contrasting interpretations of evidence

If children below eight years are capable of such understanding, why have studies shown these children have poor competence at various cognitive tasks in experimental settings? There has been a shift in emphasis within theoretical perspectives on young children's cognitive capacities and capabilities. Researchers coming from a Piagetian perspective concentrated on the limitations of young children (i.e., focusing on what they could not do). The methodologies used by such researchers favoured tasks that relied heavily on specific knowledge and relevant experiences, and if children showed poor competence at such tasks it was assumed they were incapable of understanding the given domain/concept (Siegal & Peterson, 1996). Studies using tasks and procedures that exceeded children's capacities may have under-estimated young children's cognitive abilities. For example, repeated questioning, unfamiliar contexts, or unconventional language may have hampered their ability to demonstrate what they know (Siegal, 1997). Trabasso (1977) argued that the level of knowledge and/or competence children reveal during tasks relates directly to their understanding of the task itself, and whether children understand the questions they are asked is perhaps the most important predictor of how much they report (Trabassso, 1977).

There has been a shift towards studying what young children can do, and on seeing children as 'active theory builders' rather than 'passive recipients of information' (Kalish, 1997), and this shift in emphasis has motivated researchers develop new methods and tasks to gain as much information on young children's abilities as possible (Flavell et al., 2002). As more sensitive child-centred techniques have been developed (i.e., tasks which minimise any extra demands, are set in familiar settings, and use clear, relevant, and explicit information), researchers have found evidence of advanced cognitive skills in children below eight years (e.g., Siegal, 1997, Flavell, 1999). Using such methodologies researchers have found evidence that children under eight years do have the cognitive abilities to understand concepts implicit within items in self-report measures.

2.3 Children's understanding of concepts implicit in self-report measures

Young children's understanding of the following concepts was considered:

i. emotions

ii. mental representations

iii. self

iv. social comparisons

v. biology (in relation to health and illness understanding).

Understanding of the above concepts is necessary for answering self-report items in child QOL measures. Previous child QOL measures have included items asking on children's emotions (e.g., 'How often are you happy and smiling?' from the Generic Children's Quality of Life measure, Collier et al., 2000); mental functioning (e.g., 'How good are you at remembering things?' from the Child Health and Illness Profile – Child Edition, Rebok, Riley, Forrest, Starfield, Green, Robertson, & Tambor, 2001); self and others (e.g., 'Here is Nick watching the others play. He finds it hard to make friends with other kids. How much are you like Nick?' from Exqol, Eiser, Vance, & Seamark, 2000); and biology or physical health ('How often during the past week did you have a headache or tummy ache?', Ravens-Sieberer, Thomas, Kluth, Teschke, Lilienthal, & Bullinger, 2001). The age at which children acquire such knowledge has clear implications for the lower age boundaries of these instruments. The evidence for children's understanding of each of these concepts has been discussed in the following sections.

Understanding of emotions

Young children's understanding of emotions will directly impact on their ability to self-report their thoughts and feelings. Children need to understand and appreciate various aspects of emotions, from the realisation of the link between intentions and emotions, to an understanding of the subjective nature of emotions as internal mental states as well as being external reactions to objects. A number of researchers have maintained that understanding emotions does not begin until around school age (e.g., Harris & Lipian, 1989, Stone & Lemanek, 1990). However other researchers have argued that at the age of two or three years children can express and understand emotions (e.g., Josephs, 1994).

The ability to recognise emotional states from facial expressions, such as happiness, sadness, and anger, has been identified in children as young as four months old (e.g., Saarni & Harris, 1989). Researchers have shown that the use of emotion-descriptive language develops around two years old and increases significantly after this, with words such as 'happy', 'sad', and 'scared' being used initially (e.g., Alridge & Wood, 1997, Bretherton & Beeghly, 1982, Bretherton, Fritz, Zahnwexler, & Ridgeway, 1986). For example, Aldridge and Wood (1997) reported that at five years old children's emotion vocabulary is limited to five main words ('happy', 'alright', 'hurt', 'unhappy', & 'sad'), and after seven years old this increases to more complex terms (such as 'loneliness', 'anxiety', & 'pride').

Wellman, Harris, Banerjee, and Sinclair (1995) argued that by preschool age children show some understanding that actions and expressions are cues as to the emotional experiences of individuals. The ability to recognise more complex emotions such as pride and shame improves throughout childhood (Stone & Lemanek, 1990). However young children may have difficulty recognising that more than one emotion can occur at a time (Flavell & Miller, 1998, Harris, 1994). This difficulty may be linked to their inability to simultaneously keep in mind two or more concepts (Harter & Whitesell, 1989). Therefore young children may be less able to understand the co-occurance of different emotion states, and this should be considered by researchers designing items for self-report measures.

In addition, researchers have suggested that young children's understanding of their own emotions is likely to exceed their understanding of other people's emotions (Stone & Lemanek, 1990). This is not only because they have more personal experience of their own emotions, but also as they tend to take other's emotions at face value (Saarni, 1984). This may be linked to the argument that young children have not yet learnt the display rules that guide emotional displays in social situations (Hochschild, 1979). However, although they may not display full understanding of these rules, young children may still be able to use these rules in everyday situations (Saarni, 1984). Saarni (1984) reported that preschoolers would attempt to hide their disappointment at receiving a gift they did not like in the presence of another adult.

A critical feature of emotion conception is the realisation that emotions are related to an individual's expectations for an event (i.e., if reality matches your expectations you feel happy, if it differs you feel sad/cross/surprised, Harris, 1994). Appreciating this aspect of emotions includes understanding that mental acts can be formed and these can guide behaviour (Flavell, 1999). Researchers have shown that three year olds understand that emotional reactions can depend on the desires that people have (Hadwin & Perner, 1991, Wellman & Banerjee, 1991, Wellman et al., 1995). Wellman and Woolley (1990) have shown that two and a half year olds can understand the link between desire and emotion. In addition Meltzoff (1995) reported that older infants recognise what a person is trying to do even if they do not succeed in achieving it. Shultz (1980) showed that by three years of age children can distinguish intended actions from non-intended ones like mistakes. The spontaneous conversations of children have also shown that they can appreciate that people have different emotional reactions to the same target (Wellman et al., 1995).

Understanding the issues associated with emotions is also linked to whether children have an understanding of the appearance-reality distinction (Banerjee, 1997). This distinction refers to emotions not only being reactions to external objects or situations, but also internal mental states (Banerjee, 1997). Understanding of this distinction helps to explain why intentions mediate people's reactions, and why people try to hide their feelings in certain situations. Researchers have argued that children below six years old have problems appreciating this distinction (e.g., Harris, Otholf, & Meerum-Terwogt, 1981, Harris, Donnelly, Guz, & Pitt Watson, 1986, Gnepp, 1983). For example, Harris et al. (1981) reported that when six year olds were asked if another person could experience an emotion in their presence without them noticing they frequently reported that this was not possible. However other researchers have shown that four and five year olds could understand this distinction as well as older children (e.g., Josephs, 1994).

In summary, some researchers have argued that children can show some early understanding of various emotional states from as young as two years old (e.g., Wellman et al., 1995, Wellman & Woolley, 1990). However, there are various important aspects of emotions that children do not acquire until later (Josephs, 1994).

Understanding of mental representations

Items in QOL measures may require children to make judgements about their own or others mental states, and therefore some understanding of mental representations is necessary for accurate self-reports on such instruments.

By two years old the content of children's spontaneous speech suggests that they have an awareness of other people's minds (Bretherton, et al., 1986). By three years old children understand that they have mental experiences that are distinct from, but connected to, their everyday physical surroundings (Flavell, 2000, Wellman & Gelman, 1992). Flavell, Green and Flavell (1995) have argued that children gain various skills about thinking during preschool. They begin to realise that only humans can engage in thinking, and that mental states like thoughts are internal and different from physical actions or objects. They also understand that the brain is essential for thinking to take place, and that thoughts can guide and effect behaviours. However, preschoolers understanding of mental representations is still limited. For example, preschoolers tend to under-estimate the amount of mental activity that people engage in, and do not yet realise that people are constantly experiencing an ever-flowing stream of mental activities (Flavell, 1999).

Theory-of-mind researchers have attempted to find out what children know about mental states, using methods like false beliefs tasks. How much children understand about the beliefs of others and how well they appreciate that beliefs guide behaviour have both been studied extensively in young children (Wellman & Gelman, 1992). Specifically false belief tasks provide evidence on how well children understand causal mental states. For example, a version of the false belief task involves the use of a box with a picture of smarties on it, and children are asked what they think is in the box, and then shown that the box actually contains crayons (Flavell, 1999). Five year old children report that they thought there were smarties in the box before they saw in it, however three year olds claim that they thought the crayons were in the box all the time (Flavell, 1999). Using the results from this and other similar false belief tasks, researchers have concluded that a change in children's ability to think about correct and incorrect representations of the world occurs around four years of age, and that preschoolers do not have mental representations of the world and do not understand that the way the world is represented mentally can differ from how the world is physically (Flavell, 1999).

In summary, many researchers have argued that children's knowledge of mental representations develop after four years of age, and during the early school age period children acquire greater knowledge about mental states (Flavell & Miller, 1998, Perner, 1991, Taylor, 1996). Early school age children are adept at reporting how and when they came to know recently acquired facts, whereas preschoolers only have a vague understanding of how knowledge has been gained (Taylor, 1996).

Understanding of self

Linked to the ability to form and hold mental representations, an understanding of self is also necessary for children to answer items in QOL measures (which may ask children about their personalities, their friendships, and relationships to other people). There has been some attention on the emergence of self understanding in young children (Marsh, Craven, & Debus, 1998). The focus has been to not only to establish at what age children develop self-concepts, but also to clarify when self-concepts become differentiated (i.e., information about the self is divided into specific areas, such as social, academic, physical, and emotional components).

Researchers supporting the Piagetian viewpoint (e.g., Rosenberg, 1979, 1986) maintained that children's sense of self develops around seven to eight years of age, and view children before this age as 'little behaviourists' relying on behaviours and external appearances to make sense of themselves. Other researchers have argued that children's ideas of self were linked to descriptions based on overt behaviours, physical abilities, activities, or possessions (e.g., Stone & Lemanek, 1990). Stone and Lemanek (1990) maintained that between seven and eight years a change occurs in children's self-concepts. Between these two years the psychological conception of self emerges, after which children are able to make distinctions between the physical and mental aspects of self. Along a similar vein Damon and Hart (1988) proposed a developmental model of self understanding, moving from a behaviourist view, to a more comparative view, to finally a more abstract psychological concept of self.

Other researchers have shown that children below eight years are capable of holding psychological self-concepts, which are an early version of a theory of self (e.g. Youngstrom & Goodman, 2001, Harter, 1998). Youngstrom and Goodman (2001) argued that two and a half year olds are capable of holding self-conceptions, and that cognitive maturation requires further development of these abilities rather than the acquisition of completely new ones. Howe and Courage (1997) have provided evidence that the cognitive self emerges about two years old, which in turn allows autobiographical memories to be stored (i.e., things that have happened to 'me') at this young age.

Researchers such as Harter (1986, 1998) and Stipek (1981) have argued that children are capable of forming and holding self-concepts before eight years, but that these concepts are global and undifferentiated (i.e., information on the self is held as an overall concept which is either good or bad). Harter (1998) also proposed a developmental model of self-understanding, where self-concepts become more complex with age as information about the self is stored in relation to different abilities and areas. Harter (1998) argued that only by middle childhood (i.e., after eight years of age) do children have the cognitive skills necessary to divide information about themselves into different areas and to hold both positive and negative self-descriptions. Therefore children's self-concepts under eight years may be global in nature due to these cognitive deficits (Harter, 1998).

However, other researchers such as Marsh, Craven, and Debus (1991) have shown that children can hold both differentiated and evaluative self-concepts below eight years of age (i.e., meaning that information about the self is stored in relation to different areas, and self-descriptions can be altered as a result of new information). Marsh et al. (1991, 1998) adapted their Self-Description Questionnaire (SDQ) for use with children below eight years, and reported evidence of clearly defined self-concepts in four to eight year olds that were divided into distinct areas such as physical, social, and emotional functioning. Chapman, Tumner and Prochnow (2000) also provided evidence for differentiated self-concepts in five to seven year olds that reflected both positive and negative self-evaluations of both ability and attitude. Chapman et al. (1995, 2000) developed a measure of young children's self-concept in relation to their reading ability, and provided longitudinal data showing that children's self-attitudes remained broadly constant across the first five years of school. Eder (1989) argued that very young children have the cognitive ability to hold a stable, psychological record of who they are, that is not tied to specific contexts. Eder (1989) characterised preschool children as "emergent psychologists" who have a rudimentary understanding of the psychological aspects of self. These researchers have maintained that children as young as four years can hold differentiated and meaningful self-concepts (Chapman & Tunmer, 1995, Eder, & Mangesdoff, 1997, Marsh et al., 1991, 1998).

Why have researchers come to such different conclusions as to the existence of self-concepts in children below eight years? As we discussed earlier in this chapter (see p. 18-19), these differences in interpretation may be due to methodological issues. The types of questions that researchers have asked children and the ways that they have presented information to children has influenced the extent of understanding that has been found in young children. Eder and Mangesdoff (1997) pointed out that researchers have tended to use open-ended questions with young children that led them to conclusions that children at such ages lacked psychological self-conceptions. Eder and Mangesdoff (1997) argued that when researchers have used items with specified answers and props as presentation aids young children have shown evidence of more complex self-conceptions. For example, Eder's (1989) puppet interview method (using two puppets to present items to children, where children could respond non-verbally or verbally) was successful in showing that three and a half year olds could hold dispositional self-descriptions that formed meaningful psychological groups.

In summary, some researchers have maintained that young children's self-conceptions are global and undifferentiated (e.g., Damon & Hart, 1998, Stone & Lemanek, 1990). However, other researchers (e.g., Chapman et al., 2000, Marsh et al., 1998) have found evidence for the existence of differentiated, evaluative self-concepts in children below eight years using innovative methodologies (e.g., interviewing children with the aid of puppets or props).

Comparing self and others (social comparisons)

The age at which children are able to make social comparisons (i.e., compare their own abilities and functioning to others, e.g. their peers) is relevant to QOL measures. The concept of QOL involves in part making judgements on how good ones' own life is compared to others' lives (i.e., individuals need a point of comparison to determine how good or bad their own abilities, behaviours, or feelings are). Not only do children need to be capable of forming and holding a concept of self, they also need to compare their self-descriptions to others around them. The question here was at what age are young children cognitively capable of making such social comparisons?

Butler (1998) argued that the ability to make social comparisons emerge at two or three years old due to the nature of the cognitive processes needed for such comparisons. While other types of comparisons may require complex mental operations (e.g., temporal comparisons of oneself over time requires comparing both concrete and abstract entities), social comparisons involve the simultaneous comparison of self to other (both of which are concrete outcomes).

Following from this, some researchers have argued that children as young as three years are adept at viewing themselves in relation to their peers (e.g., Chafel, 1986, 1991, Ruble, Boggiano, Feldman, & Loebl, 1980, Ruble, 1983, Ruble, Eisenberg, & Higgins, 1994). For example, Ruble et al. (1980) provided evidence that when children were in familiar settings they used social comparisons from as young as preschool. Chafel and Bahr (1998) also showed that three year olds were capable of making basic comparisons between themselves and their playmates. Hames (1998) argued that social comparison abilities are present in children at two years, and showed that at such ages children were sensitive to differences between themselves and others (e.g., two and three year old children adjusted their speech and behaviours when interacting with younger siblings and friends). Hames (1998) argued that preschoolers are just as able to make social comparisons as older children, however they may focus on different aspects when making comparisons (i.e., younger children are more concerned with similarities to peers than differences as they need to determine what are the normal ways to perform). Younger children may make comparisons to concrete entities, abilities and/or possessions, whereas older children may compare internal attributes such as thoughts and feelings (Ruble et al., 1980).

In summary, researchers have provided evidence that children as young as three years can make comparisons to their peers (e.g., Chafel & Bahr, 1998, Hames, 1998). However, preschool children may make different types of comparisons to those made by older children.

Understanding of biology (in relation to health and illness)

There has also been some research into the age at which children develop an understanding of biology and biological processes that are separate from their psychological and social explanations of the world. Researchers have focused on studying children's understanding of inheritance and of health and illness as examples of how young children develop a theory of biology (Siegal & Peterson, 1999). The debate has focused on the age at which children develop a causal framework (i.e., where information is causally linked together rather than being simply a list of facts about biological processes, Kalish, 1999). Here we focus on children's understanding of health and illness, as this domain of knowledge is important for self-reports on QOL items (which typically ask children about either their general health and well-being, or about specific illness-related symptoms).

Children's understanding of health and illness has often been couched in Piagetian terms, specifically in relation to Piaget's cognitive developmental stages (Rushforth, 1999). Researchers coming from Piaget's theoretical framework have argued that children below seven years are incapable of understanding health and illness in biological terms, and that it is not until ten or eleven years that children can have a good understanding of this cognitive domain (e.g., Bibace & Walsh, 1981, Burbach & Peterson, 1986, Perrin & Gerrity, 1981). Bibace and Walsh (1980, 1981) argued that children below six years have an immature understanding of cause and effect in relation to health and illness. Children may offer explanations of illness that involve magic (e.g., when asked "How do people get colds?" children of this age would answer "From the sun/god", Bibace & Walsh, 1981). They may also see illness as a punishment for wrongdoings or misbehaviour (e.g., "I was naughty and did not come in when I was told to so I got a cold", Burbach & Peterson, 1986).

The work of researchers such as Bibace and Walsh (1981) that rely on stage models have been criticised for various reasons. First, researchers have found evidence for overlaps between different stages of development, and have shown that children's understanding may not always be limited by their age or developmental stage (Kister & Patterson, 1980, Meadows, 1993). Second, stage models do not account for the effect of personal experience of specific illnesses on children's understanding that has been found by various researchers (e.g., Crisp, Ungerer, & Goodnow, 1996, McQuaid, Howard, Kopel, Rosenbaum, & Bibace, 2002, Paterson Moss-Morris, & Butler, 1999, Schmidt & Weishaupt, 1990). Third, there is increasing evidence that children below seven years can show detailed and sophisticated understanding of health and illness, and that children of this age are not limited to magical explanations (Kalish, 1996, Siegal, 1988).

Alternative approaches have been put forward to help explain how and when children develop a biological understanding of health and illness. One approach incorporates the idea of scripts or schemas (Nelson, 1985). Nelson (1985) argued that children build schemas of the events surrounding illness, and by arranging these events in a logical, temporal sequence these events acquire a meaning. Researchers supporting this theory have argued that age differences in children's understanding of their health are due to differences in the organisation and complexity of their schema for these concepts (on the basis of first hand experience with health-related actions and messages children receive from adults about health, Normandeau, Kalnins, Jutras, & Hanigan, 1998). Another model put forward by Carey (1985) was based on the idea of conceptual change. Carey (1985, 1995) argued that children's understanding of the body and of health moves from a human, social perspective (i.e., you eat/wash because you are told to by your parents), to a more biological basis (i.e., you eat/wash to keep your body strong/well). Researchers supporting this approach have argued that that children do not have a full causal understanding of health and illness until ten years of age (Carey, 1985, 1995, Solomon, Johnson, Zaitchik, & Carey, 1996).

While both of these approaches may be more appropriate than stage approaches, they do not account for the fact that chronological age, access to appropriate information, and experience all play a part in children's level of illness understanding (Bird & Podmore, 1990, Charman & Chandiramini, 1995, Eiser, 1989). In addition the Piagetian stage model, schema model, and conceptual change model do not fit well with evidence that children can develop a biological understanding of health and illness younger than eight years old (e.g., Hatano & Inagaki, 1994, Hergenrather & Rabinowitz, 1991, Inagaki & Hatano, 1993, Paterson et al., 1999, Peltzer & Promotussanon, 2003).

Theories based around the idea of causal frameworks may explain when and how young children develop a full understanding health and illness better than the above models (Paterson et al., 1999). Researchers have argued that children possess a basic causal framework for understanding health and illness from an early age - as young as two years old (e.g. Keil, 1994, Simons & Keil, 1995, Wellman & Gelman, 1992). Researchers such as Wellman & Gelman (1992) and Morris, Taplin, and Gelman (2000) have argued that preschool children have the framework needed for a biological theory, and this framework is elaborated on with increasing experience.

What remains in question for supporters of the idea of causal frameworks is exactly how coherent young children's models of health and illness are (as the construction of a causal explanatory theory not only requires the learning of facts but also the co-ordination of these facts into a coherent system, Solomon & Cassimatis, 1999). Kalish (1999), amongst other researchers, considered this issue in relation to children's understanding of contamination and contagion, and argued that young children may have some form of biological model for these processes without understanding the specific details (i.e., young children can recognise that illness is a physical process, but may not fully understand all the bodily processes involved in illness). Kalish (1999) argued that young children are capable of holding a physical model of health and illness (i.e., infection caused by the physical transfer of materials), which forms the basis for the development of a differentiated biological model (i.e., agents of infection are understood as distinct types with a more detailed understanding of bodily processes).

In summary, work has shown that preschoolers and young children do have more understanding of health and illness than has been previously thought, but there are still limitations to their knowledge and their cognitive capacities (e.g., they may have a sophisticated understanding of the causes of illness, but a poor understanding of the bodily processes involved in illness and of the concept of time between cause and onset of symptoms, Williams & Binnie, 2002). There is still debate as to the extent and coherency of young children's understanding of health and illness. Researchers supporting the idea of causal frameworks have argued that young children's cognitive representations of illness are not qualitatively different from those of adults, although the content may be less mature and less detailed (Goldman, Whitney-Saltiel, Granger, & Rodin, 1991, Kalish, 1999).

2.4 Implications for the content development of the TedQL measure

This chapter considered whether children below eight years would be capable of answering items in QOL measures. We considered the evidence for young children's understanding of five concepts implicit behind many QOL items: emotions; mental representations; self; social comparisons; and biology (in relation to health and illness). These concepts are particularly relevant to QOL measurement as we have defined QOL as a multidimensional concept encompassing physical, psychological, social, and cognitive aspects of functioning, and also involving subjective judgements (i.e., the meaning and importance individuals place on aspects of their lives is relevant to their assessment of their own QOL, see Chapter 1, p. 8-9). Therefore our QOL measure based on the above definition of QOL could include items that require children to have knowledge of their own emotions; the existence of a sense of self; the ability to make social comparisons between themselves and their peers; and an understanding of their own physical and psychological functioning.

Emotions

Based on our review of the literature we felt that developing a QOL measure with items asking about their own emotions would be viable with children as young as four years of age. Our review showed that children understand the link between expectations and emotions and the appearance-reality distinction by four or five years old (see p. 22). The evidence we reviewed in this chapter showed that by two years old children can understand emotion words such as happy, sad, cross, and scared (see p. 21). Children's emotion vocabulary extends to include other emotion descriptors such as lonely, anxious, and proud at around five years of age (see p. 21). Therefore our measure could include items that asked about situations when they had felt happy or sad, and cross or angry.

Mental representations

Our review of the literature in relation to understanding mental representations revealed that four to five year old children would be capable of answering items about their thought processes (see p. 23-4). Researchers have found that two year olds are capable of understanding some aspects of thinking, and by three years they can make a clear distinction between the mental and physical environment (see p. 23). We found that there was consensus that children's understanding of mental states increases significantly during and after four years of age (see p. 23). Therefore we argued that including items asking children about their psychological and cognitive functioning would be viable with children as young as four years of age.

Self-concepts

We found evidence that children under eight years can understand different aspects of their self, and are capable of forming both positive and negative self-descriptors (see p. 25-6). Relating our review of the literature on children's development of self-concepts to the development of our QOL measure, we maintain that children below eight years would be capable of holding a sense of self (seep. 25-6). Therefore we could include items asking children about what sort of person they were in our QOL measure (e.g., whether they have a lot or a few friends, whether they like to boss their friends around, whether they are good or bad at activities such as running, climbing, or bike riding).

Social comparisons

Children need to be able to compare their own abilities and functioning to others around them to make a judgment on how good their own lives are. Social comparisons play an important part in this process, and are therefore particularly relevant to QOL. The age at which children are capable of making social comparisons will have implications for the development of our QOL measure. Our review of the literature revealed that by three and four years of age children can and do make comparisons between themselves and others (see p. 27). The main differences between the types of comparisons that younger and older children make is that younger children concentrate on similarities rather than differences, and younger children make comparisons of concrete entities, abilities, and possessions as opposed to internal attributes, thoughts, and feelings (see p. 27-8). We could include items that required comparisons to peers (e.g., making a judgment of how good they are at running or bike riding, or at writing and reading) in our QOL measure for children as young as three years of age.

Biology (in relation to health and illness)

Although there is still debate as to the extent and coherency of young children's understanding of biology (see p. 29-30), our review showed that preschoolers and young children may have more understanding of health and illness than has been previously thought (see p. 30). It was evident that by three to four years children can understand the physical aspects of illness and show some appreciation of the biological causes of sickness (i.e., germs and contagion, see p. 30-1). The implications of this work for our QOL measure were that children below eight years would be capable of answering simple items about their own everyday health (e.g., having tummy aches, or feeling too tired to play).

Based on our review of the developmental literature, we have taken an optimistic view of young children's capacities to self-report on their lives, abilities, thoughts, and feelings in this thesis. We maintain that attempting to develop a self-report measure for children below eight years would be a viable aim, and that children of this age can understand items that require them to self-report on their emotions, mental states, self-concepts, and health.

The following chapter (Chapter 3) reports the results of a review of the self-report measures currently available for use with children below eight years. This review provided context for the development of our new child QOL measure.

Chapter 3: Self-report measures for children aged three to eight years – a review.

Summary

We conducted a systematic review of self-report measures for children aged between three and eight years. Measures of QOL, self-esteem, self-concept, mental health, and pain measures were included. The aims were to compare the response scales and presentation styles used, to compare the ages that measures were targeted at, and to critically evaluate the item generation stages and the reliability, validity, and responsiveness data reported by authors. A search strategy was devised using appropriate keywords (e.g., child, children, scale, measure, self-report). One hundred and ninety-nine papers were included in the review. From these papers 105 measures were identified: pain (n=34), self-esteem/concept (n=32), QOL (n=22), mental health (n=17).

The response scales used were Likert, graphic, facial expression, and visual analogue, with Likert scales used most frequently (n=48). Items were read aloud to children (n=39), or presented as pictures (n=30) or in a written format (n=28), or presented using computers (n=5) or three-dimensional props (n=3). The measures were targeted at a variety of ages, with some developed for wide age ranges (e.g., 6-14 years) and others focusing on a narrow age range (e.g., 6-8 years). Items for measures were typically generated from the children themselves (e.g., using interviews or focus groups). Authors reported internal reliability data (n=57) more frequently than reproducibility data (n=38). Authors typically did not report all three types of validity. The most popular type of validity was construct validity – both convergent (n=50) and discriminant validity (n=39). The authors of four measures reported sensitivity data for their instruments.

Measurement issues that need to be addressed in developing and validating child self-report instruments are discussed. The results of our review have been used to guide the development of our child self-report measure.

3.1 Introduction

3.1.1 Systematic reviews

Systematic reviews are increasingly recommended in psychological and health science research. This methodology was originally pioneered in medical research. The value of reviewing and synthesising research becomes apparent when we consider the volume of studies that are published each year in both the medical and psychological literature (for example over two million articles are published annually in the biomedical literature in over 20,000 journals). The CRD Report (1996) gives a good explanation of the role of systematic reviews in research: "Systematic reviews are a scientific tool which can be used to summarise, appraise, and communicate the results and implications of otherwise unmanageable quantities of research".

Systematic reviews differ from other reviews in that they are based on strict criteria which should ensure the collection, inclusion, and consideration of *all* the available evidence rather than a selection of the published literature (CRAG, 1996). This review methodology also includes critical appraisal of the literature in attempt to: "weigh up the evidence critically to assess its validity and usefulness" (adapted from Sackett & Haynes, 1995).

Established guidelines for conducting systematic reviews have been developed by researchers (e.g., Clarke & Oman, 2000). There are ten phases which are normally recommended for systematic reviews, and these phases can be seen in Table 3.1.

Table 3.1 Stages involved in undertaking a systematic review

Phase	Purpose
0	Identification of the need for the review (show why this review is needed and that it has not been done before)
1	Problem specification/proposal (develop a well-built, structured, answerable question)

2	Review protocol (establish clearly the aims and scope of the review)
3	Identification of the literature (using literature searches)
4	Selection of studies (using clear inclusion and exclusion criteria)
5	Study quality assessment (using appraisal criteria)
6	Data extraction (extract key data from selected papers)
7	Data synthesis (bring all the findings, value and limitations from the selected studies together)
8	Report and recommendations
9	Getting evidence into practice

These recommended phases were followed in this review.

3.2 Review methodology

3.2.1 Identification of the need for the review (Phase 0)

The purpose of this review was to identify and critically evaluate the self-report measures available for children aged between three and eight years. These included QOL, self-esteem, self-concept, mental health, and pain measures. The Cochrane database was searched to identify any systematic reviews that had considered this topic. One review was identified: Quality-of-life measures in chronic diseases of childhood (Eiser & Morse, 2001). This Health Technology Assessment (HTA) report identified QOL measures for children aged from 0-18 years that had been published between 1980 and 2000. Both child and proxy (e.g., parent or health professional) report measures were included. Literature reviews of QOL (Annett, 2001, Bullinger & Ravens-Sieberer, 1995, Connolly & Johnson, 1999, Eiser, Mohay, & Morse, 2000, Gill & Feinstein, 1994, Garratt, Schmidt, Mackintosh, & Fitzpatrick, 2002), self-esteem/concept (Davis-Kean & Sandler, 2001), and pain (Erickson, 1990) measures were also identified from other databases. Despite the identification of these papers, the need for our review was justified as follows:

i. QOL, self-esteem, self-concept, mental health, and pain child self-report measures had not been reviewed using the *systematic* review methodology

ii. no review had focused specifically on *child self-report* measures for children aged between three and eight years

iii. no current review had collectively evaluated, or attempted to integrate, the methodologies used in developing QOL, self-esteem, self-concept, mental health, and pain child measures.

3.2.2 Problem specification (Phase 1) and review protocol (Phase 2)

This review focused on identifying self-report measures for children aged between three and eight years. QOL self-report measures currently available for this age group were considered. Measures of health status, functional status, and well-being were also included on the grounds these terms have been used interchangeably with QOL (Jenney et al., 1995). Measures of self-esteem and self-concept, mental health, and pain were also included. The rationale for the inclusion of these related concepts were:

i. the overlap in theoretical conceptualisation (for example, QOL and self-esteem/self-concept measures often include similar domains and items. A QOL measure may include ratings of family and peer relationships, and a self-esteem measure may also include such domains. QOL can be defined as how 'good your life and functioning is', and self-esteem may be defined as 'how you think about yourself and your abilities')

ii. comparison of the methodologies employed in different areas developing self-report instruments for children in this age group.

The aims were to:

i. compare the response scales used
ii. compare the presentation styles used
iii. compare the age groups at which the measures have been targeted (e.g., whether authors have developed measures for wide or narrow age ranges)
iv. critically evaluate the item generation stages
v. critically evaluate the reliability, validity, and responsiveness data reported.

3.2.3 Identification of the literature (Phase 3) and selection of studies (Phase 4)

Search procedure and strategy

The following databases were searched from 1970 to September 2003:

- PsycINFO via WEBSPIRS (formerly known as PsycLIT)
- MEDLINE via WEBSPIRS
- ISI Web of Science
- CCTR
- Embase
- ERIC-AT Test locator (URL: http://www.ericae.net/testcol.htm)

Internet search engines were used to identify links to any additional databases or research by entering keywords. The following websites were identified:

- American Thoracic Society – Quality of Life (http://www.atsqol.org/)
- Australian Centre on Quality of life (http://acqol.deakin.edu.au/index.htm)
- International Society for Quality of Life Research (http://www.isoqol.org)
- Online Guide for Quality of Life Assessment http://www.olga-qol.com/)
- Quality of Life Instruments Database, version 1.8 (http://www.qolid.org/)

The following search strategy was devised, and applied using the appropriate keywords and logical operators specified by each database:

1. (child) or (child*) or (children) or (children*)
2. (measure*) or (scale) or (index)
3. 1 and 2
4. (self-report) or (self report) or (self assessment) or (self-assessment) or (child* report)
5. 3 and 4

Search 3 produced general papers on measures used with children. This result was narrowed to identify papers which had developed or used *self-report* measures for children (search 5).

Abstracts were screened to assess the relevance of studies, and papers that clearly met the exclusion criteria were excluded at this point. Where abstracts were ambiguous, the full papers were obtained. All other references were downloaded into Endnote (Macintosh, version 5) and the papers were retrieved. Review articles were used to help identify and obtain additional papers. Authors of key papers were also contacted personally to enquire on additional work that may have been unpublished, in press, or in preparation.

Inclusion and exclusion criteria

The inclusion and exclusion criteria were as follows:

Inclusion criteria

- required age for self-report was between three and eight years
- measures of: QOL, self-esteem, self-concept, mental health, or pain
- written in English

Exclusion criteria

- measure targeted only at children over eight years
- measure developed solely for proxy report of the child's functioning (e.g., parent, teacher or nurse report)
- review article on self-report measures, or comments on the measurement of self-report in children.

3.2.4 Study quality assessment (Phase 5) and data extraction (Phase 6)

The measures from the included papers were compared using the checklist below designed by the author.

Content and format of measures

i. response scales (i.e., Likert, graphic, facial, or visual analogue response scales)
ii. justification/evidence for choice of response scale
iii. presentation styles (i.e., verbal, written, pictorial, props, or computerised)
iv. age ranges (i.e., narrow or wide age range)

Quality of measures

i. item generation methods (i.e., how the content of items were developed e.g., from literature reviews, existing child measures, existing adolescent or adult measures, expert panels, or from children themselves)
ii. reliability data reported
iii. validity data reported
iv. sensitivity data reported.

Key data were extracted from each paper and summarised on a data extraction form. (See Figure A1 in Appendix A for example of data extraction form). The data for each of the measures were also summarised into word tables, which enabled comparisons across the different areas (i.e., QOL, self-esteem/concept, mental health and pain).

Psychometric issues involved in evaluating measures

The most frequently cited requirements for self-report measures relate to their psychometric properties, i.e., whether responses are reliable (that children will respond the same way on different occasions), valid (that measures are assessing what they say they are), and responsive (that scores can detect changes over time). There are a variety of ways to assess the reliability, validity, and responsiveness of any given measure.

Reliability can be assessed in two ways. First, the internal consistency of children's responses (internal reliability) can be assessed. This form of reliability is concerned with whether the items are all tapping the same construct (Bryant, 2000). This can be assessed by calculating the correlations between items in a measure, and the higher the correlations between items the greater the internal consistency. An internal consistency of above .70 has been recommended as a guideline for 'good' levels of internal consistency (Nunnally, 1978, Cronbach, 1951, see Study 1, p. 107).

Second, the reproducibility of children's responses (test-retest/temporal reliability) can be assessed. This form of reliability is concerned with whether a measure produces consistent responses over time. This can be assessed by calculating whether children's scores on a given measure are positively correlated at two different times. Correlation coefficients are used to calculate the level of agreement between scores over time, and these estimates must reach above the recommended criterion standard of .60 to provide evidence for reproducibility over time (Juniper, Guyatt, Streiner, & King, 1997, see Study 4, p. 227-8).

It is also necessary to assess validity as measures can be reliable but not valid (e.g., respondent's scores on an instrument can be consistent over time, but the scores may have no relationship with the concept the measure was intended to measure, Bryant, 2000). Establishing the validity of a measure is a gradual process involving the accumulation of evidence from a variety of sources. Bryant (2000) defined the validity of a measure as how thoroughly (content validity) and accurately (construct validity) it measures a theoretical concept, and how useful it is at predicting outcomes (criterion validity). Measures can only be judged as either valid or invalid, however the evidence supporting the validity of a measure may be viewed as weak or strong (Bryant, 2000).

There are three main components to validity. First, content validity is concerned with the plausibility, breadth, and depth of items in a measure (i.e., do items cover all the relevant aspects of the concept being measured?). It is a subjective judgement can be made by researchers themselves, respondents, or proxies. Content validity can be shown when an expert panel (e.g., parents or psychologists) rate the depth and breadth of items in an instrument. Researchers have also evaluated content validity using statistical procedures (such as exploratory factor analysis) to determine the domains that a measure assesses and identify how strongly the content defines each domain (Bryant & Yarnold, 1995). These techniques include principal-components analysis (PCA) and confirmatory factor analysis (CFA).

Second, criterion validity is concerned with how accurately an instrument predicts a well-accepted indicator of a given concept or criterion (i.e., how well test scores provide a current estimate of a relevant external outcome, and/or predicts future functioning on this outcome, Bryant, 2000). Criterion validity can be divided into predictive (future), concurrent (current), and retrospective (past) validity. Establishing predictive validity requires a longitudinal design to assess the relationship between initial tests scores (e.g., self-motivation) and a criterion measure collected at a later time (e.g., career success). Concurrent validity concerns whether tests scores are correlated to a criterion measure assessed at the same time, and retrospective validity relates to whether test scores are correlated to a criterion from the past (e.g., recollections of previous experiences). Both concurrent and retrospective validity provide weaker evidence of the validity of a measure when compared to prospective validity (Bryant, 2000). Concurrent validity is weaker because the relationship between the criterion and the test scores may be artificially inflated due to respondent's desires to answer consistently (as both the test scores and criterion scores are collected at the same time point, Cook & Campbell, 1979). Retrospective validity is weaker because an individual's knowledge of the present can distort their recall of the past.

Third, construct validity is the most difficult aspect of validity to establish and is concerned with whether a measure actually assesses the underlying construct that it is intended to measure (i.e., the match between the underlying construct and the operational definition used for an instrument, Bryant, 2000). The first aspect of establishing construct validity requires researchers to specify the specific components of the construct being measured and distinguish it from other related but separate constructs. A related aspect is face validity which has been defined as the degree to which a measure 'appears' to measure what it is intended to measure (Bryant, 2000). Face validity can be assessed by asking respondents to indicate how appropriate they feel items are to the concept being measured. Evidence for this can also be shown when an expert panel agree that items are asking what they claim to measure.

Construct validity can be dived into two categories – discriminant and convergent validity. Convergent validity is concerned with the degree to which multiple measures of the same or related constructs show convergence or agreement (i.e., test scores should correlate highly with scores from other related measures). Discriminant validity is usually assessed at the same time, and relates to the degree to which multiple measures of different concepts are distinct from each other (i.e., scores on measures assessing different constructs should not be correlated). Another means of assessing discriminant validity is to evaluate whether an instrument can discriminate between groups of individuals known to differ on an accepted criterion (e.g., does a measure of mental health discriminate between clinically depressed and normal controls?). This aspect of validity can also be termed clinical validation when the criterion groups consist of individuals with psychological disorders (such as depression or schizophrenia, Bryant, 2000).

Another important aspect of a measure is whether it is sensitive to changes over time, which has been termed 'responsiveness' or 'sensitivity to change'. While it is important to provide evidence that a measure can produce reproducible scores over a short time period (i.e., test-retest reliability, see p. 43), a measure also needs to be able to detect important changes over time (Terwee, Dekker, Wiersinga, Prummel, & Bossuyt, 2003). For example, a pain measure should be able to pick up a change in pain levels before and after anaesthesia, or a depression measure should show a difference in ratings before and after an intervention or therapy sessions. It is also important that measures can detect changes that are meaningful and important to individuals (e.g., asking questions like 'is a change of 5 to 10 points equally meaningful as a change of 10 to 15 points?' Terwee et al., 2003).

3.3 Review results

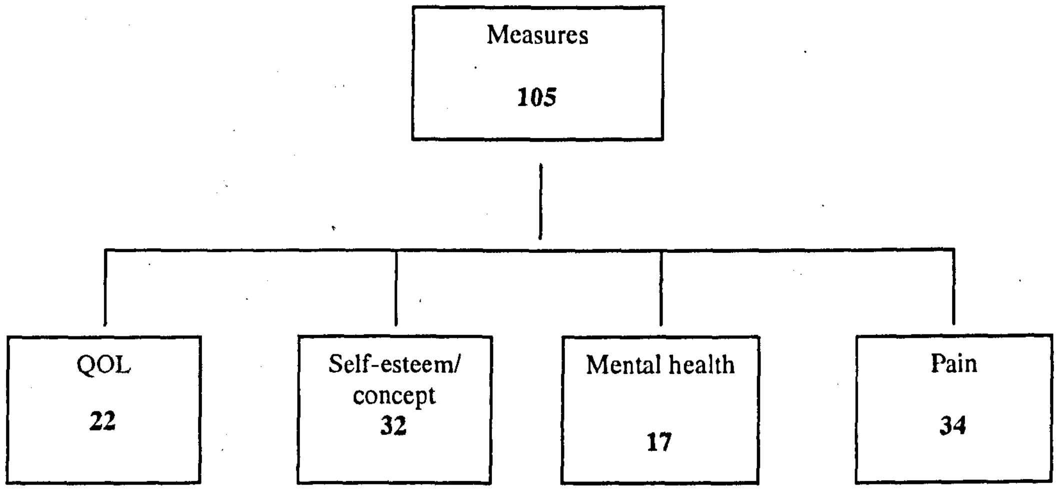
As a result of the initial screening of the abstracts, 259 potentially relevant papers were identified and obtained. Ninety papers were excluded due to operation of exclusion criteria. (See Table A1 in Appendix A for excluded papers). An additional 30 references were obtained from requests for articles that were unpublished, in press or in preparation.

Therefore 199 papers were included in the review, and 105 measures were identified from these papers (i.e., as some measures were reported in more than one paper). These measures and their psychometric properties have been summarised alphabetically in Table 3.2 (at the end of this chapter, see p. 75-90). Each measure was given a number, so that measures could be referred to by their measure number in square brackets (for example [1] is the About my Asthma questionnaire, Mishoe, Baker, Poole, Harrell, Arant, & Rupp, 1998).

Of the included papers 131 were published in the US and Canada, 43 were published in the UK, and 25 were published elsewhere (Australia and New Zealand: n=5, Netherlands: n=10, Sweden: n=2, Germany: n=8).

The included measures were divided into the different areas identified. Assessment of QOL included measures of QOL, health status, functional status, and well-being. Measurement of children's self-esteem and their self-concepts included instruments measuring academic, physical, social, and emotional competence. Mental health assessment included measures of fear, anxiety, and depression. More self-esteem/concept (n=32) and pain (n=34) measures were identified for children below eight years, compared to the number of QOL measures (n=22) and mental health measures (n=17) identified for this age group (see Figure 3.1). The number of domains (i.e., sub-groups of items, e.g., psychological functioning, cognitive functioning) included in measures ranged from one (various measures) to nine [7]. The total number of items ranged from one (various pain measures) to 137 [65].

Figure 3.1: Breakdown of the number of measures included by area



3.3.1 Data synthesis (Phase 7)

Content and format of measures

a. Response scales

The scales chosen to represent response choices of items to children were classified into four categories:

i. Likert (i.e., written linear scale anchored at various points with numbers and/or words)
ii. Graphic (i.e., three-dimensional or pictorial scale with/without word/number anchors)
iii. Facial expression (i.e., pictorial linear scale anchored at various points with cartoon or photographic faces)
iv. Visual analogue (i.e., visual linear scale anchored at each end with numbers).

The most commonly used scale type was Likert response scales (n=48), with 15 QOL and 21 self-esteem/concept measures using Likert scales. Twenty-eight used graphic response scales, and 15 of these were pain measures. Twenty-three measures used facial expression scales, and 11 of these were assessing pain. The least commonly used scale type was visual analogue scales, with six measures employing visual analogue scales. (See Figure A2 in Appendix A for breakdown of measures by scale type, see Figures 3.5 - 3.13 at end of chapter for examples of each scale, p. 91-2).

b. Justification/evidence for choice of response scale

There were a variety of arguments put forward by authors as to the value of each different type of scale for use in child self-report measures. Given the variety of response scales that have been used with young children, and the arguments for and against the different types, it was evident that authors needed to provide justification for their choice of scales.

However it was not always clear how authors made their choice of response scale type, and whether this choice was based on evidence. The authors of twenty-four measures (23%) provided justification for their scale choice, and the reasons they gave can be summarised as:

i. comparing the psychometric properties of similar (or the same) measures across different scales
ii. directly testing children's understanding of a given response scale
iii. using a response scale from an existing child measure
iv. modification of the response scale from an adolescent self-report measure
v. basing their scale choice on children's own preferences.

These reasons are discussed below.

Comparing the psychometric properties of measures across different response scales

The authors of seven measures [5, 73, 79, 81, 86, 94, 97] compared the psychometric properties of the same measure across different response scales, to provide evidence for the effectiveness of their chosen type (i.e., to assess which scale produced the most consistent and reliable responses from children). Such testing allowed a judgement of the relative value of response scale types to be made. For example, Champion, Goodenough, Wu, Chua, Taplin, and Ziegler (2000) compared children's ratings of their pain across six scales: the Adjectival Rating Scale [73], the Coloured Analogue Scale [79], the Faces Pain Scale [81], the Finger Span [86], the Poker Chip Tool [94], and the Sydney Animated Facial Expression Scale [97]. These researchers considered whether children's scores were correlated across these scales to assess their equivalence (Champion et al., 2000).

Testing children's understanding of a given response scale

The appropriateness of a scale can also be judged by testing children's understanding of the chosen response scale. This can be achieved using hypothetical questions (e.g., asking a question where the child's answer can be predicted and seeing whether they chose an appropriate point on the scale) or by requiring children to rate their answer twice (e.g., getting them to rate their answer once on the chosen response scale and then on an equivalent scale, and seeing whether their choices are consistent). Such testing was incorporated by authors in the development of six measures [3, 7, 12, 25, 28, 66]. An example of the use of this technique was provided by Quittner, Sweeney, Watrous, Munzenberger, Bears, Nitza, Fisher, and Arcos (2000) when developing the Cystic Fibrosis Questionnaire – Child version [7]. These researchers required children to choose their response on a Likert scale and then to indicate their response again on an additional thermometer scale (Quittner et al., 2000).

Using response scales from existing child measures

The authors of two QOL [1, 6] and four self-esteem/concept measures [23, 33, 36, 47] reported that their response scales were taken from existing child self-report measures. For example, Measelle, Ablow, Cowan, and Cowan (1998) presented their response options using a similar graphic bipolar response scale to Harter and Pike's (1984) measure [40]. Measelle et al. (1998) argued that this helped maximise children's comprehension of the scale and minimise socially desirable responding. However basing the scale choice for a new measure on a scale used previously may not be helpful, as the authors of the previous measure may not have had any justification for their choice of scale type in the first place.

Modification of the response scale from adolescent measures

Four QOL measures were a downward extension of existing adolescent measures [5, 14, 18, 19]. In three of these the authors used the same response scale with the younger children in a simplified form, for example the 5-point Likert response scale in the KINDL was reduced to 3-points for the Kiddy-KINDL [14] (Ravens-Sieberer & Bullinger, 1998).

Basing response scale choice on children's preferences

One set of researchers used children's own preferences to direct their choice of response scale. Rebok et al. (2001) asked children whether they preferred a linear or circular scale for answering items in their measure [5]. They found 74% of children preferred the circular scale and these children also reported that it helped them to understand the scale (Rebok et al., 2001).

c. Presentation styles

The styles that authors chose to present their items to children were classified into five categories:

i. Verbal (i.e., interviewer reads written items aloud to child)
ii. Pictorial (i.e., items presented to child using visual aids, such as cartoons, drawings or photographs)
iii. Written (i.e., child reads items from a questionnaire)
iv. Props (i.e., items presented by interviewer verbally using three-dimensional aids, such as puppets, teddy bears, dolls)
v. Computerised (i.e., items presented to child with the aid of a computer).

(See Figure A3 in Appendix A for a breakdown of measures by presentation styles).

Verbal presentation

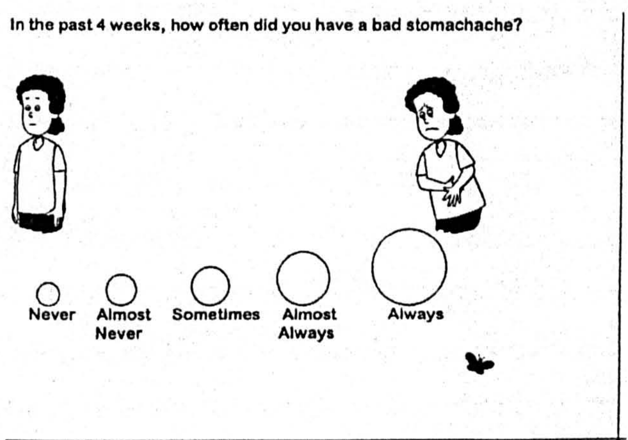
In 39 of the measures authors chose to read the items aloud to children, and 32 of these assessed pain. Three QOL, two self-esteem/concept, and two mental health measures required items to be read aloud. This preference for verbal presentation in pain measures may have been related to the fact that many of these instruments only involve one item (i.e., 'how much pain are you feeling?). Verbal presentation does have the advantage that such measures can be given to pre-literate children. Some researchers have argued that asking children direct questions is the best way to obtain reliable information from young children (e.g., Marsh et al., 1998).

Pictorial aids

Presenting items in a pictorial format (either by cartoon or photographs) was used in 30 measures, and 20 of these were self-esteem/concept measures. In 28 of these 30 measures the pictures were presented as cartoons. The authors of the other two measures [24, 29] used photographs of children to illustrate items. These cartoons and photographs were presented as a visual aids to children to help them understand and interpret the items (Harter & Pike, 1984). Cartoons and pictures can have an advantage over photographs as photographs might imply a specific person whereas cartoons can be less personal.

Three QOL measures [5, 12, 21] used cartoons to support their items. Figure 3.2 shows an example of an item from one of these; the Child Health and Illness Profile – Child Edition [5] with the cartoon drawings at each side help for an item asking whether they have had stomach aches in the last four weeks.

Figure 3.2: Cartoon from the Child Health and Illness Profile – Child Edition [5] measure



Written presentation

In 28 of the measures the items were presented in a written format, with 13 of these assessing QOL. For this presentation style the children are required to read items themselves and answer the items as a questionnaire. The use of written items alone has the disadvantage that the measure can only be targeted at children who are capable of reading them. The majority of the measures using written items were developed for children over six years of age (n=18). However ten measures were identified where authors used written items with children younger than six years. The authors of one self-concept [27] and mental health [70] measure stated that their items were appropriate for children as young as three years (Boger & Knight, 1969, Eccles, Wigfield, Harold, & Blumenfeld, 1993). However it is possible that children of such a young age could not read and understand items.

Computerised presentation

Five measures employed computers to present items to children. These were three QOL measures [2,10,11], one mental health [59], and one pain measure [97]. These all used animated cartoons or faces to illustrate items to children. Children used a mouse (or a touch screen) to make their response choices.

Props as three-dimensional visual aids

Using props (n=3) to present items to children has been less common. The three measures that used props as aids for presenting items were all self-concept measures [23, 45, 50]. Authors used hand puppets to 'talk' to children and illustrate items to them. These measures have been used successfully with children as young as three years. The Berkeley Puppet Interview (BPI) [23] has been used with four to eight year old children. Measelle et al. (1998) provided evidence for the reliability, sensitivity, and validity of the BPI measure (e.g., convergence between child, parent, and teacher reports). The Self-Interview [50] has been used with children aged three to eight years, and demonstrated acceptable levels of internal consistency and reproducibility in children's responses (Eder, 1990). Verschueren, Buyck, and Marcoen (2001) provided evidence for the predictive validity of the Puppet Interview [45] in relation to children's self-representations at five years of age. Verscheuren et al. (2001) found that children with positive self-representations at five years showed higher acceptance by peers and higher global self-worth at eight years of age.

d. Age ranges

Although the majority of authors targeted their measures at children aged from four to ten years, there was some variation in age ranges. The largest age range was found in a pain measure [104] (3-18 years), whereas measures such as the Kiddy-KINDL [14], I Feel-Me Feel [31], and the Vertical Scale [100] have been targeted at narrower age ranges, 4-7 years [14], 3-6 years [31], and 3-8 years [100].

Despite this variation some patterns were evident when considering different types of measures (i.e., QOL, self-esteem/concept, mental health, and pain). Over half of the QOL measures were targeted for use with children aged six years and older (n=12). Measures designed to assess self-esteem/concepts and pain in children were generally developed for lower ages than QOL measures. Eighteen of the self-esteem/concept and 28 of the pain measures have been targeted at children as young as three and four years.

Quality of measures

The relevant issues are what makes a 'good' child self-report measure, and how can such a judgment be made. There are a variety of qualities that are considered to make a 'good' measure (e.g., reporting reliability, validity, and sensitivity data). There are also other characteristics which researchers have used as indictors of quality: such as availability of proxy (i.e., parent or teacher) and child forms, containing age-sensitive versions, and having a theoretical basis (e.g., Eiser & Morse, 2001). In this review we used four criteria to judge the quality of the instruments: a) the item generation methods, b) the reliability data, c) the validity data, and d) the sensitivity/responsiveness data reported by authors.

a. Item generation methods

The way authors developed the content of their measures can impact directly on the face validity of these instruments. For 58 of the measures (55%), the item generation methods were unclear as the authors did not report the origin of their items. This omission did not necessarily mean the authors had no rationale or justification for the content of their measures, however it was not reported in their papers.

For 47 of the measures (45%), authors did report how they developed the content of their measures. These were categorised into four main techniques. Items were either:

i. generated by the children themselves (i.e., from interviews or pilot work with children)
ii. modified from existing measures (i.e., items adapted or altered from child or adolescent measures)
iii. generated and rated by an expert panel (i.e., by health professionals, clinicians, or psychologists)
iv. generated by researchers themselves (i.e., by researchers as a result of a literature search).

(See Table A2 in Appendix A for a summary of item generation methods. Note: Some authors used more than one item generation method).

Items generated/altered by the children themselves

The authors of 29 measures reported using information from children themselves to inform the item content. Researchers have used techniques such as interviewing children and asking them to list, or talk about, relevant topics [e.g., 5, 50]. Other authors reported using children's spontaneous comments from pilot work specifically to develop the vocabulary and wording of their items [e.g., 50, 72, 75, 104]. For example, Abu-Saad (1990) used children's word descriptors of pain from their piloting develop their items for the Abu-Saad Paediatric Pain Assessment Tool [72] by sorting them into categories and themes. The authors of 24 measures reported piloting their items on children to assess the acceptability and appropriateness of the language and wording chosen (see Appendix A, Table A2). For example, Lewis-Jones & Finlay (1995) altered the wording of their items in the Children's Dermatology Life Quality Index [4] to improve clarity following piloting. Bridgeman and Shipman (1978) used pilot work with children to identify unfamiliar vocabulary in the Brown IDS Self-Concepts Referents Test [24]. Edelsohn, Ialongo, Werthamer-Larsson, Crockett, and Kellam (1992) added symbols to help explain the items in the Children's Depression Inventory [56] as a result of pilot work which showed younger children (6-7 years) had difficulty understanding some of the language used in their measure.

Items modified from existing measures

The authors of 15 measures reported adapting items from adolescent or child measures. In these instances authors usually adapted a measure developed for an older age group for use with younger children (e.g., Beyer & Arandine, 1988, Christie et al., 1993, Hughes & Leatherman, 1982). This approach can be useful in providing ready-made items for researchers, however it may be inappropriate to adapt items developed for older children. For example, in relation to measuring QOL or self-esteem, there are differences in what are considered 'normal' developmental goals in relation to skills acquisition (La Greca, 1990), and issues in eight year olds' lives may differ from those in sixteen year olds' lives (Rosenbaum, Cadman, & Kirpalani, 1990). In addition young children may not able to understand the concepts implicit in items that have been used successfully with older children due to their cognitive development (Shahinfar et al., 2000). As discussed in Chapter 2 (see p. 20-30), researchers need to consider children's developmental stage when choosing and developing items for their instruments.

Items generated and rated by an expert panel

The authors of 14 measures used an expert panel to generate and rate a pool of items as appropriate and relevant for the target group of children. Such panels usually consisted of health professionals, clinicians, and/or psychologists. While this may be a useful way to establish the content validity for items in measures, and a source of experienced opinions, adults may not be able to identify all the issues important and relevant to children. As discussed in Chapter 1 (see p. 5-6), adults may not be aware of the issues in young children's lives, and children may also be adept at hiding their thoughts and feelings from adults. We also discussed the evidence for a lack of concordance between parent and child reports of QOL in Chapter 1 (see p. 5-6) and Chapter 4 (see p. 98).

Items generated by researchers based on literature searches

For nine measures authors generated the content of their items themselves basing items on the results of a literature search of relevant papers and existing instruments (e.g., Eiser et al., 2000, Perez, 1982, Valla, 2000, Varni et al., 1998). Some researchers have argued that this method enabled them to identify relevant issues for young children (e.g., Eiser et al., 2000).

Combining item generation techniques

The authors of 16 measures used a combination of methods to develop the item content of their measures. An example is the Pediatric Rhinoconjunctivitis Quality of Life Questionnaire [17]. Juniper, Guyatt, Epstein, Ferrie, Jaeschke, and Hiller (1992) generated a list of possible items from existing adolescent and adult measures and a review of the literature, and then altered, deleted, and added items as a result of discussions with adult clinicians and children themselves.

b. Reliability, validity, and responsiveness data reported

There are three main ways to assess the psychometric properties of self-report measures (i.e., their reliability, validity, and responsiveness). We compared the psychometric properties reported by authors across the instruments included in this review.

Reliability data reported –

Internal consistency

Table 3.3: Summary of internal reliability statistics used by authors

Area	N of measures	Internal reliability data (n)					Range of internal consistency values
		α	Hoyt	K_{20}	Split -half	No data	
QOL	22	14	1	1	0	6	0.46 – 0.97
Self-esteem/concept	32	21	0	2	3	6	0.36 – 0.92
Mental health	17	7	0	2	0	8	0.54 – 0.89
Pain	34	6	0	0	0	28	0.54 – 0.74

The authors of 57 measures (54%) assessed the internal reliability of children's scores on their instruments, and reported reliability coefficients in their papers. Our review showed that authors have used a variety of coefficient values to assess the internal consistency of children's scores on their measures (e.g., Cronbach's alpha α, Kuder-Richardson K_{20}, Split-Half). As shown by Table 3.3, the most commonly used statistic for assessing internal consistency was the Cronbach's alpha coefficient (n=48). The range of values for internal reliability were comparable across mental health and pain measures (see Table 3.3). The range of values were also comparable across QOL and self-esteem/concept measures (see Table 3.3). The lowest internal consistency value was reported by McDowell and Lindholm (1986) for a self-concept measure - the Primary Self-Concept Inventory [44] (α = .36, see Table 3.2 at end of chapter).

Table 3.4: Measures with internal consistency data by scale type

Type of scale used in measure	N of measures	Internal reliability values reported % (n)	Values above .70 standard % (n)
Likert	48	71 (34)	65 (31)
Graphic	28	43 (12)	32 (9)
Facial expression	23	30 (7)	22 (5)
Visual analogue	6	66 (4)	50 (3)

We compared the internal reliability data across scale and presentation type used by authors. First, in relation to scale type, the authors of nearly three-quarters of measures using Likert scales (71%) reported internal reliability data (see Table 3.4). For two-thirds of these measures the values were above the recommended value of .70. The authors of two-thirds of the measures (66%) using visual analogue scales reported reliability values, and of these half were above .70 (see Table 3.4). The authors of two-fifths of measures (43%) using graphic scales reported reliability values, and one-third of these were above the .70 standard (see Table 3.4). One-third of the measures employing facial expression scales reported internal consistency data for children's responses, and five of these were above .70 (see Table 3.4).

Table 3.5: Measures with internal consistency data by presentation type

Type of presentation style	N of measures	Internal reliability values reported % (n)	Values above .70 standard % (n)
Verbal	39	21 (8)	13 (5)
Pictorial	30	83 (25)	73 (22)
Written	28	72 (20)	64 (18)
Computerised	5	40 (2)	20 (1)
Props	3	66 (2)	66 (2)

Second, in relation to presentation type, 83% of the pictorial measures provided information on the internal consistency of children's responses to items (see Table 3.5). Of these pictorial measures three-quarters (73%) reported values above the recommended value of .70. The authors of 72% of the written measures reported reliability data, and two-thirds reported consistency values above .70 (see Table 3.5). The authors of one-fifth of the measures (21%) using verbal presentation of items reported internal reliability data. 13% of these instruments the reliability values were above .70 (see Table 3.5). The authors of two (out of the three) measures [23, 50] using props to present their items to children reported internal consistency data, and for both of these measures the consistency values were above .70 (see Table 3.4). The authors of two (out of the five) measures [11, 59] using computerised presentation reported internal consistency values, and these values were above .70 for one of these instruments [59].

Test-retest reliability (reproducibility)

Table 3.6: Summary of reproducibility statistics used by authors

Area	N of measures	Test-retest reliability data (n) ρ	ρ±	No data	Range of internal consistency values
QOL	22	6	5	11	0.35 – 0.93
Self-esteem/concept	32	13	0	19	0.38 – 0.94
Mental health	17	8	2	7	0.39 – 0.89
Pain	34	4	0	30	0.35 – 0.92

The authors of 38 measures (36%) assessed the reproducibility of the children's scores on their instruments. Authors used two types of coefficient statistics to calculate test-retest reliability – either Spearman's correlation coefficients (ρ) or Intra-class coefficients (ρ±). Overall researchers reported internal reliability data for their measures more frequently than test-retest reliability data. This was shown by the fact that the authors of 38 measures reported reproducibility data (see Table 3.6), compared to 57 measures with internal reliability data (see Table 3.3).

Table 3.7: Measures with reproducibility data by scale type

Type of scale used in measure	N of measures	Test-retest reliability values reported % (n)	Values above .60 standard % (n)
Likert	48	54 (26)	50 (24)
Graphic	28	21 (6)	21 (6)
Facial expression	23	17 (4)	17 (4)
Visual analogue	6	33 (2)	33 (2)

The authors half of the measures using Likert scales (54%) reported test-retest reliability data for their measures, and of these 50% were above the recommended value of .60 (see Table 3.7). The authors of one third (two out of six) of measures employing visual analogue scales reported test-retest reliability data, and all of these were above 0.60 (see Table 3.7). The authors of one-fifth of measures (21%) using graphic scales reported reproducibility data for their instruments, and all of these were above the .60 standard (see Table 3.7). The authors of 17% of the measures using facial expression scales gave reproducibility data for their instruments – all above .60 (see Table 3.7).

Table 3.8: Measures with reproducibility data by presentation type

Type of presentation style	N of measures	Test-retest reliability values reported (n)	Values above .60 standard (n)
Verbal	39	3 (5)	3 (5)
Pictorial	30	43 (13)	43 (13)
Written	28	64 (18)	57 (16)
Computerised	5	20 (1)	20 (1)
Props	3	33 (1)	33 (1)

For presentation style, the authors of two-thirds of written measures (64%) reported test-retest reliability data for their instruments, and 57% were above the recommended value of .60 (see Table 3.8). For two-fifths of the measures (43%) using pictorial format the authors reported reproducibility data, and all of these were above .60 (see Table 3.7). For measures using props or computerised presentation for items, two authors reported test-retest reliability data with values above .60 (see Table 3.8). The authors 3% of measures using verbal presentation provided reproducibility data – all above .60 (see Table 3.8).

Validity data reported -

Content validity

Table 3.9: Measures with content validity data by scale type

Type of scale used in measure	N of measures	Content validity data reported % (n)
Likert	48	25 (12)
Graphic	28	14 (4)
Facial expression	23	22 (5)
Visual analogue	6	17 (1)

Table 3.10: Measures with content validity data by presentation type

Type of presentation style	N of measures	Content validity data reported % (n)
Verbal	39	13 (5)
Pictorial	30	27 (8)
Written	28	29 (8)
Computerised	5	0 (0)
Props	3	33 (1)

As shown by Tables 3.9 and 3.10, the authors of 22 measures (21%) attempted to establish content validity. Authors of these instruments used three techniques for establishing content validity. Items were either rated for relevance, depth, or breadth using an expert panel usually consisting of psychologists or other health professionals (n=3); or the content validity of items was supported by information obtained directly from children themselves (n=6); or established using statistical methods (such as factor analysis or PCA, n=13).

Criterion validity (concurrent, predictive, and retrospective)

No authors attempted to assess the retrospective validity of their measures using outcome measures that involve recollection of experiences. Authors of one measure attempted to assess the predictive validity. Verschueren et al. (2001) provided support for the predictive validity of the Puppet Interview [45] by showing that children scores on this measure at 5 years of age were related to various outcome measures at 8 years of age (e.g., acceptance by peers, global self-worth, teacher-rated independence).

Table 3.11: Measures with concurrent validity data by scale type

Type of scale used in measure	N of measures	Concurrent validity data reported % (n)
Likert	48	6 (3)
Graphic	28	7 (2)
Facial expression	23	13 (3)
Visual analogue	6	17 (1)

Table 3.12: Measures with concurrent validity data by presentation type

Type of presentation style	N of measures	Concurrent validity data reported % (n)
Verbal	39	15 (6)
Pictorial	30	3 (1)
Written	28	4 (1)
Computerised	5	0 (0)
Props	3	33 (1)

As shown by Tables 3.11 and 3.12, the authors of eight measures (8%) attempted to establish concurrent validity. Authors of these instruments used either observed behaviour (n=7); or diary symptom scores (n=1); or achievement tests (n=1) as outcome measures for comparisons to scores on their measures.

Construct validity (convergent and discriminant)

Convergent validity

The authors of 50 measures (48%) assessed the convergent validity of their instruments. Convergent validity was assessed by comparing scores on a given measure to scores on other measures also hypothesised to be measuring the same construct. Of the 50 instruments with convergent validity data, 22 of these were pain measures. Authors of pain measures generally placed more emphasis on validity (convergent) than reliability data (Beyer & Knapp, 1986). This may be because reliability is harder to assess for pain measures compared to QOL or self-esteem/concept measures. This could be for two main reasons. First, as many pain measures consist of only one single item, assessment of internal consistency may not always be possible as there may not be enough items to make up a 'scale' as such (Erickson, 1990). Second, pain is a state that often varies dramatically over short time periods and therefore ratings should not be expected to be the same if measured over different time points (Erickson, 1990).

Table 3.13: Measures with convergent validity data by scale type

Type of scale used in measure	N of measures	Convergent validity data reported % (n)
Likert	48	46 (22)
Graphic	28	39 (11)
Facial expression	23	61 (14)
Visual analogue	6	50 (3)

In relation to scale type, two-thirds of the authors of measures using facial expression scales provided convergent validity data (61%, see Table 3.13). Half of the authors of visual analogue measures investigated the convergent validity of their instruments (50%, see Table 3.13). Just under half of the authors of measures using Likert and graphic scales reported convergent validity data (46% & 39%, see Table 3.13).

Table 3.14: Measures with convergent validity data by presentation type

Type of presentation style	N of measures	Convergent validity data reported % (n)
Verbal	39	62 (24)
Pictorial	30	47 (14)
Written	28	39 (11)
Computerised	5	20 (1)
Props	3	0 (0)

In relation to presentation style, two-thirds of the authors of verbal measures provided convergent validity data for their instruments (62%, see Table 3.14). Just under half of the authors of measures using pictorial or written presentation for items reported convergent validity data (47% & 39%, see Table 3.14).

Discriminant validity

Authors of 29 measures (28%) provided data on the discriminant validity of their instruments. Discriminant validity was assessed in two ways – whether scores on a given measure were unrelated to other measures assessing different concepts, and whether a given measure can distinguish between groups known to differ. The majority of the authors assessed the latter for their instruments (n=23). The authors used clinical (e.g., children with different ratings on DSM-III for depression) and physical (e.g., children's pain levels pre- and post- analgesia) indicators to group children to allow assessment of the discriminative value of scores on their measures.

Table 3.15: Measures with discriminant validity data by scale type

Type of scale used in measure	N of measures	Discriminant validity data reported % (n)
Likert	48	23 (11)
Graphic	28	32 (9)
Facial expression	23	26 (6)
Visual analogue	6	50 (3)

In relation to scale type, half of the authors of measures using visual analogue scales provided discriminant validity data for their instruments (50%, see Table 3.14). One third of the measures using graphic response scales had evidence for discrminant validity (32%, see Table 3.15). Roughly a quarter of the authors of measures using Likert or facial expression scales reported discriminant validity data (23% & 26%, see Table 3.15).

Table 3.16: Measures with discriminant validity data by presentation type

Type of presentation style	N of measures	Discriminant validity data reported % (n)
Verbal	39	18 (7)
Pictorial	30	20 (6)
Written	28	46 (13)
Computerised	5	60 (3)
Props	3	0 (0)

In relation to presentation style, two-thirds of the authors of computerised measures provided discriminant validity data for their instruments (60%, see Table 3.16). Just under half of the authors of measures using written presentation for items reported convergent validity data (46%, see Table 3.16). One fifth of authors of measures using verbal or pictorial presentation reported evidence of discriminant validity (18% & 20%, see Table 3.16).

Responsiveness/sensitivity data reported –

The number of researchers who reported investigation of the responsiveness of their measures was lower than the number who reported validity or reliability data. Many authors reported evidence for the reproducibility children's scores on their measures over a short time period (i.e., test-retest reliability over one or two weeks), but did not consider whether their measures were sensitive to change over a longer period of time [e.g., 4, 9, 59, 66, 69].

The authors of four measures (3%) [1, 13, 19, 65] evaluated the sensitivity of their measures. Of these measures, the authors of three instruments evaluated the sensitivity by looking for mean changes over time due to expected changes in status. For example, Le Coq et al. (2000) examined whether their QOL measure [13] was responsive to a change in children's asthma status between two time points (where a change in symptoms was reported by parents). Le Coq et al. (2000) used t-tests to see if the mean scores were different between the time points. The authors of the Pediatric Rhinoconjunctivitis Quality of Life Questionnaire measure [19] calculated a responsiveness index to assess the sensitivity of their instrument using a minimal importance difference in treatment score and variance in subjects (Juniper, Howland, Roberts, Thompson, & King, 1998). Juniper et al. (1998) used this index to examine whether their measure was sensitive to a change in children's rhinoconjunctivitis at two times (where a change was judged by a global rating of change by child themselves).

3.4 Review discussion

Report and recommendations (Phase 8)

Self-report measures for use with children below eight years to assess children's QOL, self-esteem/concepts, mental health, and pain were identified, compared, and critically evaluated in this review. In this review, we focused on examining the content and format of measures, and on comparing the quality of these different instruments.

The content and format of measures were compared on the following four criteria:

i. response scales (i.e., Likert, graphic, facial, or visual analogue response scales)
ii. justification/evidence for choice of response scale
iii. presentation styles (i.e., verbal, written, pictorial, props, or computerised)
iv. age ranges (i.e., narrow or wide age range).

Likert response scales were most commonly employed by authors to represent response choices to children (see p. 46). However a variety of arguments have been out forward as to the value of the other response scale types (graphic, facial expression, & visual analogue). All four types of response scales have been used successfully with young children in existing instruments (see Table 3.2, p. 75-90), and therefore researchers need to be able to justify their response scale choice (Wallander, Schmidt, & Koot, 2001).

The authors of 24 measures (23%) provided justification or evidence for their choice of response scale type (see p. 47). The most common method for providing justification for scale choice was compare the psychometric properties of measures across different response scales. The authors of seven measures assessed the relative value of different response scales for their measures (see p. 47). Similar studies would be useful to provide clear evidence for response scale choices for child self-report measures. We have discussed these issues in more detail in Chapter 7 (see p. 199-201).

The most common way to present items to children was to read them aloud (n=39, see p. 49). While this presentation method has been used successfully by researchers in their measures for young children, this review has also highlighted the potential value of using pictures (n=30) or props (n=3) for child self-report measures. Researchers have shown that presenting items using visual aids can be useful in facilitating children's understanding (Ernst et al., 1994, Harter & Pike, 1984, Martini et al., 1990), reducing demands on their memory (Patterson, 1995, Salmon, Bidrose, & Pipe, 1995), engaging and maintaining their attention (Mize & Ladd, 1988), and avoiding reliance on verbal or reading skills that may be lacking in younger children (e.g., preschool children, Measelle et al., 1998). We have discussed these issues further in Chapter 6 (see p. 171-2).

The measures identified in this review were targeted at a variety of ages, with some being aimed for wide age ranges (e.g., 6-14 years) and others focusing in on a narrow age range (e.g., 6-8 years, see p. 52). This review included any child measures that were targeted for children aged between three and eight years. This meant that measures met the inclusion criteria if they were targeted at for example children aged between seven and thirteen years. Examples of this were shown for the most part in QOL and mental health measures, where age ranges for instruments began around six or seven years and spanned all the way to fifteen or sixteen years. For example, the How are you? measure [13] had a target age range of seven to thirteen years, and the Short Children's Depression Inventory [68] can be used with seven to seventeen year olds.

We argued that measures developed for a large age range may be less appropriate and understandable for the younger aged children within the chosen age group, as the content of items may not have been developed specifically with children under eight years in mind. As we discussed in Chapter 2, there are differences in the cognitive capabilities and linguistic skills of children under eight and children over thirteen years, which raises the issue as to whether any one measure will be appropriate for such a wide age range (see p. 16-19). Chapter 2 considered what children aged between three and eight years can understand, in relation to concepts implicit behind items in self-report measures (e.g., emotions, self-understanding, mental representations, see p. 20-30). The differences in children's cognitive capacities and understanding across ages will have implications for the content of items included in self-report measures (see Chapter 2, p. 31-3). Measures targeted for wide age ranges may include items that are inappropriate for the younger children or use wording that children of a young age can not understand.

The quality of measures were compared on the following four criteria:

i. the methods authors used to generate items
ii. reliability data reported
iii. validity data reported
iv. responsiveness data reported.

The technique most commonly used was to allow children themselves to generate items (see p. 53-4). This was achieved by conducting child interviews, and using storybooks or pictures to facilitate discussions, or by using comments from children during piloting to adapt items. Using information from children themselves to generate items means that the content of measures are driven by the population they are targeted at (i.e., as items in child measures are based on children's own thoughts and views rather than the assumptions of adult researchers or proxies). Eiser and Morse (2001) in their review of QOL measures for children with chronic illnesses recommended that children should be involved in the development stages of instruments. We designed Study 3 (interview study) to incorporate children's own views and thoughts in the content development of our TedQL measure (see Chapter 5, p. 136-7). However the use of other item generation techniques, such as literature searching or using an expert panel to generate items are still useful and valid methods.

There are two ways to assess the reliability of a measure: assessing the internal consistency and the test-retest reliability of children's responses. Authors assessed the internal reliability of children's responses on their instruments more frequently than the test-retest reliability of responses (see p. 56 & p. 58). This may have been because the latter requires assessment at more than one time point, which can be time consuming and costly for researchers. However it is important for authors to show evidence that their measures produce reproducible responses from children over short time periods. In relation to internal reliability, the most popular statistic was Cronbach's alpha coefficient (α). We compared the amount of measures reporting internal reliability values for their measures across scale type and presentation style. Authors of measures using Likert or visual analogue scales reported the highest amount of internal consistency values above the .70 standard (see p. 57). Measures using a written format, or pictures or props to present their items to children had the highest amount of internal consistency values above .70 (see p. 57). In relation to test-retest reliability, the most popular statistic was Spearman's correlation coefficient (ρ). Again measures using Likert or visual analogue scales reported the most amount of reproducibility values above the .60 standard (see p. 58). Authors of measures using written or pictorial presentation styles reported the highest amount of reproducibility values above .60 (see p. 59). It was difficult to make a judgement on which scale or presentation type produced measures with the 'best' psychometric properties as not all authors reported reliability data for their measures.

There are three ways to assess the validity of a given measure (i.e., content, criterion, & construct), and ideally authors should provide evidence for all of these as they represent different aspects of any instrument. However typically researchers did not provide evidence for all the types of validity in their papers (see p. 59-64). The most common form of validity assessed was construct validity (both convergent and discriminant). The authors of 50 measures reported convergent validity data for their measures by comparing scores on their instrument with another measure hypothesised to be measuring the same construct (see p. 61-2). The number of measures for which authors reported convergent validity data was broadly the same across all four scale types (see p. 61-2). In relation to presentation style, the number of measures with convergent validity data was highest for verbal measures (see p. 62). The authors of 29 measures provided evidence for the discriminant validity of their instruments, and the majority of these authors did this by testing whether scores in their measure were uncorrelated to measures of different concepts (see p. 63).

The authors of 18 measures provided evidence for the content validity of their measures, and the most common techniques for doing this was using statistical methods such as factor analysis (n=6, see p. 60). The number of researchers who reported content validity was broadly the same across all four scale types (see p. 59-60).

The type of validity that was most frequently neglected was criterion validity (which includes concurrent, predictive, & retrospective). No authors attempted to assess the retrospective validity of their instruments (see p. 60-1), and only one group assessed the predictive validity of their measure (Verschueren et al., 2001). The authors of eight measures provided evidence for concurrent validity, and the most popular outcome measure was observed behaviour (see p. 60-1). There is a need for more authors to assess this aspect of validity. Assessing the responsiveness or sensitivity of measures was also neglected by the authors of measures included in this review. We identified four measures where authors had reported evidence for the sensitivity of their instruments (see p. 63-4). It would be useful for more researchers to attempt to assess whether their measures are sensitive to changes over time.

Getting evidence into practice (Phase 9)

The aim of our work was to develop a generic child self-report QOL measure for children below eight years. We designed our measure based around the recommendations from the literature reviewed in this and previous chapters (Chapters 1 and 2), and developed the content and format of the instrument from information obtained directly from children themselves over a series of studies that we report in this thesis. In the following sections we have highlighted how the literature we reviewed led to the initial development of our measure – the teddy bear QOL measure (TedQL).

a. Description of the measure

Presentation style

We decided to present the items in our TedQL measure to children using an interview format. Our review of child self-report measures in this review revealed that the most common way to present items to children was verbally (see p. 50). Our review also highlighted the potential value of using pictures or props to administer items to young children (see p. 51). However our review showed that no QOL measures had used props to present items to children (see p. 51). We decided to use two presentation styles for our measure – verbal and props.

We chose two identical teddy bears (40 cm high) which were only differentiated by their name badges to interview children (see Figure 5.3). We used the names *Iggy* and *Ziggy* for the teddy bears, which originated from the names used by Measelle et al. (1998) for their hand puppets for their self-perception scale – the Berkeley Puppet Interview (BPI).

Figure 3.3: Photograph of the two teddy bears (TedQL.1)

Previous work has shown that young children may be unwilling to say they resemble puppets/props/toys that were perceived to be of the opposite sex (Eder, 1990), therefore the teddy bears used in the TedQL measure were referred to as either female or male depending on the sex of the child.

Task design

The design of our interview task was based on two existing child self-concept/perception measures – the PSPCSA (Harter & Pike, 1984) and the BPI (Measelle et al., 1998). Harter and Pike (1984) gave children a description of a set of two pictures and asked them to chose which of the two pictures most closely resembled themselves. For example, a child was shown two pictures – one showing a child who is good at puzzles and one showing a child who is not very good at puzzles, and they were asked to indicate which they are most like out of the two. Once children made the first decision, they were then asked to think about whether they are a lot like their chosen pictures or a little bit. Responses are based on the choice of circles, where a big circle = *a lot like that*, and a little circle = *a little bit like that*. Measelle et al. (1998) used two hand puppets (tan-coloured puppy dogs) to interview children about their self-perceptions. The puppets were used to describe both sides of an item, e.g., one puppet would say "I have lots of friends at school" and the other would say "I don't have lots of friends at school", and then they asked the child which puppet they were like.

We developed an interview method for our instrument similar to Harter and Pike's (1984) and Measelle et al.'s (1998) measures where the bears were first described, and then children were asked to recognise themselves. We used a forced recognition task as used by Harter and Pike (1984). Specifically one teddy bear would represent the positive side of the statement (e.g., "Iggy/Ziggy is *good* at running"). The other would represent the negative side of the statement (e.g., "Iggy/Ziggy is *not good* at running"). In the first version of the TedQL measure (the TedQL.1), children were first asked which bear was *most like* them, and then they were asked how they *felt* about this level of functioning/ability. First, children were asked to point to which bear they were *most like*. Second, children were asked to point to a picture showing how *happy* or *sad* they felt about how they were at the described activity or behaviour.

Recall period

We chose a specific recall time period for children to use when answering items in our TedQL measure. In all versions of our TedQL measure the children were asked to think about how they had been during the last week when answering items. This decision was based partly on existing child QOL measures, such as the PedsQLTM4.0 measure which asks children to think about have they have been doing in the last few weeks (Varni, Seid, Knight, Burwinkle, Brown, & Szer, 2002).

b. Item content development

The specific content of the items in the TedQL.1 measure was guided by a review of the literature (see Chapter 2, p. 31-3), from measures currently available for young children (Eder, 1989, 1990, Guyatt et al., 1997, Harter & Pike, 1984, Stone & Lemanek, 1990) and previous experience with children (working one to one with children in nurseries and schools).

Emotion-related items

The literature reviewed in Chapter 2 showed that by two years old children can understand emotion words such as happy, sad and cross (see p. 21). Based on this review, we felt justified in including an item asking children how much they felt cross or angry in the TedQL.1 (as used in Study 1). When we expanded our measure to include more items (the TedQL.2 used in Study 2) we added two more emotion-related items – How much they felt happy or sad, and how much they worried about losing their things.

Mental representation related items

Based on our review of the developmental literature in Chapter 2, we established that researchers have argued that children's understanding of mental states increases significantly during and after four year of age (see p. 23-4). Therefore we felt able to include items asking children about their memory and psychological functioning in our TedQL measure. We included two items within the TedQL.1 and 2 that asked children about their functioning in this area – how good they were at remembering things they were told to do, and how much they had bad dreams at night.

Self-concept items

The review in Chapter 2 also provided evidence that children below eight years are capable of forming, holding and expressing self-concepts (see p 24-6). Based on the results of our literature review in Chapter 2 we felt that we could include items asking children about how they viewed themselves, their abilities, and their relationships. The TedQL.1 included three items asking children to rate their ability to run, play on swings, and to read. The TedQL.2 was expanded to include more items asking about their physical and cognitive abilities (e.g., hopping, climbing, tying shoelaces, writing, and drawing). The TedQL.1 also included items asking children how well they got on with their peers and their family (e.g., how much they get bossed around at school, and how much they get told off at home). In Study 2, the TedQL.2 was expanded to include more items asking about friends and family relations (e.g., how much they like to play with their friends, and how much they tell their mum/dad about what they have been doing at school).

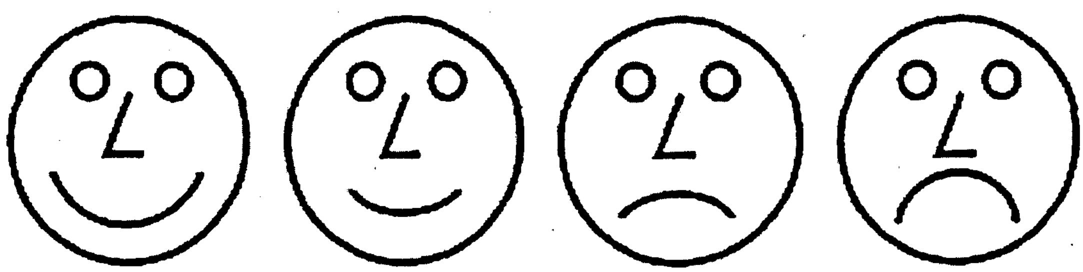
Social comparison-related items

Children need to be able to not only hold self-concepts but also be capable of comparing their own abilities and functioning to other people (e.g., their peers) to make a judgement of how good their own lives are. As discussed in Chapter 2, the ability to perform social comparisons is particularly relevant to QOL measures (see p. 26-8). The review in Chapter 2 found evidence that by three and four years of age children can make social comparisons. Based on the conclusions made in Chapter 2, we included items in the TedQL.1 and 2 that may require children to compare themselves to others (e.g., rating how good they were at a given activity by comparing themselves to other children in their class, or making a judgement on how good they would like to be at something).

c. Response scale choice

For the initial version of our measure (the TedQL.1) children's responses were made using a 4-point facial expression scale, as used in an existing child QOL measure (Christie et al., 1993, French, Christie, & Sowden, 1994). An example of the response scale used can be seen in Figure 3.4.

Figure 3.4: Image of the faces response scale (TedQL.1)



A facial expression scale was chosen to facilitate children in rating how they felt about their abilities and functioning (i.e., using happy and sad faces to rate whether they were happy or sad about a given item).

d. Establishing the procedure

First, we recognised the value of establishing rapport with children, and building up trust between the researcher and the child interviewees. Researchers such as Siegal (1997) have argued that that when children feel safe and relaxed, they will be more willing to report their thoughts and feelings. In all versions of the TedQL measure the interviewer spent at least one session in the classroom in the preceding days spending time with the children and helping them with their schoolwork, in order to help establish rapport with them beforehand.

Second, we included a training period at the start of the measure to help children become familiar with the rating task. Some researchers have shown that training periods may actually serve to increase the reliability of children's responses to self-report items (e.g., Harris, Guz, Lipian, & Man-Shu, 1985). Training periods can include both practise items and hypothetical questions. Practise items can help children gain experience in using the response scales and understanding the task. Hypothetical questions can be used to assess whether children can understand and use a given response scale correctly (see Chapter 7, p. 202). Due to these recommendations, we used a practise item at the start of the TedQL.1 ("how good are you at singing?", see p. 104) and added a hypothetical question to the TedQL.2 ("how much they liked their favourite sweet", see p. 122).

Third, we wanted to ensure that children did not build up a preference for one teddy bear over another, or think that one teddy bear was better than the other. Therefore the two teddy bears were counterbalanced for whether they represented a positive or negative statement, to help ensure no bear would be seen as particularly 'good' or 'bad' (i.e., half of the time, Iggy represented the positive side, and half of the time Ziggy represented the negative side, and vice-versa, Hughes, 1984).

The following chapters (Section 2, Chapters 4 & 5) introduce our child self-report measure (the TedQL), report the results of preliminary studies using this instrument (see Chapter 4), and further content development of the measure (see Chapter 5).

Table 3.2: Descriptive characteristics of the measures included in systematic review of child self-report measures

ID	Name of measure (study)	Target age group	Presentation style	Response scale type; no. of points in scale	Reliability data (internal consistency: α, K_{20}, split-half; test-retest reliability: ρ, ρ±)	Validity data (content; criterion: concurrent, predictive, retrospective; construct: convergent, discriminant; face)
Quality of life measures						
1	About my Asthma (Mishoe et al., 1998)	6-12	Written	Likert; 4	α=0.93	Convergent: r=0.41; Content: items panel rated
2	Animated Computer Program (Buller, 1999)	5-12	Computerised	Likert;		Pilot: children could understand and use scale
3	Childhood Asthma Questionnaire (Christie et al., 1993, Christie & French, 1994, French et al., 1994, 1998)	4-7	Written	Facial; 4	α=0.56, 0.63, ρ± = 0.59, 0.62	Discriminant: discriminates between children
4	Children's Dermatology Life Quality Index (Lewis-Jones & Finlay, 1995)	3-16	Written	Likert; 4	ρ=0.86	Discriminant: discriminates between children
5	Child Health and Illness Profile - Child Edition (Rebok et al., 2001, Riley et al., 2004)	6-11	Pictorial	Graphic; 5	α=0.64-0.83, ρ± = 0.35-0.69	Content: Only 6.5% found items hard to answer, factor analysis supported framework of the measure; Convergent: r=0.44-0.63
6	Child's Health Self-Concept Scale (Hester, 1990)	7-13	Written	Likert; 4	Hoyt=0.48-0.80, 0.86 total, α=0.70, ρ=0.44-0.58	Convergent: partial relation between child and parent reports; Content: items rated by panel of

						experts
7	Cystic Fibrosis Questionnaire - Child Version (Quittner et al., 2000, Henry et al., 1996)	6-13	Written	Likert; 4	α=0.46-0.71	
8	C-QOL (Jirojanakul & Skevington, 2000)	5-8	Written	Likert; 5	α=0.86, ρ=0.91	Face: asked three children; Content: items expert panel rated
9	DUCATQOL (Koopman et al., 1996, 1999, submitted)	6-16	Written	Visual analogue; 5	α=0.91, 0.92, ρ=0.84, 0.90	Discriminant: discriminates between children with different complaints
10	DUX-25 (Koopman et al., 2002, submitted)	5-16	Computerised	Facial; 5	Internal reliability and test-retest reliability at a good level (values not reported)	Discriminant: discriminates between children with different conditions
11	Exeter Health-Related Quality of Life measure (Eiser et al., 2000)	6-12	Computerised	Visual analogue; 0-100	α=0.50-0.69	Discriminant: discriminates between children
12	Generic Children's Quality of Life Questionnaire (Collier, 1997, Collier et al., 2000)	6-14	Pictorial	Likert; 5	α=0.74, 0.78	Content: items child generated
13	How are you? (Bruil, 1999, Le Coq et al., 2000)	7-13	Written	Likert; 4	α=0.77-0.86, ρ± = 0.46-0.83	Convergent: r=0.53-0.60; Discriminant: discriminates between children
14	Kiddy KINDL (Ravens-Sieberer & Bullinger, 1998, 2002, Ravens-Sieberer et al., 2001)	4-7	Verbal	Likert; 3		
15	Life Activities	5-17	Written	Likert; 5	α=0.97, ρ=0.76	Content: items child generated

	Questionnaire for Childhood Asthma (Creer et al., 1993)					during interviews
16	Nordic Quality of Life Questionnaire for Children (Lindstrom, 1993, Lindstrom & Eriksson, 1993)	6-18	Written	Likert;		
17	Pediatric Asthma Quality of Life Questionnaire (Juniper et al., 1992, 1996)	7-17	Written	Likert; 7	ρ± = 0.84-0.93	Convergent: r=0.30-0.58
18	Pediatric Quality of Life Questionnaire (Varni et al., 1998, 1999, 2002, submitted)	5-18	Written	Likert; 3, 5	α=0.59-0.85	Convergent: correlated to related scales; Discriminant: discriminates between children
19	Pediatric Rhinoconjunctivitis Quality of Life Questionnaire (Juniper et al., 1998)	6-12	Verbal	Likert; 7	ρ± = 0.93	Concurrent: correlation to diary symptom scores
20	Personal Adjustment and Role Skills Scale (Ellsworth , 1981, Stein & Jessop, 1990, Walker et al., 1990)	5-18	Verbal	Likert; 4	α=0.71-0.90	Content: factor analysis showed six clear factors, discriminates between children; Convergent: r=0.74-0.80 with other related scales,
21	Pictured Child's Quality of Life Self Questionnaire (Manificat et al., 2000)	4-12	Pictorial	Facial; 4	α=0.64, 0.71	Content: Factor analysis showed four clear factors, Discriminant: discriminates between children; Face: acceptable to children;

						Convergent: r=0.30
22	Quality of Life Scale - 'Three wishes' and 'What worries you the most?' (Neff & Dale, 1990)	7-12	Written	Facial; 5	K_{20}=0.49-0.85	
Self-esteem/self concept measures						
23	Berkeley Puppet Interview (Measelle et al., 1998)	4.5 up	Props	Likert; 7	α=0.62-0.76, correlated over time	Concurrent: scores correlated with behaviour tests & achievement tests
24	Brown IDS Self-Concepts Referents Test (Boger & Knight, 1969, Brown, 1969, Bridgeman & Shipman, 1978, Walker et al., 1973)	3.5-6.5	Pictorial	Likert; 2	α=0.59-9.82, ρ=0.55-0.76	Convergent: r=0.23-0.34
25	Children's Physical Self-Concept Scale (Stein et al., 1998)	6 up	Written	Graphic; 4	α=0.60-0.81, 0.77 total, ρ=0.69-0.94, 0.88 total	Discriminant: discriminates between different children
26	Children's Self-Concept Index (Boger & Knight, 1969, Helms et al., 1968)	5.5 up	Pictorial	Likert; 2	α=0.80	
27	Children's Self-Social Construct Test: Preschool Form (Biller, 1968, Boger & Knight, 1969, Flammer, 1971, Long & Henderson, 1968, 1970, McDowell, & Lindholm, 1986,	3 up	Written	Graphic; 5	Split half=0.65-0.77, α=0.62	Discriminant: discriminates between different children

	Raizen & Bobrow, 1974)					
28	Competence Perceptions Subjective Task Value Beliefs (Eccles et al., 1993)	6.5 up	Written	Graphic; 7	α=0.67-0.82, & 0.55-0.86	Content: evidence of consistent factors reflecting discriminations across domains and between concepts
29	Experimental Photographic Self Concept Test (Boger & Knight, 1969)	3 up	Pictorial	Facial;		
30	I Think I am (Reichenberg & Broberg, 2002)	7-9	Written	Likert; 2		
31	I Feel-Me Feel (White & Human, 1976)	3-6	Pictorial	Facial; 5	α=0.73	Convergent: no correlation between child and parent or nurse ratings; Content: factor analysis revealed one main factor (self/social contruct)
32	Joseph Preschool and Primary Self-Concept Screening Test (Joseph, 1979)	5.5 up	Pictorial	Likert; 2	α=0.73	Convergent: r=0.51
33	McDaniel-Piers Young Children's Self-Concept Scale (McDaniel & Leddick, 1978, McDaniel, 1973)	4 up	Written	Likert;	KR_{20}=0.72-0.92, ρ=0.65	Convergent: r=0.26-0.54; Content: factor analysis identified three factors linked to domains
34	Martinek-Zaichowsky Self-Concept Scale (Martinek & Zaichowsky, 1975)	6.5 up	Pictorial	Likert; 2	α=0.88	Convergent: r=0.49

35	Maryland Pre-School Self-Concept Scale (Smith, 1978, Hughes, 1981, Hughes & Leatherman, 1982)	4-6	Pictorial	Likert; 2	α=0.58-0.67, ρ=0.77	Convergent: r=0.42
36	Maryland Pre-School Self-Concept Scale - Revised (Hughes, 1981, 1984)	4-6	Pictorial	Likert; 2	α=0.66-0.77, ρ=0.62	Convergent: r=0.29-0.54
37	North York Primary Self-Concept Test (Crawford, 1977)	6 up	Pictorial	Facial; 2	Split-half=0.80, K_{20}=0.80-0.85	
38	Perez Self-Concept Inventory (Perez, 1982)	5.5 up	Pictorial	Likert; 2	α=0.80, ρ=0.77	Convergent: r=0.05-0.46
39	Pictorial Self-Concept Scale (Bolea et al., 1971)	5-9	Pictorial	Likert; 2	Split half=0.85	Convergent: r=0.42
40	Pictorial Scale of Perceived Competence and Social Acceptance (Harter & Pike, 1984, Fantuzzo et al., 1996)	4-7	Pictorial	Graphic; 4	α=0.50-0.85	Face: asked children reasons for their answers; Content: factor analysis supported two out of the four hypothesised factors; Discriminant; discriminates between different children
41	Piers Preschool Pictorial Self-Concept Scale (Jensen, 1983, 1985)	4 up	Pictorial	Likert; 3	α=0.65, ρ=0.84	Convergent: low correlation to teachers ratings; Content: based on children's comments, factor analysis identified four factors linked to domains
42	Preschool Self-Concept Picture Test (Boger & Knight, 1969, Woolner,	4 up	Pictorial	Likert; 2	α=0.70, & 0.85-0.93	

	1966, McDowell & Lindholm, 1986)					
43	Preschool and Primary Self-Concept Scale (Stager & Young, 1982)	4-9	Pictorial	Likert; 4	α=0.72, 55% stable	
44	Primary Self-Concept Inventory (Pomerance Torshen et al., 1977, McDowell & Lindholm, 1986)	4.5-8	Verbal	Graphic; 3	α=0.36, ρ=0.38-0.73	Content: factor analysis revealed seven factors congruent with seven subscales in measure
45	Puppet Interview (Cassidy, 1988, Verschueren et al., 2001)	5-7	Props	Likert; 2		Predictive: scores at 5 yrs old were related to outcomes scores at 8 yrs old (peer acceptance, global self-wroth, etc)
46	Purdue Self-Concept Scale for Preschool Children (Cicirelli, 1974, Samuels & Griffore, 1979)	4 up	Pictorial	Likert;	α=0.86, ρ=0.70	Convergent: correlations between child and teacher ratings
47	Reading Self-Concept Scale (Chapman & Tunmer, 1995)	5-8	Verbal	Likert; 5	α=0.82-0.88	
48	Self-Concept and Motivation Inventory (McDowell & Lindholm, 1986, Milchus, Farrah & Reitz, 1968, Davis & Johnson, 1987)	5 up	Written	Facial; 3	α=0.56-0.69, ρ=0.38-0.66	
49	Self-Description Questionnaire-I (Marsh	5-8	Written	Likert;	α=0.50-0.78, & 0.83-0.95, ρ=0.32-0.47	Convergent: r=0.38, r=0.52; Content: factor analysis revealed

	et al., 1991, 1998)					eight factors linked to domains
50	Self-Interview (Eder, 1990)	3-8	Props	Likert;	α=0.75-0.80, ρ=0.47-0.65	Content: items rated by panel, factor analysis identified three factors linked to subscales
51	The Self-Social Constructs Test - Self-Esteem Scale (Long et al., 1969, Boger & Knight, 1969)	4.5 up	Pictorial	Graphic;	Split-half=0.48-0.77	Discriminant: discriminates between different children
52	Thomas Self-Concept Values Test (Michael, 1972, Suinn 1972)	4.5 up	Pictorial	Likert; 2	α=0.73, ρ=0.78	
53	U-Scale Self-Concept Test (Ozehosky & Clark, 1971)	5.5 up	Pictorial	Likert;	α=0.67, ρ=0.67	
54	What Face do you Wear? (Davis & Johnston, 1987)	4 up	Pictorial	Facial; 3		
Mental health measures						
55	Children's Dental Fear Picture Test (Klingberg et al., 1995)	4 up	Pictorial	Graphic; 4		Discriminant: discriminates between different types of children (categorised on fear levels based on parent's ratings)
56	Children's Depression Inventory (Carey et al., 1987, Edelsohn et al., 1992, Ialongo et al., 1993, Saylor et al., 1984)	6 up	Written	Likert; 3	α=0.81-0.84, ρ=0.59 and 0.77	Convergent: correlation to teacher and peer ratings; Discriminant: discriminates between clinic-referred and 'normal' children
57	Children's Manifest	6-19	Written	Likert; 2	K_{20}=0.83, ρ=0.58-0.68	Convergent: correlations to

	Anxiety Scale (La Greca et al., 1988, Reynolds, 1980, 1981, 1982, Reynolds & Richmond, 1978, 1979)					anxiety scores; Discriminant: discriminates between different children; Content: factor analysis identified three factor solution fitting with domains in measure
58	Children's Self-Report Questionnaire (Beitchman et al., 1985, 1987, 1989)	7-12	Written	Likert;	ρ=0.77 and 0.67	Discriminant: discriminates between different children (psychiatric status)
59	Dominic Questionnaire-R (Valla, 2000, Breton et al., 1999)	6-11	Computerised	Likert; 2	α=0.64-0.83, ρ± = 0.71-0.81	Discriminant: discriminates between different children
60	Fear Faces Scale (Katz et al., 1982, 1982)	4 up	Pictorial	Facial; 7		Convergent: correlated to measures of distress and other self-report anxiety and pain measures
61	Hospital Fears Rating Scale (Melamed & Lumley ,1988)	6 up	Written	Graphic; 5	ρ=0.75	Discriminant: discriminates between children having intervention to help prepare them for surgery and controls
62	Hospital Fears Questionnaire (Roberts et al., 1981)	5 up	Verbal	Graphic; 5, 8		Convergent: scores correlated to other anxiety scales
63	Levonn Scale (Martinez & Richters, 1990, Richters & Martinez, 1993, Richters et al., 1990, Shaninfar et al., 2000)	3.5 up	Pictorial	Graphic; 3	α=0.87 and 0.89, ρ=0.60	Convergent: child scores correlated to parent scores on behaviour measure and other violence exposure scale (child-report)
64	Medical Fear	4 up	Written	Likert; 3	α=0.84	Concurrent: scores correlated

	Questionnaire (Broome, 1986)					with children's observed behaviour
65	Pictorial Instrument for Child and Adolescent Psychiatry (Ernst et al., 1994)	6-16	Pictorial	Graphic; 5	α=0.54-0.86	Discriminant: scores discriminated between children with range of different childhood disorders (DFA used to assess this)
66	Preschool Symptom Self-Report (Martini et al., 1990)	3 up	Pictorial	Likert; 2	α=0.89, $\rho \pm$ = 0.86	Content: scores linked to spontaneous comments from children
67	Scare Scale (Beyer & Arandine, 1988)	3-13	Verbal	Graphic; 5		Convergent: correlation to other fear scale, Discriminant: no relation to pain measures
68	Short Children's Depression Inventory (Carlson & Cantell, 1979, 1980, Edelsohn et al., 1992)	7-17	Written	Likert; 4		Discriminant: discriminates between children with different ratings on DSM-III for depression
69	Social Anxiety Scale for Children Revised (La Greca et al., 1988, La Greca & Lopez, 1998, La Greca & Stone, 1993)	7 up	Written	Likert; 3	α=0.63-0.83, ρ=0.39-0.70	Convergent: scores correlated to other anxiety scales; Discriminant: discriminates between different children; Content: factor analysis identified two factors & items not loading over 0.30 removed
70	The Anxiety Svale (Boger & Knight, 1969)	3 up	Written	Facial;	ρ=0.57-0.89	Convergent: no correlation with teacher rating of anxiety
71	Venham Picture Test (Venham & Gaulin-Kremer, 1979)	4 up	Pictorial	Facial; 2	K_{20}=0.84, ρ=0.70	Content: items based on children's spontaneous comments; Concurrent: scores

						on measure consistent to observed behaviour; Discriminant: mean anxiety score was higher immediately before dental visit than directly after visit
Pain measures						
72	Abu-Saad Paediatric Pain Assessment Tool (Abu-Saad, 1990, Abu-Saad et al., 1994)	5-15	Verbal	Visual analogue; 10	α=0.77 & 0.83	Content: factor analysis revealed two factor structure; Convergent: child scores correlated to parent and nurse ratings; Discriminant: no correlation to fear ratings; Concurrent: scores correlated to other indicators of pain (observed behaviour)
73	Adjectival Rating Scale (Champion et al., 1998, Goodenough et al., 1997)	4-12	Verbal	Likert; 6		Face: least preferred pain scale; Convergent: correlated to other pain scales (CAS, poker chip tool)
74	Algocube (Poulain et al., 1993)	5 up	Verbal	Graphic; 5		
75	Block Scale (Lehmann, 1990)	3-8	Verbal	Graphic; 6	Not over 80% stable when under 7 years old	
76	Charleston Pediatric Pain Pictures (Adesman & Walco, 1992, Belter et al., 1988)	3-9.5	Pictorial	Visual analogue; 5	α=0.54-0.74, ρ=0.53-0.72	Convergent: scores correlated to VAS and CAS pain scales
77	Children's Anxiety and	3-17	Verbal	Facial; 2		Content: children's comments

	Pain Scale (Kuttner & Lepage, 1983, 1989, Goodenough et al., 1998, 2000, Chambers et al., 1999)					used for items; Concurrent: scores related to observational measures of distress; Convergent: poor relationship between parent and child pain ratings, scores correlated to correlated to other pain scales
78	Children's Global Rating Scale (Carpenter, 1990)	4 up	Verbal	Graphic; 5		Convergent: correlation between child and parent ratings; Discriminant: children's pain and fear ratings highly correlated
79	Coloured Analogue Scale (McGrath et al., 1996, Hicks et al., 2001, Champion et al., 2000, Colwell et al., 1996)	3-15	Verbal	Graphic; 0-100		Convergent: scores correlated to other pain scales (faces and SAFE); Face: most preferred pain scale
80	Faces Scale (Tyler et al., 1993)	6 up	Verbal	Facial; 5		Discriminant: scores differed as hypothesised when compared pre- and post-operatively
81	Faces Pain Scale (Bieri et al., 1990, Hunter et al., 2000, Goodenough et al., 1997, 1999, 2000, Chambers et al., 1999)	3.5 up	Verbal	Facial; 7	ρ=0.35-0.81 & 0.79	Content: established during development stages, using piloting & children's preferences; Convergent: scores correlated to other pain scales (FAS, CAS, SAFE), poor correlations found between child and pain ratings
82	Faces Rating Scale (Wong & Baker, 1988, Chambers et al., 1999,	5-13	Verbal	Facial; 6		Convergent: scores correlated to other pain scales (FPS), moderate correlation to nurse

	West et al., 1994, Whaley & Wong, 1991)					ratings, poor correlation to parent ratings; Face; majority of parents and children preferred this faces scale over other faces scales
83	Facial Affective Scale (McGrath, 1987, 1990, 1990, McGrath et al., 1985, Goodenough et al., 1999)	4-10	Verbal	Facial; 9		Convergent: scores correlated to other pain scales; Concurrent: scores correlated to observational measures of distress
84	Facial Expression Scale (Maunuksela et al., 1987, Chambers et al., 1999)	3 up	Verbal	Facial; 5		Convergent: poor relationship between pain and child pain ratings (kappa=0.21)
85	Facial Expression Scale (Zelter et al., 1988)	3 up	Verbal	Facial;		Convergent: poor relationship between pain and child pain ratings (kappa=0.36)
86	Finger Span (Champion et al., 2000)	4-12	Verbal	Graphic; 5		Convergent: poor correlation to other pain scales; Face; least preferred scale, rated hardest and confusing
87	Glasses Scale (Whaley & Wong, 1991, Wong & Baker, 1988)	3-18	Verbal	Graphic; 6		Face: preferred faces scale over this scale
88	Horizontal Scale (Lehmann, 1990)	3-8	Verbal	Graphic; 5	Not over 80% stable when under 7 years old	
89	Mechanical Visual Analogue Scale (Price, 1994, Price et al., 1994)	3 up	Verbal	Graphic; 15		Convergent: scores related to scores on faces pain scales (FAS and FPA)

90	Number-Thermometer Scale (Szyfelbein et al., 1985)	3 up	Verbal	Graphic; 10		
91	Oucher (Beyer & Arandine, 1986, 1988)	3-12	Verbal	Facial; 100		Discriminant: scores differed as hypothesised when compared by pre- and post-analgesia; Convergent: scores correlated to other pain scales (Poker Chip tool); Discriminant: low correlation to fear scales
92	Pain Ladder (Hester et al., 1990, 1989)	4-13	Verbal	Graphic; 10	α=0.66	Convergent: related to variances; Discriminant: related to variances
93	Pain Thermometer (Szyfelbein et al., 1985, Belter et al., 1988)	3 up	Verbal	Graphic; 8-10	α=0.56-0.69, ρ=0.57-0.70	Convergent: scores highly correlated to other pain scales (poker Chip tool, Oucher, faces)
94	Poker Chip Tool (Hester et al., 1990, Hester, 1979, Aradine et al., 1988, Beyer & Arandine, 1986, 1988, Champion et al., 2000, Goodenough et al., 2000, St-Laurent-Gagnon et al., 1999)	4-13	Verbal	Graphic; 5	α=0.70	Concurrent: vocal behaviours related to scores; Convergent: scores correlated to other pain scales; Face: preferred over other pain scales
95	Red & White Visual Analogue Scale (Maunuksela et al., 1987)	3 up	Verbal	Graphic; 11		
96	Sheffield Children's	6.5 up	Verbal	Facial; 5		Convergent: scores correlated to

	Hospital Pain Assessment Tool (Goddard, 1996)					nurse ratings
97	Sydney Animated Facial Expression Scale (Goodenough et al., 2000, Champion et al., 2000. Hicks et al., 2001)	4 up	Computerised	Facial; 100		Convergent: correlated to other faces pain scales (FPS) and pain scales (VAS)
98	Smiley Faces Scale (Jylli & Ollson, 1995, Pothman, 1990)	3-18	Verbal	Facial; 5		Convergent: scores highly correlated to other pain scale (VAS), low correlation between parent and child ratings
99	Verbal Analogue Scale (Hicks et al., 2001)	4-12	Verbal	Visual analogue; 0-200		Convergent: scores correlated to CAS pain scale
100	Verbal Analogue Scale (Thomas et al., 1997)	3-15	Verbal	Visual analogue;		
101	Vertical Scale (Lehmann, 1990)	3-8	Verbal	Graphic; 5	Not over 80% stable when under 7 years old	
102	Visual Analogue Toy (Arts et al., 1994, Goodenough et al., 2000)	4-16	Verbal	Graphic; 10		Convergent: scores highly correlated to other pain scales (faces); Concurrent: scores correlated to observed behaviour; Construct: PCA loadings reported to show all pain scales loaded on same factor
103	Waldron/Varni Pediatric Pain Coping Inventory	5-18	Verbal	Likert; 3	α=0.57-0.67, 0.85 for total	

	(Varni et al., 1996)					
104	Word Descriptor Scale (Fogel Keck et al., 1996)	3-18	Verbal	Likert; 5	ρ=0.92	Convergent: correlated to VAS and numeric pain scales; Discriminant: scores discriminates between children not in pain before surgery and in pain after surgery
105	Word Graphic Scale (Savedra & Tesler, 1989, Savedra et al., 1990)	3-15	Verbal	Likert; 5		

Figure 3.5: Example of the Likert scale from the Kiddy KINDL [14] QOL measure

3. ... and how you have been feeling about yourself.

During the past week...	never	sometimes	very often
1. ... I was proud of myself	☐	☐	☐
2. ... I felt pleased with myself	☐	☐	☐

Figure 3.6: Example of the Likert scale from the How are you? [13] QOL measure

Have you remembered what you learned at school during the past seven days?
No, never
Yes, sometimes
Yes, often
Yes, very often

Figure 3.7: Example of the graphic scale from the Pain Thermometer [93] pain measure

Rate how much pain you have on the thermometer below:

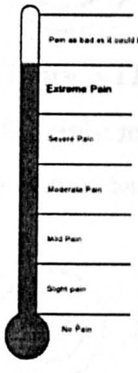


Figure 3.8: Example of the graphic scale from the Glasses Scale [87] pain measure

How much pain do you feel? Rate using these glasses below:

None.......................................Alot

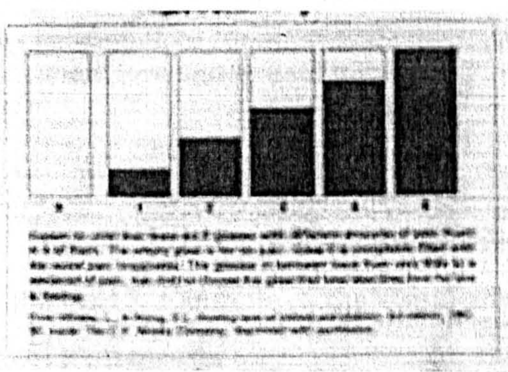
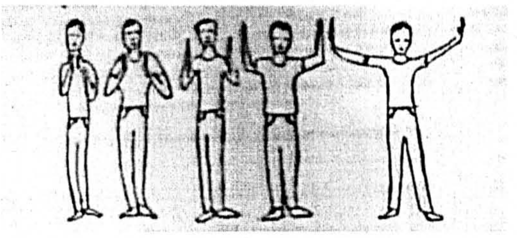


Figure 3.9: Example of the graphic scale from the Pictorial Instrument for Child and Adolescent Psychiatry [65] mental health measure

How anxious do you feel?



None at all................................A lot

Figure 3.10: Example of the facial expression scale from the Childhood Asthma Questionnaire [3] QOL measure

Do you go to the swimming pool? Yes/No

How do you feel when you go to the swimming pool?

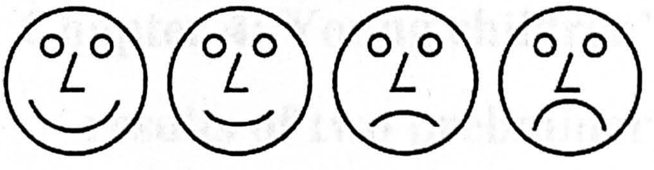


Figure 3.11: Example of the facial expression scale from the Faces Rating Scale [82] pain measure

How much hurt do you have?

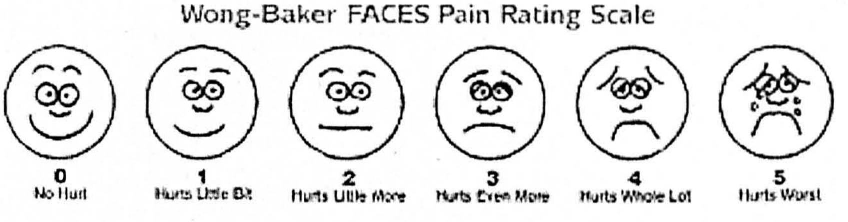


Figure 3.12: Example of the visual analogue scale from the Word Descriptor Scale [104] pain measure

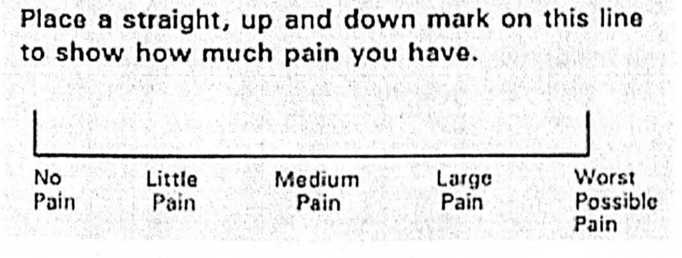
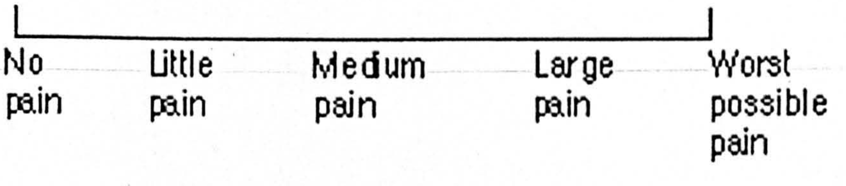


Figure 3.13: Example of the visual analogue scale from the Word Graphic Scale [105] pain measure



Chapter 4: Young children's self-reported QOL – results of two preliminary studies (Studies 1 and 2).

Summary

Study 1

Aims

Study 1 aimed to: i) compare the psychometric properties of our new measure (TedQL.1) to an established measure (PedsQLTM4.0), ii) investigate the relationship between these two measures, and iii) explore the relationship between child and parent rated child QOL.

Sample

Thirty-six children (5.0-8.0 years) completed two QOL measures (TedQL.1, PedsQLTM4.0). Twenty-four of their parents completed the PedsQLTM4.0 for their child.

Results

Children's responses were more consistent for the PedsQLTM4.0 compared with the TedQL.1. The TedQL.1 was reported to be easier and more enjoyable than the PedsQLTM4.0. There was a positive correlation between the children's scores on the TedQL.1 and on the PedsQLTM4.0. There was no significant relation between children's TedQL.1 and parents' PedsQLTM4.0 scores. However children's PedsQLTM4.0 scores were positively correlated to parents' PedsQLTM4.0 scores.

Implications

The TedQL.1 measure was related to an established QOL measure meaning and therefore we hypothesised that both measures were assessing a similar underlying construct. The TedQL.1 measure was expanded to include more items in attempt to raise the internal consistency of children's responses on this instrument.

Study 2

Aims

Study 2 aimed to: i) investigate the impact of scale type (i.e., circular versus linear) on children's TedQL.2 responses, and ii) explore the relationship between child and parent rated child QOL.

Sample

Twenty-eight children (3.0-5.0 years) completed the TedQL.2 measure, using one of two different response scales (circles or linear). Twenty-one of their parents completed the PedsQL$_{TM}$4.0 for their child.

Results

Children's responses using the circles scale had higher internal consistency than children using the linear scale. Children using the circles response scale took less time to complete the TedQL.2 items than children using the linear scale. Children's TedQL.2 scores were not significantly correlated to parents' PedsQL$_{TM}$4.0 scores.

Implications

The results confirmed that scale type can impact on the psychometric properties of a measure. Child and parent rated child QOL were not related when they were using different measures. As a result of Studies 1 and 2, we realised the need to develop a parent report version of the TedQL measure.

4.1 Introduction (Study 1: Preliminary validation of the TedQL.1 measure)

The purpose of the empirical studies reported in this thesis was to develop a generic QOL measure that could be used to gain self-reports from children below eight years. This chapter reports the results of the first two studies, both of which were preliminary studies in the development of a new child QOL measure (the TedQL.1, & TedQL.2).

4.1.1 Comparing the psychometric properties of the TedQL.1 to an established measure (PedsQLTM4.0)

Researchers have argued that part of the psychometric validation of new instruments involves comparison to existing measures (e.g., Graham et al., 1997, Langeveld et al., 1996). Therefore the first aim of this study was to compare the psychometric properties of the initial version of our new measure (TedQL.1) with an established measure that has been used successfully with young children.

We felt it would be useful to use the PedsQLTM4.0 measure for this comparison, as it is a well-validated instrument that has been used successfully with healthy children and those with chronic diseases, and their parents (Varni et al., 2002). The PedsQLTM4.0 measure was developed over a series of studies conducted during the 1990's by Varni and colleagues (Varni et al., 2002). This measure was originally designed as a generic HRQOL instrument to be used with all types of paediatric populations (Varni, Seid, Jacobs, & Rode, 1999). Disease-specific modules have also been developed to use with the generic core, to measure QOL in specific populations (e.g., cancer, asthma, Varni, Rode, Seid, Katz, Freidman-Bender, & Quiggins, 1999, Varni, Seid, Jacobs, & Rode, 2000). The measure has also been adapted for use with children below eight years, by incorporating a narrower response scale (i.e., from a 5-point to a 3-point Likert scale) and re-wording items for a lower reading age (e.g., 'low energy level' changed to 'feeling too tired to play').

Despite both the TedQL.1 and PedsQLTM4.0 being measures of QOL these two measures differed in the way items are presented to children. The PedsQLTM4.0 measure uses a questionnaire format to present items verbally to children, whereas the TedQL.1 uses teddy bears as three-dimensional props to illustrate items to children. As discussed in Chapter 6 (see p. 179-80), researchers such as La Greca (1990) have argued that children may have difficulty reading and understanding questions when presented in a questionnaire format, and that they may not respond well to formal data collection methods that have been used successfully with adults. There is also evidence that measures using pictures and/or props appeal to young children, and help maintain their attention better than written measures (Ceci & Bruck, 1993, Measelle et al., 1998, Harter & Pike, 1984). Such aids also serve to clarify and concretise items to children (Ernst et al., 1994, Harter & Pike, 1984), and therefore lead to more meaningful responding (Mize & Ladd, 1988).

Based on the literature reviewed in Chapter 6 (see p.179-81) as summarised above, we felt that children would respond to and understand the TedQL.1 measure better than the PedsQLTM4.0 measure, as the TedQL.1 uses three-dimensional props to present items whereas the PedsQLTM4.0 items are presented verbally to children. Therefore we predicted that the psychometric properties of the TedQL.1 measure would be better than the PedsQLTM4.0 measure (see Chapter 6, p. 179-81). We predicted that the internal reliability of children's responses on TedQL.1 would be higher than their responses on the PedsQLTM4.0. We also predicted that more children would rate the TedQL.1 measure as the easiest and most enjoyable instrument to use, when compared to the PedsQLTM4.0 measure.

4.1.2 Investigating the relationship between the TedQL.1 and the PedsQLTM4.0 measures

The second aim of this study was to explore the relationship between the TedQL.1 and PedsQLTM4.0 measures. This aim was achieved by investigating whether children's TedQL.1 scores were correlated to their PedsQLTM4.0 scores.

Both the TedQL.1 and the PedsQLTM4.0 were developed to be self-report measures of child QOL. The PedsQLTM4.0 measures QOL by asking children about their actual functioning and abilities (e.g., "how much of a problem have you had with running in the last week?"). The TedQL.1 measures QOL by asking children how they are at a given activity or about their level of functioning, and then asking how they feel about this level (i.e., *happy* or *sad*). In this way the TedQL.1 measure includes individual 'happiness' with functioning or abilities rather than just asking children about their actual levels of functioning (Lawford, Volavka, & Eiser, 2001). Although the TedQL.1 'happiness' scores take individual preferences into account, these scores are in essence still measuring QOL and therefore should still be related to scores on the PedsQLTM4.0 measure (which measures actual functioning and abilities). Based on the argument that scores from instruments that are measuring the similar constructs should be correlated to each other, we predicted children's happiness/QOL (TedQL.1) scores would be positively correlated to their total QOL (PedsQLTM4.0) scores (i.e., children's TedQL.1 scores should go up, as their PedsQLTM4.0 scores go up).

4.1.3 Exploring the relationship between child and parent rated child QOL

Researchers have advocated that a necessary requirement for the validation of new child measures is moderate agreement between proxy and child reports (e.g., Theunissen, Vogels, Koopman, Verrips, Zwinderman, Verloove-Vanhorick, & Wit, 1998, Varni et al., 1999, see Chapter 1, p. 5). Therefore the third aim of this study was explore the relationship between child and parent rated child QOL. As the TedQL.1 measure did not have a parent report version, we assessed parent-child agreement when children used the TedQL.1 and parents used the PedsQLTM4.0. We also assessed parent-child agreement when children and parents were using the same measure (PedsQLTM4.0).

Despite the requirement for moderate parent-child agreement for the validation of new measures, a number of researchers have shown that the level of agreement may be quite poor (e.g., Langeveld et al., 1997, Vogels et al., 1998). As discussed in Chapter 1 (p. 5-6), there may be reasons for this lack of agreement between children and their parents on measures of child QOL. Vance et al. (2001) have argued that parents may not always have enough information about their children's internal states to give an accurate proxy rating, and therefore report their child's QOL from their own perspective as both an adult and a parent. Guyatt et al. (1997) reported that parents of young children do not have a good idea of the subjective aspects of their children's world, and therefore gain information on their children's lives from more easily observable aspects (i.e., from their child's actual overt external behaviour). Based on the literature highlighted above, we predicted that children's and parents' ratings of child QOL would not be significantly correlated, across either the TedQL.1 or the PedsQLTM4.0 measures.

4.1.4 Summary of the aims and predictions of Study 1

The first aim of Study 1 was to compare the psychometric properties of our new measure (TedQL.1) to an established measure (PedsQLTM4.0). We predicted that the psychometric properties of the TedQL.1 measure would be better than the PedsQLTM4.0 measure. Specifically we predicted that the internal reliability of children's responses on TedQL.1 would be higher than their responses on the PedsQLTM4.0. We also predicted that more children would rate the TedQL.1 measure as the easiest and most enjoyable instrument to use.

The second aim was to investigate the relationship between these two measures. We predicted children's TedQL.1 scores would be positively correlated to their PedsQLTM4.0 scores.

The third aim was to explore the relationship between child and parent rated child QOL. We predicted children's and parents' ratings of child QOL would not be correlated across either the TedQL.1 or the PedsQLTM4.0 measures.

4.2 Methodology (Study 1)

Sample

Ethics approval was obtained from the Department of Psychology Ethics Committee at the University of Sheffield. Forty participants aged 4-8 years were identified from a primary school in Sheffield. Children took a letter home to their parents explaining the study. Their parents were asked to complete a permission slip for their child to take part. Two children were excluded because permission was not granted, and two others owing to learning difficulties. Thirty-six children (15 females & 21 males; 5.0-8.0 years) completed the study. The mean age of the children was 5.91 years (SD= 0.31 years). 24 (67%) of the children were Caucasian and 12 were of Asian origin (33%).

Questionnaires for parent completion were sent home with all children who had participated in the study (n=36). Twenty-four parents returned their questionnaires giving 24 parent-child dyads in Study 1.

4.2.1 Child-data

Measures

TedQL.1 measure

The children were interviewed using the two identical teddy bears as described in Chapter 3, which were referred to as either female or male depending on the sex of the child (see Figure 3.3, p. 70).

Ten items were selected for the TedQL.1 version which were thought to best represent children's lives and concerns in a general way; for example "Does your mum or dad tell you off at home?" was used instead of "Do your brothers or sisters fight with you?" as not all children have siblings. Eder (1990) has provided evidence that preschoolers and young children exhibit a predominance of general self-descriptions (e.g., "I usually play with my mum") and also that they describe themselves in terms of typical behaviours and activities rather than trait adjectives. Therefore we developed items that described activities and behaviours in general terms. The initial version of our measure consisted of 10 items within 5 areas. These are listed in Table 4.1.

Table 4.1: Items in the TedQL.1 measure

Area	Items
Physical Competence	PC1: good at running PC2: good at swinging
Peer Acceptance	PA1: having lots of friends PA2: getting bossed around at school (N)
Maternal Acceptance	FA1: mum/dad telling them off at home (N) FA3: going on trips with mum/dad
Psychological Functioning	PF1: getting cross/angry (N) PF2: having bad dreams at night (N)
Cognitive Functioning	CF1: good at reading CF2: remember what teacher asked him/her to do

Note. (N)= negatively scored item, scores were reversed.

As described in Chapter 3, a forced recognition task was used where the bears were first described and then children were asked to recognise themselves (see p. 70). As described in Chapter 3, one teddy bear represented the positive side of the statement and the other represented the negative side of the statement (see p. 70). First, children were asked to point to which bear they were *most like*. Second, children were asked to point to a picture showing how *happy* or *sad* they felt about how they were at the described activity or behaviour. The children were asked to think about how they had been during the last week when answering the items. Responses were made using a 4-point facial expression response scale, as used in an existing child QOL measure (Christie et al., 1993, French, Christie & Sowden, 1994). An example of the response scale used can be seen in Chapter 3 (see Figure 3.4, p. 73).

PedsQLTM4.0 measure

The children also completed the PedsQLTM4.0 measure (Varni, Katz, Seid, Quiggins, Friedman-Bender & Castro, 1998). This measure was presented in questionnaire format where items were read aloud to children, and their responses recorded by the interviewer. Parallel child self-report and parent proxy-report versions exist for this measure. There were different versions for each age group: from 5-7 years (young child version); 8-12 years (child version); and 13-18 years (adolescent version). The young children version (5-7 years) was used in this study (see Appendix B for full child measure). This version had been adapted for younger children by Varni et al (1998) who simplified the wording of some items and used a Likert response scale with fewer points.

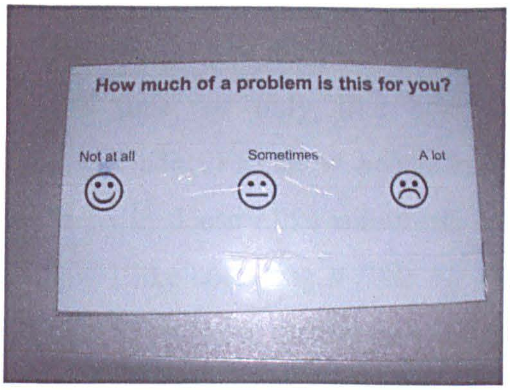
This version of the PedsQLTM4.0 measure consisted of 23 items divided into four domains of functioning: physical (n=8); emotional (n=5); social (n=5); school (n=5). These items and domains are listed in Table 4.2.

Table 4.2: Domains and items in the PedsQLTM4.0 measure

Domain	Item
Physical functioning	Walking
	Running
	Playing sports or exercise
	Picking up big things
	Taking a bath or shower
	Doing chores (like picking up your toys)
	Having hurts or aches
	Feeling too tired to play
Emotional functioning	Feeling scared
	Feeling sad
	Feeling mad
	Having trouble sleeping
	Worrying about what will happen to you
Social functioning	Getting along with other kids
	Other kids saying they don't want to play with you
	Other kids teasing you
	Other kids being able to do things that you cannot do
	Keeping up with other kids when playing
School functioning	Paying attention in school
	Forgetting things
	Keeping up with schoolwork
	Missing school as not feeling good
	Missing school as having to go to doctor's or hospital

Children's answers were given using a 3-point Likert response scale (anchored with: *not at all*; *sometimes*; *a lot of a problem*). The children are given one training item, using smiley faces to help them understand how to use the response scale: "Is it hard for you to snap your fingers?". Figure 4.1 shows the smiley faces used to help children understand the response scale.

Figure 4.1: Photograph of the smiley faces used to aid understanding (PedsQLTM4.0)



This measure has been used successfully to assess QOL in children and their parents, with both healthy and ill populations. Normative data has been published for American children (Varni et al., 2002, Varni, Seid, & Kurtin, 2001). High levels of internal consistency (child: α= .88 total QOL; .80 physical health summary scores; .83 psychosocial health summary scores; parent: α= .90 total QOL; .88 physical health summary scores; .86 psychosocial health summary scores) have been reported by Varni et al. (2002). This measure has also demonstrated clinical validity (total QOL scores distinguish between children on and off treatment), and construct validity (total QOL scores distinguish between healthy and ill children, Varni et al., 2002).

Procedure

As described in Chapter 3, the interviewer spent a 2-hour long session in the classroom in the preceding days helping the children with their schoolwork and establishing rapport with them (see p. 73). The children were interviewed individually in a room separate to the main classroom.

The measure was administered as described in Chapter 3 (see p. 70), with the bears used in the TedQL.1 measure being placed opposite the child, and the experimenter sat adjacent to the child. The children were asked to take part in a game, and they were told: "I am going to ask you to play a game with me. I have two teddy bears that are called Iggy and Ziggy. They look the same but they like doing different things. Look here, Iggy really likes singing so he sits here on this side, and Ziggy really does not like singing so he sits here on the other side. I am going to tell you what Iggy and Ziggy like to do and how they feel about this, and then I am going to ask you what you like to do and you feel. So, if Iggy likes colouring and Ziggy doesn't like colouring, I want you to tell me whether you are the same as Iggy or Ziggy. You may like *really* like colouring all the time, or only like colouring *a little bit.* So Iggy sits here by this (point/circle/face) to show he/she *really* likes colouring, and Ziggy sits here to show he/she *really* doesn't like colouring. I am going to point to this point/smaller circle/face to show I like colouring *a little bit* – not as much as Iggy. See? Which bear are you most like?" (Indicate that they should choose a bear).

The study was explained to the children, and they were asked for their verbal assent. Next the children were shown how to use the response scales during a training period. As described in Chapter 3, during this training period children were given one practise item for the TedQL.1 ("How good are you at singing?", see p. 74). This item was given to assess whether they all understood the task and could use the appropriate response options. All the children responded accurately to the practise item.

The children were then given all 10 items in the TedQL.1 measure. As described in Chapter 3, the two teddy bears were counterbalanced for whether they represented a positive or negative statement (see p. 74). The children were encouraged to restate their choice following selection (e.g., "I am like Ziggy, I have lots of friends to play with, and I feel really happy about this").

The children were given the PedsQLTM4.0 measure (which was administered as directed by Varni et al., 1998) following the completion of the TedQL.1 measure. The children were asked: "Think about how you have been doing for the last few weeks. Please listen carefully to each sentence and tell me how much of a problem this is for you". The children were then given all 23 items in the PedsQLTM4.0 measure. After completing both measures the children were asked which measure they found easier to use and which they found more enjoyable to use. The children were told: "Now I want you to think about the two things that you have just done. I asked you some questions using these teddies bears and these faces, and then I asked you some questions by reading them out loud and using these smiley faces. Which one did you find easier to use – watching the teddy bears or listening to me reading? Which one did you like using the most?". The children chose a sticker at the end of the session to acknowledge their participation.

Scoring

TedQL.1 measure

Children's responses to the items on the TedQL.1 measure were recorded as numerical scores on a response sheet by the interviewer (0=*very sad*; 1=*just a little bit sad*; 2=*just a little bit happy*; 3=*very happy*). The children's answers to the second part of the items (i.e., how they *felt* about this level of ability/functioning) provided their 'happiness' or QOL scores for the TedQL.1 measure. Appendix B gives an example of the response sheet used. The TedQL.1 data was entered into a statistics programme - SPSS version 10. The children's responses were entered to give scores for their 'happiness' with their abilities and functioning (i.e., their QOL). These QOL/happiness scores ranged from 0 to +3. Negatively scored items were reversed so that higher scores represented higher levels of QOL.

PedsQLTM4.0 measure

Children's responses to the items on the PedsQLTM4.0 measure were recorded as numerical scores on the response sheet by the interviewer (0=*not at all*; 2=*sometimes*; 3=*a lot of a problem*). Appendix B gives the response sheet used in this study. The PedsQLTM4.0 data was entered into SPSS. The scores ranged from 0 to +4. The scores for the items were reverse coded and linearly transformed to a 0-100 scale as directed by Varni et al. (2002), so that higher scores represented higher QOL (i.e., 0=0, 1=25, 2=50, 3=75, 4=100).

These scores were used to compute total scores (total QOL: all 23 items); physical health summary scores (PH sub-scale score): 8 physical functioning items; and psychosocial health summary scores (PS sub-scale score): 5 emotional, 5 social, and 5 school functioning items).

4.2.2 Parent-data

Measures

PedsQLTM4.0 measure

Parents were asked to complete the proxy version of the PedsQLTM4.0 measure in relation to their child. As with the children, the young child (5-7 years) version was used. This measure consisted of the same 23 items as the children answered (see Table 4.2, p. 102), however parents answered using a 5-point Likert response scale (0=*never*; 1=*almost never*; 2=*sometimes*; 3=*often*; 4=*almost always a problem*).

Procedure

Parents were sent the PedsQLTM4.0 questionnaire and a letter explaining the study in more detail (see Appendix B for the letter sent to parents). They were asked to complete the questionnaire, and return it to the school.

The parents were given written instructions at the start of the questionnaire as follows: "On the following page is a list of things that might be a problems for your child. Please tell us how much of a problems each one has been for your child during the past one month." (see Appendix B for the full parent questionnaire).

Scoring

PedsQLTM4.0 measure

Parents recorded their answers to the items on the PedsQLTM4.0 measure as numerical scores on the set response sheet provided. Appendix B shows the response sheet given to parents. The PedsQLTM4.0 data was entered into SPSS. The scores ranged from 0 to +5. As with the child data the scores for the items were transformed to a 0-100 scale, and the same total scores were calculated (total QOL; PH sub-scale; PS sub-scale).

4.2.3 Overall treatment of data and statistical analyses

The distributional properties of the children's scores on the TedQL.1 and PedsQLTM4.0 measures, and parents' scores on the PedsQLTM4.0 measure were examined. Assessment of skew and kurtosis were made, using the criterion that if the associated *z* score was above 1.00 the scores were significantly skewed or curved (kurtosis) (Howitt & Cramer, 1997). Kolmogorov-Smirnov tests were used to assess whether the distributions of children's and parents' scores were normally distributed. This test compares any set of scores in a sample to a normally distributed set of scores with the same mean and standard deviation, and shows whether the observed scores differ significantly from this normal distribution (Field, 2000). Where data appeared to be significantly skewed, curved, and/or different from a normal distribution, non-parametric statistics and tests were used (Spearman's rank correlations, Mann-Whitney-U, Wilcoxon and Kruskall-Wallis H tests) as recommended by various researchers (e.g., Conover, 1971, Dancey & Reidy, 1999, Gibbons, 1976, Siegel, 1956). We recognised that non-parametric statistical tests are less powerful than their parametric counterparts (as the analysis is carried out on rank-order data as opposed to the actual data, Field, 2000), but where the data showed significant problems with skew, kurtosis, and/or normality we used non-parametric tests throughout the whole analysis for consistency.

Analysis was conducted between children's TedQL.1 and PedsQLTM4.0 scores, and parent's PedsQLTM4.0 scores, and one independent variable (gender) to consider whether their scores differed systematically by this variable.

The following analyses were conducted to address the specific hypotheses made, in relation to each of three aims of this study:

1) Comparing the psychometric properties of the TedQL.1 to an established measure (PedsQLTM4.0)

a. The internal reliability (or consistency) of the children's responses were assessed using Cronbach's alpha statistics (α) and compared across the two QOL measures (TedQL.1 and PedsQLTM4.0). The internal reliability of parents' responses on the PedsQLTM4.0 were also calculated. Cronbach's alpha assesses the correlations between items in a measure. The reliability of a measure is related to the homogeneity of the items to each other (Breakwell, Hammond, & Fife Schaw, 1995), i.e., the higher the correlations between items, the greater the internal consistency. An internal consistency of above .70 has been recommended as a guideline for 'good' levels of internal consistency (Nunnally, 1978, Cronbach, 1951). We used the standard of .70 in Study 1.

b. The acceptability of the TedQL.1 and PedsQLTM4.0 measures were assessed by asking children which measure they found easiest to use, and which they found more enjoyable to use (see procedure section, p. 104-5). The percentages of children's preferences were compared across the two measures.

2) Investigating the relationship between the TedQL.1 and PedsQLTM4.0 measures

a. The correlations between children's QOL/happiness scores on the TedQL.1 and their total QOL scores on the PedsQLTM4.0 were assessed, using Spearman's rank order correlation coefficients (ρ).

3) Exploring the relationship between child and parent rated child QOL

a. The agreement between child-rated and parent-rated child QOL was assessed. Spearman's correlation coefficients (ρ) were used to assess whether children and parents' QOL scores were correlated to each other. The children reported their QOL using the TedQL.1 and PedsQLTM4.0, and the parents rated their child's QOL using the PedsQLTM4.0.

4.3 Results (Study 1)

4.3.1 Data screening analysis

Item descriptives: means, range and assessment of skew/kurtosis

Children's QOL/happiness scores on the TedQL.1 measure were skewed towards higher QOL, with high levels of kurtosis (see Table 4.3). Children's PS sub-scale scores on the PedsQLTM4.0 measure were also skewed towards higher QOL, with high levels of kurtosis (see Table 4.3). Parents' scores (when rating their child's QOL using the PedsQLTM4.0 measure) were significantly skewed towards higher child QOL (see Table 4.3). Kolmogorov-Smirnov tests showed that parents' PedsQLTM4.0 scores were significantly different from normal (D=0.17-0.21, p <.05, see Table 4.3, see Appendix B for normal Q-Q plots). Non-parametric tests were used for analyses because of these distributions (see Appendix B for full details of skew and kurtosis calculations).

Table 4.3: Descriptives for children's and parents' scores

Measure	N	Mean (SD)	Range	Skew	Kurtosis	Normality test (D)
Children's scores						
TedQL.1:						
QOL/happiness	36	2.43 (0.34)	1.80 – 3.00	0.14	-0.89**	0.14
PedsQL4.0:						
Total QOL	36	72.83 (13.33)	45.65 – 100.00	0.13	-0.68	0.09
PH sub-scale	36	78.82 (14.20)	50.00 – 100.00	-0.26	-0.78	0.13
PS sub-scale	36	69.63 (16.44)	40.00 – 100.00	-0.60	1.29**	0.11
Parents' scores						
PedsQL4.0:						
Total QOL	24	84.28 (10.57)	57.61 – 84.28	-0.80**	-0.02	0.18*
PH sub-scale	24	89.94 (8.64)	65.63 – 100.00	-1.11**	1.30**	0.21*
PS sub-scale	24	81.25 (12.53)	53.33 – 96.67	-0.64**	-0.66	0.17*

*Note. Significance level of skew, kurtosis, and normality: ** p<.01; * p<.05*

Potential item bias: gender

A Kruskall-Wallis H test revealed that there was no effect of gender on children's or parents' QOL scores using either of the two measures (TedQL.1 or PedsQLTM4.0) (see Table 4.4).

Table 4.4: Effect of gender on children's and parents' scores

Gender of child	**Male**		**Female**	
Measure	**n**	**Median (Mean)**	**n**	**Median (Mean)**
Child:				
TedQL.1 measure: QOL/happiness	21	2.50 (2.47)	15	2.30 (2.37)
PedsQL measure: Total QOL	21	69.57 (69.46)	15	78.26 (77.54)
Parent:				
PedsQL measure: Total QOL	12	77.72 (80.34)	12	89.67 (88.22)

Note. (Means are reported in brackets for comparison)

4.3.2 Comparing the psychometric properties of the TedQL.1 to an established measure (PedsQLTM4.0)

Internal reliability (consistency)

The internal consistency of children's and parents' responses on the TedQL.1 and PedsQLTM4.0 measures was calculated using Cronbach's alpha statistics. The children's reports on the TedQL.1 measure produced a low value of consistency, with an alpha value below the .70 standard (see Table 4.5).

Both the children's and the parents' scores on the PedsQLTM4.0 produced good levels of consistency, with alpha values above the .70 standard for total QOL and PS sub-scale (see Table 4.5). These internal reliability values were comparable to those reported by Varni et al. (2002). Overall, the consistency of the children's and the parents' responses were higher when answering using the PedsQLTM4.0 than the TedsQL.1 measure (see Table 4.5).

Table 4.5: Internal consistency (α) of children's and parents' responses

	Children			**Parents**		
Measure	**N**	**No. of items**	**α**	**n**	**No. of Items**	**α**
TedQL.1:						
QOL/happiness	36	10	0.35			
PedsQL4.0:						
Total QOL	36	23	0.75	24	23	0.90
PH sub-scale	36	5	0.44	24	8	0.67
PS sub-scale	36	8	0.74	24	15	0.89

Acceptability

The children were asked which measure they found easier to use and which measure they found more enjoyable to use. 72% (n=26) of the children reported that they found the TedQL.1 measure easier to use, and 89% (n=32) reported finding the TedQL.1 measure more enjoyable to use compared to the PedsQLTM4.0 measure (see Table 4.6).

Table 4.6: Children's individual preferences for measures

	Percentage preferences: % (n)	
Measure	**Easier to use**	**More enjoyable**
TedQL.1	72% (26)	89% (32)
PedsQL4.0	28% (10)	11% (4)

4.3.3 Investigating the relationship between the TedQL.1 and PedsQLTM4.0 measures

The relationship between children's QOL/happiness scores on the TedQL.1 and their total QOL scores on the PedsQLTM4.0 measure was explored using Spearman's correlation coefficients. The children's QOL/happiness TedQL.1 scores were positively correlated with their total QOL PedsQLTM4.0 scores (see Table 4.7).

Table 4.7: Relationship between children's scores on the TedQL.1 and PedsQLTM4.0 measures

	PedsQL4.0: Total QOL	
	N	**ρ**
TedQL.1: QOL/happiness	36	0.33*

*Note. * p<.05*

4.3.4 Exploring the relationship between child and parent rated child QOL

The relationship between child- and parent-rated child QOL was assessed using Spearman's correlation coefficients. First, the relationship between children's and parents' scores was assessed when children were reporting their QOL using the TedQL.1 measure, and parents' were rating their child's QOL using the PedsQLTM4.0 measure. The correlation between children's TedQL.1 scores and parents' PedsQLTM4.0 scores was not significant (see Table 4.8).

Table 4.8: Relationship between children's TedQL.1 scores and parents' PedsQLTM4.0 scores

	Parent: PedsQL4.0: Total QOL	
	n	**ρ**
Child:		
TedQL.1: QOL/happiness	24	-0.10

Second, the relationship between child- and parent-rated child QOL was assessed when children and parents were both using the same measure (PedsQLTM4.0). The correlation between children's and parents' PedsQLTM4.0 scores was significant (see Table 4.9).

Table 4.9: Relationship between children's and parents' PedsQLTM4.0 scores

	Parent: PedsQL4.0: Total QOL	
	n	**ρ**
Child:		
PedsQL4.0: Total QOL	24	0.38*

Note. * *p<.05*

4.4 Discussion (Study 1)

The first aim of this study was to compare the psychometric properties of the new TedQL.1 to an established measure (PedsQLTM4.0). The PedsQLTM4.0 was administered to children verbally in a questionnaire-style format, whereas the TedQL.1 was presented to children in the form of a game using teddy bears as props. We made two predictions in relation to comparing the psychometric properties of these two measures to each other.

First, we predicted that the internal reliability of children's responses on TedQL.1 would be higher than their responses on the PedsQLTM4.0. This prediction was not confirmed. Overall the consistency of the children's responses was higher when answering using the PedsQLTM4.0 (α = .75) than the TedsQL.1 (α = .35) measure (see p. 111).

Second, we predicted that more children would rate the TedQL.1 as the easier and more enjoyable measure to use. This hypothesis was supported. 89% of the children found the TedQL1 more enjoyable to use than the PedsQLTM4.0 (see p. 112). Children clearly preferred the TedQL.1 measure (presented in a game format with props) over the PedsQLTM4.0. Our finding supports the work of researchers such as Ceci & Bruck (1993) and Measelle et al. (1998) who argued that props increase the appeal of measures to children. 72% of the children reported the TedQL.1 as easier to use than the PedsQLTM4.0. This result also supports the argument that props can help clarify material to children (Ernst et al., 1994, Harter & Pike, 1984).

This study considered the relative merits of the TedQL.1 measure with the PedsQLTM4.0 measure. While children preferred using the TedQL.1 measure over the PedsQLTM4.0, the internal consistency of children's responses to the TedQL.1 items were lower than their responses to the PedsQLTM4.0 items (see p. 111). This finding might be explained in terms of the relative length of the two measures. The TedQL.1 measure contained only 10 items (compared to the 23 items in the PedsQLTM4.0) and this may have accounted for the low reliability statistic (α = .35, see p. 111). Some researchers have argued that the length of measures (i.e., number of items) may assist children by giving them time to understand and learn how to respond accurately to the questions, in turn leading to more reliable responses from children (e.g., Davis-Kean & Sandler, 2001, Marsh et al., 1991, Marsh et al., 1998). For example, Marsh et al. (1998) found that the items at the end of their self-concept measure (the SDI) were psychometrically stronger than those at the start. Marsh et al. (1998) findings are contrary to the view that the quality of children's responses will deteriorate on items towards the end of instruments. Marsh et al. (1998) have suggested that the use of short instruments may be counterproductive, and argued that their results may account for why researchers have had difficulties obtaining responses from young children with good psychometric properties.

As the children's reports on the TedQL.1 measure produced a low value of consistency, with an alpha value below the .70 standard (α = .35, see p. 111), we expanded the number of items in our measure for Study 2 (from 10 to 23 items, see Table 4.10, p. 121) to produce a more reliable self-report instrument. These additional items were based on comments children had spontaneously volunteered in Study 1 about what they liked to do, who they played with, and who was important to them (recorded anecdotally by researcher). A new version of the TedQL measure was developed for use in Study 2 (the TedQL.2).

As the TedQL.1 was expanded to include more items, we felt that children could only cope with answering one main question for each item in the new version. In the TedQL.1, children were first asked which bear was *most like* them, and then they were asked how they *felt* about this level of functioning/ability. We changed the TedQL.2 to ask children to rate their ability/functioning levels on the 4-point scale, using the teddy bears to help understand the rating task. The TedQL.1 had used a faces scale to rate their feelings (happiness) about any given activity, and it was noted by the researcher that some of children in Study 1 had expressed confusion on the second part of the question. We felt that the idea of individual preferences could still be incorporated within a later version of the TedQL following the completion of Study 2.

We need to compare the psychometric properties we obtained using the PedsQLTM4.0 measure to existing published studies using this measure. The internal consistency of children's responses to the PedsQLTM4.0 items was lower than the values reported in previous published studies using this measure (young child version). For example the alpha values reported by Varni, Seid, and Kurtin (2001) for healthy children are shown in brackets against the values we found in Study 1: total QOL: α = .75 (.88); PH sub-scale: α = .44 (.80); PS sub-scale: α = .75 (.83). However the internal consistency of parent's responses to the PedsQLTM4.0 items were comparable (for total QOL and PS sub-scale) to the values reported by Varni et al. (2001): total QOL: α = .90 (.90); PH sub-scale: α = .67 (.88); PS sub-scale: α = .89 (.86). We were unsure why our alpha values for children's responses were lower than those reported by Varni et al. (2001) for the young children version of the PedsQLTM4.0. However, we recognised that we used different samples of children to those used by Varni et al. (2001), specifically we were testing in the U.K. and Varni et al. (2001) collected data from American children.

The second aim of this study was to investigate the relationship between the TedQL1 and the PedsQLTM4.0 measures. This aim was achieved by assessing whether children's TedQL.1 scores were correlated with their PedsQLTM4.0 scores. Our prediction, that children's TedQL.1 scores would be positively correlated to their PedsQLTM4.0 scores, was confirmed. The children's scores were correlated across these two measures (ρ= .33, see p. 112). This result confirms that the PedsQLTM4.0 and the TedQL.1 were measuring a similar construct (i.e., QOL), because scores on both measures were correlated across children's ratings. The results showed that our new TedQL.1 measure was related to an established measure of QOL (PedsQLTM4.0), and therefore we hypothesised that both instruments were measuring a similar construct (i.e., QOL).

The third aim of this study was to explore the relationship between child and parent rated child QOL. As the TedQL.1 measure did not have a parent report version we assessed parent-child agreement when children and parents were using different measures (TedQL.1, PedsQLTM4.0), and also when children and parents were using the PedsQLTM4.0 measure (PedsQLTM4.0). We predicted that children's and parents' scores would not be correlated with each other across the two measures, and also when using the same measure. Our prediction was partly confirmed. Children's TedQL.1 scores were not correlated to parents' PedsQLTM4.0 scores (ρ= -.10, see p. 113). However when children and parents were both using the PedsQLTM4.0 measure to rate child QOL, their scores were correlated to each other (ρ= .38, see p. 113). Our results showed that when children and parents are using the same measure to rate child QOL their scores are more likely to be related to each other, than when using two different measures to rate child QOL.

4.5 Introduction (Study 2: Further validation of the TedQL.2 measure)

4.5.1 Investigation the impact of response scale type on children's TedQL.2 responses

As discussed in the review in Chapter 3 (see p. 47 & p. 65), research from the pain literature has shown the value of comparing different response scales to establish the 'best' format for child measures (e.g., Chambers & Craig, 1998, Goodenough, Addicoat, Champion, McInerney, Young, Juniper, & Ziegler, 1997). Chambers and Craig's (1998) findings suggest that pain ratings vary depending on the type of faces scale used, and also note that the ability of children below five years to use faces as a response option has not been well established. Some researchers have begun to consider these issues in relation to child self-reported health below eight years (e.g., Chambers & Johnston, 2002, Rebok et al., 2001, Riley et al., 2004). Rebok et al. (2001) investigated the effect of response scale type for a self-report health measure. They presented evidence that children below eight years old understood graphic response scales (such as circles of varying sizes) better than linear response scales (Rebok et al., 2001).

We felt that the impact of response scale type needed to be considered in the development stages of our new child self-report measure (the TedQL.2). Therefore the first aim of Study 2 was to investigate whether scale type would impact on children's TedQL.2 responses. This aim was achieved by comparing the internal reliability of children's TedQL.2 responses across two different types of response scales. We chose two scales that had been used in previous child self-report measures (see Table 3.2 in Chapter 3, see p. 75-90). We chose a circles (graphic) scale (Harter & Pike, 1984, Rebok et al., 2001) and a linear (visual analogue) scale (Hicks et al. 2001, Koopman, Kamphuis, Verrips, Vogels, Theunissen, Fekkes, Verloove-Vanhorick, & Wit, 1997, Thomas, Goodneough, Von Baeyer, & Champion, 1997).

Based on the findings of Rebok et al. (2001), we predicted that children using the circles response scale to answer the TedQL.2 items would produce responses with higher internal reliability (consistency), and take less time to complete the items compared to children using the linear response scale.

4.5.2 Further exploration of the relationship between child and parent rated child QOL

The second aim of this study was to explore the relationship between child and parent rated child QOL. As in Study 1 (due to the lack of a proxy report version of the TedQL measure), we assessed parent-child agreement when children used the new version of the TedQL (TedQL.2) and parents used the PedsQLTM4.0. In Study 2 we assessed child-parent agreement with an even younger age group than in Study 1 (3-5 year olds compared to 5-8 year olds).

Based on the literature reviewed in Study 1 (p. 98) and the results of Study 1 (p. 113), we predicted that children's and parents' QOL scores would not be correlated when rating child QOL.

4.5.3 Summary of the aims and predictions of Study 2

The first aim of Study 2 was to investigate the impact of scale type of children's TedQL.2 responses. We predicted that children using the circles response scale to answer the TedQL.2 items would produce responses with higher internal reliability (consistency), and take less time to complete the items, compared to children using the linear response scale.

The second aim was to explore the relationship between child and parent rated QOL. We predicted that children's and parents' child QOL ratings would not be correlated.

4.6 Methodology (Study 2)

Sample

Ethics approval was obtained as in Study 1 (p. 100), and 32 participants aged 3-5 years were identified from a nursery in Sheffield. Their parents were given a letter explaining the study. Their parents were asked to complete a permission slip for their child to take part. Three children were excluded from the analyses because they completed fewer than 50% of the study questions, and one child was absent on the assessment day. Twenty-eight children (15 females and 13 males; 3.0-5.0 years) completed the study. The mean age of the children was 3.88 years (SD= 0.62 years). Twenty-four (86%) of the children were Caucasian and four were of Asian origin (14%).

Questionnaires for parent completion were sent home with all children who had participated in the study (n=24). Twenty-one parents returned their questionnaires giving 21 parent-child dyads for this study.

4.6.1 Child-data

Measures

TedQL.2 measure

The children were interviewed using the two identical teddy bears used in Study 1, which where referred to as either female or male depending on the sex of the child (see Chapter 3, Figure 3.3, p. 70).

The TedQL.2 measure consisted of 23 items within 5 domains/areas. These are listed in Table 4.10.

Table 4.10: Domains and items in the TedQL.2 measure

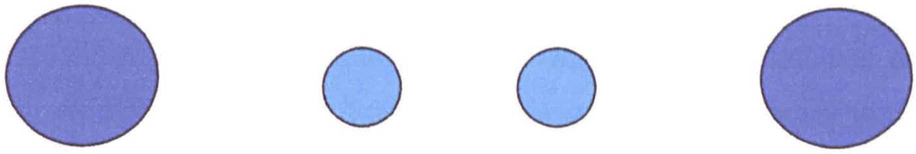
Domain	Items
Physical Competence	PC1: good at swinging PC2: good at running PC3: good at playing with balls PC4: good at climbing high things PC5: can tie shoes PC6: good at hopping
Peer Acceptance	PA1: has lots of friends PA2: likes playing with friends PA3: likes to tell friends what to do (N) PA4: has friends to play with
Maternal Acceptance	FA1: likes to play with mum FA2: mum talks to them FA3: likes to tell mum what been doing FA4: mum tells them off a lot
Psychological Functioning	PF1: is happy PF2: has bad dreams at night PF3: worries about losing their things (N) PF4: some things make them really cross/angry (N)
Cognitive Functioning	CF1: gets upset if can't do work (N) CF2: remembers what people tell them to do CF3: good at reading CF4: good at writing CF5: getting better at drawing

Note. (N)= negatively scored item, scores were reversed.

As in Study 1 a forced recognition task was used where the bears were first described and then children were to choose which bear was *most like them.* The children were then probed for whether they were *really* like this or *just a little bit.* The children were asked to think about how they had been during the last week when answering the items. Responses were made using one of two 4-point response scales (circles or linear). Examples of the two types of response scales used are shown in Figure 4.2.

Figure 4.2: Examples of the two response scales (TedQL.2)

Circles:



Linear:

Procedure

The interviewer spent two 3-hour long sessions in the nursery in the preceding days playing with the children and establishing rapport with them. The children were interviewed individually in a separate room. The same set-up and protocol was used as in Study 1, where the bears were placed opposite the child and they were asked to take part in a game (see p. 103-4). Thirteen of the children were assigned to the circles response scale, and 15 to the linear scale. The children were randomly assigned to two groups by their nursery teacher, and these groupings were used to determine which children completed the TedQL.2 with which response scale.

The children were first asked for their verbal assent to the task, and then the children were shown how to use the response scales during a training period. During this training period children were given one practise item (how much they liked doing colouring) and one hypothetical question (how much they liked their favourite sweet). These were used to assess whether they all understood the task and could use the appropriate response options.

The children were then given all 23 items in the TedQL.2 measure. The two teddy bears were counterbalanced for whether they represented a positive or negative statement. As in Study 1 the children were encouraged to restate their choice following selection (see p. 104). The children chose a sticker at the end to acknowledge their participation in the study.

Scoring

TedQL.2 measure

Children's responses to the items on the TedQL.2 measure were recorded as numerical scores on a response sheet by the interviewer (0=*really bad/not at all good/dislike*; 1=*just a little bit bad/not good/dislike*; 2=*just a little bit good/like*; 3=*really good/a lot good/always good/like*). Appendix C gives an example section of the response sheet used for this study. The QOL scores ranged from 0 to +3. This raw data was entered into SPSS. Negatively scored items were reversed so that all items were scored such that higher scores represented higher levels of QOL. These scores were used to calculate mean QOL (total scale) scores for all the children.

4.6.2 Parent-data

Measures

PedsQLTM4.0 measure

Parents were asked to complete the proxy toddler (3-5 years) version of the PedsQLTM4.0 measure in relation to their child. This version of the measure consisted of the same items as the parents answered in Study 1 (see Table 4.2, p. 102), however, it contained 21 items (as opposed to 23 items). The items in the physical functioning, emotional functioning and social functioning domains were the same as the young child (5-7 years) version used in Study 1. The difference in the toddler version was that the school functioning domain was changed to nursery functioning, with two items removed ("paying attention in school" and "forgetting things"), and one item being re-worded from "keeping up with schoolwork" to "doing the same activities as peers". The parents' answers were given using a 5-point Likert response scale (anchored with: *never*; *almost never*; *sometimes*; *often*; *almost always a problem*).

Procedure

Parents were sent the PedsQLTM4.0 questionnaire and a letter explaining the study in more detail (see Appendix C for the letter sent to parents). They were asked to complete the questionnaire, and return it to the nursery (see Appendix C for the full parent questionnaire).

Scoring

PedsQLTM4.0 measure

Parents recorded their answers to the items on the PedsQLTM4.0 measure as numerical scores on the set response sheet provided in their questionnaire packs (in the same way as Study 1, see p. 107). Appendix C shows the response sheet given to parents. This raw data was entered into SPSS. As in Study 1 the scores for the items were transformed to a 0-100 scale, and total scores were calculated (total QOL; PH sub-scale; PS sub-scale).

Overall treatment of data and statistical analyses

The distributional properties of the children's scores on the TedQL.2 measure, and parents' scores on the PedsQLTM4.0 measure were examined. Assessment of skew and kurtosis were made using the same criteria as Study 1 (see p. 106). Normality testing was also carried out using Kolmogorov-Smirnov tests, as in Study 1 (see p. 106-7). Where data appeared to be significantly skewed, curved and/or different to a normal distribution, non-parametric statistics and tests were used throughout the analysis, as with Study 1 (see p. 106-7). Analysis was conducted between the children's TedQL.2 scores, and the parents' PedsQLTM4.0 scores, and independent variables (age and gender) to consider whether their scores differed systematically by these variables.

The following analyses were conducted to address the specific hypotheses made, in relation to the aims of this study:

1) Investigating the impact of response scale type on children's TedQL.2 responses

a. The internal reliability of the children's responses on the TedQL.2 measure was assessed using Cronbach's alpha statistics (α), and compared across the two response scale types (circles and linear).

b. The mean time taken for children to complete the TedQL.2 measure was calculated and compared across the two response scale types.

2) Exploring the relationship between child and parent rated child QOL

a. The internal reliability of the parent's responses on the PedsQLTM4.0 measure was assessed using Cronbach's alpha statistics (α), to ensure the reliability estimates for this measure were above the .70 standard used in Study 1 (see p. 107).

b. The agreement between child-rated and parent-rated child QOL was assessed. Spearman's correlation coefficients (ρ) were used to assess whether the QOL mean scores were correlated. The children reported their QOL using the TedQL.2, and the parents rated their child's QOL using the PedsQLTM4.0.

4.7 Results (Study 2)

4.7.1 Data screening analysis

Item descriptives: means, range and assessment of skew/kurtosis

Children's mean scores on the TedQL.2 measure were skewed towards higher QOL, with high levels of kurtosis (see Table 4.11). Parents' scores, when rating their child's QOL using the PedsQLTM4.0 measure, were significantly skewed towards higher QOL, with high levels of kurtosis for total QOL and PS scores (see Table 4.11). Normality tests showed that parents' PedsQLTM4.0 scores were significantly different from normal for total QOL (D=0.11, p<0.05) and PS scores (D=0.19, p< .05, see Table 4.11, see Appendix C for normal Q-Q plots). As in Study 1 non-parametric tests were used for analyses because of these distributions (see Appendix C for full details of skew and kurtosis calculations).

Table 4.11: Descriptives for children's and parents' scores

Measure	N	Mean (SD)	Range	Skew	Kurtosis	Normality test (D)
Children's scores						
TedQL.1:						
Total scale	28	2.11 (0.29)	1.39 – 2.70	-0.27	0.57	0.12
Parents' scores						
PedsQL4.0:						
Total QOL	21	81.24 (9.57)	64.69 - 96.43	-0.11	-0.78	0.11*
PH sub-scale	21	83.93 (9.54)	65.63 - 100.00	-0.36	-0.19	0.12
PS sub-scale	21	79.58 (10.58)	63.46 - 98.08	-0.04	-1.03**	0.19*

*Note. Significance level of skew, kurtosis, and normality: ** p<.01; * p<.05*

Potential item bias: age

Spearman's correlation coefficients revealed that age was not correlated with the children's TedQL.2 scores or parents' PedsQLTM4.0 scores (see Table 4.12).

Table 4.12: Relationship between chronological age, and children's and parents' mean scores

	Chronological age	
Measure	n	ρ
Child:		
TedQL.2 measure:		
Total scale	28	-0.13
Parent:		
PedsQL4.0 measure:		
Total QOL	21	0.01

Potential item bias: gender

A Kruskall-Wallis H test revealed that there was no effect of gender on children's or parents' QOL scores using either of the two measures (TedQL.2 or PedsQLTM4.0, see Table 4.13).

Table 4.13: Effect of gender on children's and parent's scores

Gender of child	**Male**		**Female**	
Measure	**n**	**Median (Mean)**	**n**	**Median (Mean)**
Child:				
TedQL.1 measure:				
Total scale	13	2.00 (2.02)	15	2.09 (2.20)
Parent:				
PedsQL measure:				
Total QOL	10	84.52 (83.69)	11	80.95 (79.00)

Note. (Means are reported in brackets for comparison)

4.7.2 Investigating the impact of response scale type on children's TedQL.2 responses

Internal reliability (internal consistency)

The internal consistency of children's responses was calculated using Cronbach's alpha statistics, and compared across the two response scales (circles and linear). The children's responses when using the circles (graphic) scale had a higher internal consistency than when using the linear (visual analogue) scale (see Table 4.14).

Table 4.14: Internal consistency (α) of children's responses on the TedQL.2 measure across the two response scales

	Response scale			
	Circles		Linear	
Measure	No. of items	α	No. of items	α
TedQL.2:				
Total scale	23	0.70	23	0.48

Time taken

The median time children took to complete the TedQL.2 measure was calculated, and compared across the two response scales. A Mann-Whitney U test revealed that the time taken to complete the task was longer with the linear response scale, than with the circles scale (see Table 4.15).

Table 4.15: Effect of response scale type on time taken to complete the TedQL.2 measure

	Response scale			
	Circles		Linear	
Measure	n	Time taken (mins): Median	n	Time taken (mins): Median
Child:				
TedQL.1 measure:				
Total scale	13	8.00	15	12.00*

Note. * *p<.05*

4.7.3 Exploring the relationship between child and parent rated child QOL

Internal reliability (consistency)

The internal consistency of parents' responses on the PedsQLTM4.0 measures was calculated using Cronbach's alpha statistics.

The parents' scores on the PedsQLTM4.0 produced good levels of consistency, with alpha values above the .70 standard for total QOL and PS sub-scale (see Table 4.16). These internal reliability values were comparable to those reported by Varni et al. (2002).

Table 4.16: Internal consistency (α) of parents' responses on the PedsQLTM4.0

	Parents		
Measure	n	No. of Items	α
PedsQL4.0:			
Total QOL	21	21	0.86
PH sub-scale	21	8	0.70
PS sub-scale	21	13	0.80

The relationship between child and parent rated child QOL was assessed using Spearman's correlation coefficients. In this section of the data children were reporting their QOL using the TedQL.2 measure, and parents were rating their child's QOL using the PedsQLTM4.0 measure. The correlation between children's and parents' QOL scores was not significant (see Table 4.17).

Table 4.17: Relationship between children's TedQL.2 scores and parents' PedsQLTM4.0 scores

	Parent: PedsQL4.0: Total QOL	
	n	ρ
Child:		
TedQL.2: Total scale	21	0.22

4.8 Discussion (Study 2)

The first aim of this study was to investigate the impact of response scale type on children's responses on the TedQL.2 measure. Children completed the TedQL.2 items using one of two different response scales (circles i.e., graphic, or linear i.e., visual analogue). The internal reliability of children's responses and the time taken to answer the items was then compared across these two response scales.

We predicted that children using the circles response scale to answer the TedQL.2 items would produce responses with higher internal reliability. This hypothesis was confirmed by the results of this study. The children's responses when using the circles scale had higher internal consistency (α= .70) than children using the linear scale (α= .48, see p. 128).

Our second prediction, that children using the circles response scale would take less time to complete the TedQL.2 items compared to children using the linear scale, was also supported. Children using the circles response scale took less time to complete the TedQL.2 items (Mdn = 8.00 minutes), than children using the linear scale (Mdn = 12.00 minutes, see p. 128). Taking less time to answer the items using one response scale type over another could be judged as evidence that children found the measure easier to use when given this response scale to answer items with. However it does not mean children *understood* the items better when they took less time to complete them.

These results support Rebok et al.'s (2001) argument that graphic response scales may aid children in a rating task. The results of Study 2 also showed that response scale type did impact directly on the internal consistency of children's responses to items in the TedQL.2 measure. We argued that the findings of Study 2 needed following up - with larger sample sizes, different response scales (such as facial response scales), and a repeated measures design. For these reasons further investigation of the impact of response scale type on other psychometric properties of the TedQL measure (such as reproducibility) was carried out in Study 4 (see Chapter 8, p. 209-11).

The second aim was to explore the relationship between child and parent rated child QOL. This aim was achieved by assessing whether child-rated QOL (using the TedQL.2) was correlated with parent-rated child QOL (using the PedsQLTM4.0). Based on the literature discussed in Chapter 1 (see p. 5-6), we predicted that that children's and parents' scores would not significantly correlated across these two measures. The results supported our prediction revealing no significant relationship between children's and parents' ratings of child QOL ($\rho = .22$, see p. 129). However, it should be noted that children and parents were using two different measures to rate child QOL. In Study 1 we found that when children and parents were using the same measure (i.e., the PedsQLTM4.0) their scores were correlated to each other (see p. 113). Therefore the results of these two studies have shown that children and parents need to rate child QOL using the same measure, when researchers are assessing the relationship between child and parent rated child QOL.

The TedQL.1 and TedQL.2 versions of our measure did not have a parent report version to allow comparisons between proxy and child self-reports directly. It would be preferable to develop a parent report version of the TedQL for direct comparisons between child and parent reports. We therefore argued that further development of the TedQL measure should include the production and validation of a parent version of the TedQL. For these reasons we developed a parent report version of the TedQL measure in Study 5 (see Chapter 10, see p. 277).

As in Study 1 we need to compare the psychometric properties obtained in Study 2 for the PedsQLTM4.0 measure to existing published studies using this measure. The internal consistency of parent's responses to the PedsQLTM4.0 items was comparable to the values reported in previous published studies using this measure (Varni et al., 2002, report values for toddler version). The alpha values reported by Varni et al. (2002) are shown in brackets against the values we found in Study 2: total QOL: $\alpha = .86$ (.77); PH sub-scale: $\alpha = .70$ (.75); PS sub-scale: $\alpha = .80$ (.76).

4.9 General discussion

Studies 1 and 2 considered the relative merits of two measures of child QOL (the TedQL.1 & 2, & the PedsQLTM4.0). The PedsQLTM4.0 showed good levels of internal reliability across both child and parent reports within both studies. The alpha values were comparable to those in published studies using this measure (see p. 115 & p. 131). Although the psychometric properties of the TedQL.1 and 2 were lower the PedsQLTM4.0 measure, the TedQL instrument was found to be more acceptable to young children. Study 1 showed that children preferred using the TedQL (where the items were presented using teddy bears as props as opposed to verbally), and also reported that they found this instrument easier to use (see p. 112). Considering the PedsQLTM4.0 has been developed over more than fifteen years, it was encouraging that our relatively new measure (in its early stages of development) stood up so well in comparison.

These two studies also investigated the relationship between child and parent rated child QOL. The results showed that children's and parents' ratings are more likely to be related when they are rating child QOL using the same measure. In Study 1 when children and parents were both using the PedsQLTM4.0 measure, their reports were correlated with each other (see p. 113). However when children and parents were using different measures (i.e., children using the TedQL and parents using the PedsQLTM4.0) their reports were not correlated with each other (see p. 113 & 129). This result was consistent across both Studies 1 and 2. We proposed that future versions of the TedQL would include a parent report version to allow for assessment of both perspectives (see Study 5, Chapter 10, p. 272-3).

Further development of the content of the items in the TedQL.2 measure was necessary to ensure all the items included were relevant, appropriate and understandable to children below eight years. Our review of child self-report measures in Chapter 3 revealed that the most common way to generate items was to use information from children themselves (see p. 53-4). The content of our measure should be child-generated, as opposed a downward extension of a measure originally developed for older children, or containing items developed by a panel of 'professionals'. We wanted to use information directly from children themselves to inform the content development of our measure. For these reasons further development of the TedQL.2 items was carried out in Study 3 (see Chapter 5).

This measure could be criticised for lacking a clear theoretical model. A theoretical model helps to clarify the concepts that should be assessed within measures (Eiser & Morse, 2001). As discussed in Chapter 1 (p. 10-11) the absence of a theory means there is no clear way of distinguishing which factors are relevant to measurement, and makes it difficult to test the construct validity of any measure (Wallander, 1992). However there has been little empirical work on developing a theory of QOL relevant to children below eight years, or examining the appropriateness of adult or adolescent theories of QOL for younger children (Wallander, 2001). We needed to incorporate a theory of QOL our TedQL measure. One such model that could be applied to young children's QOL was based on the idea of an individual's QOL being equal to discrepancy between their 'ideal' and their 'actual' self (Bergner, 1989, Calman, 1987, see Chapter 1, p. 9). As discussed in Chapter 1 (see p. 11-12), this model relies on individual's making judgements of how much their current situation, abilities and functioning (actual self) differs from how they would like it to be (ideal self).

The literature reviewed in Chapter 2 showed that children below eight years are capable of forming and holding self-concepts (p. 24-6), and also they can make social comparisons to peers, family and other caregivers at this age (see p. 26-8). Based on this evidence we felt that children below eight years would be capable of using a QOL measure requiring them to report on both their 'actual' and 'ideal' selves. We incorporated the discrepancy model into the TedQL - specifically into the versions used in Studies 4 and 5 (see Chapter 8, p. 220 & Chapter 10, p. 274). This model allowed individual preferences for functioning and abilities to be assessed in further versions of our measure, similar to the 'happiness' scores produced by the TedQL.1 in Study 1.

The following chapter (Chapter 5) reports further development of the content of items in our TedQL measure using an interview method.

Chapter 5: Developing the content of the TedQL.2 measure - an interview method (Study 3).

Summary

Aims

Study 3 aimed use child interview data to directly inform the content development of items in the TedQL.2 measure.

Sample and method

Eighty-nine children (3.0-8.5 years) were interviewed using a pre-determined interview schedule facilitated by a storybook about a dog named Bruce. The children's answers were recorded, transcribed verbatim, and coded into 10 main themes. Content analysis was used to produce frequency data, giving information about what were the most frequently mentioned aspects of young children's lives.

Results and discussion

The older children (5.0-8.5 years) were more able to answer the questions, and provided detailed answers. The younger children (3.0-4.5 years) had difficulty with some topics covered by the interview schedule. The children were generally positive about their home and school lives. Arguing and making up with friends, siblings, and parents was a common part of life for these children. Children across all the ages were able to talk about their emotions, and could provide examples of situations that had made them feel these emotions. Based on the results of Study 3, a new version of the TedQL measure was developed containing 30 items divided into five areas (the TedQL.3). Six items were deleted, 17 items were retained with 13 of these being altered slightly, and 13 additional items were added to the TedQL.2 measure.

5.1 Introduction

Studies 1 and 2 explored the feasibility of a new child self-report QOL measure (the TedQL.1 & 2, see Chapter 4). However as discussed in Chapter 4 (see p. 132-3), both the content and format of this new measure warranted further development and investigation. Therefore in this study we considered the content development of items for the TedQL measure.

5.1.1 Potential value of qualitative methods for developing self-report items

Over the last ten years, there has been increasing interest in using qualitative methods in psychological research (Fiese & Bickham, 1998). Information obtained from such methods can help understand individual perceptions of issues that may have been overlooked in previous work (Fiese & Bickham, 1998). Interviewing children themselves about their thoughts and feelings on the concepts to be measured, and using this data to inform the content of items can help to avoid developing instruments containing items that are essentially meaningless to young children (Eiser, 1997, Ronen, Rosenbaum, Law, & Streiner, 1999). This approach can also help in establishing content validity for measures (McLaughlin & Bjornson, 1998) by ensuring items within measures are relevant to children themselves (Ronen et al., 2001).

Bradlyn, Ritchey, Harris, Moore, O'Brien, Parson, Patterson, and Pollock (1995) recommended that items in measures should be derived from the population for which the tool is to be used. Our review of self-report measures in Chapter 3 (see p. 53-5) revealed that the most common way to generate items for child instruments was to use information from children themselves. Researchers have used children's spontaneous comments from pilot work (e.g., Beyer & Arandine, 1988, Chapman & Tunmer, 1995, Fogel Keck et al., 1996), or have asked children to list, or talk about, relevant topics (e.g., Collier et al., 2000, Juniper et al., 1998, Hester, 1990). Other researchers have used interviews or focus groups to gain information from children for the items included in their measures (e.g., Damon & Hart, 1988, Ernst et al., 1994, Lewis-Jones & Finlay, 1995, Neff & Dale, 1990, Quittner et al., 2000, Rebok et al., 2001, Ronen et al., 1999). For example, Ronen et al. (1999) used a modified focus group technique with six to twelve year olds to develop a pool of items for their epilepsy HRQOL measure. Ronen et al. (1999) argued that qualitative methods have been under-utilized in health research, and their work serves to highlight the potential value of such methods for the developing the content of child self-report measures (Ronen et al., 1999).

5.1.2 Aims of Study 3

Study 3 was designed to interview healthy children aged three to eight years about their thoughts and feelings in relation to their abilities, behaviour, and their school and family lives. The aim of Study 3 was to use the interview data used to directly inform the content development of items in the our child-self report QOL measure (the TedQL.2). The information gained from the child interviews was subjected to content analysis to identify what areas were most important in the children's lives. The results were used to make decisions on which items should be removed, altered, or added to our TedQL measure.

5.2 Methodology

Sample

Ethics approval was obtained as in Study 1 (p. 100), and 90 participants aged 3-8 years were identified from a primary school and attached nursery in Faversham, Kent. Their parents were given a letter explaining the study. Their parents were asked to complete a permission slip for their child to take part in the study. Eighty-nine children (55 females and 34 males) completed the study. One child was away from school and could not take part. The children were taken from three age groups, Nursery (3.5-4.5 years; n=22), Year 1 (5.0-6.5 years; n=30), and Year 3 (7.0-8.5 years; n=37). The mean age of the children was 6.59 years (SD= 1.64 years). Eighty-seven of the children were Caucasian (98%), and two were of Asian origin (2%).

Measures

Description of the 'Bruce the dog' storybook

Children were interviewed using a storybook developed specifically to facilitate the interviews, focused around a dog called Bruce and his life, friends, family, and feelings. This book was based on material and illustrations adapted from a book originally for use with children who have been taken into care by the social service: *Bruce's Story* (Thom & Macliver, 1993). It was originally developed to help children understand what was happening in their lives, and to cope with the emotions that they may be experiencing. Permission was obtained from the author and illustrator to reproduce some illustrations from this book. We adapted the wording for use in a normal school setting to interview children about their lives, behaviour, friends, families, and feelings.

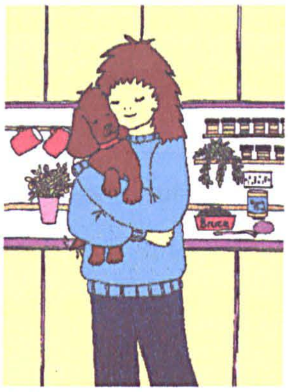
The adapted storybook consisted of ten pages covering the following ten themes:

i. personality/self-esteem
ii. about your home
iii. activities/hobbies
iv. school life
v. peer relationships
vi. family relationships
vii. feelings and thoughts
viii. earliest memories
ix. new things learning to do

x. things want to have/do

Each page contained a coloured illustration of Bruce accompanied by a set of statements relevant to each illustration, which could then be used to ask children about their lives in comparison to what 'Bruce the dog' had just told them. Figure 5.1 gives an example of a page from the storybook.

Figure 5.1: Example of an illustrated page from 'Bruce the dog' storybook



'My mum is a very special person to me. Sometimes I am naughty and she tells me off. When I make her angry it does not last long. I like it best when she gives me a cuddle and says how pleased she is that I am her dog'

(See Appendix D for full example of interview schedule and storybook).

Previous work has shown that young children may be unwilling to say they resemble puppets that were perceived to be of the opposite sex (Eder, 1990). Therefore in this study 'Bruce the dog' was referred to as either female or male compared to the sex of each child.

An interview schedule was developed, with set prompts to accompany each section in the storybook, for example the prompts used to follow the page above in Figure 5.1 were:

'Tell me about your family... What do you like to do with your mum/dad/other? Do you tell them what you have been doing at nursery/school? Do you feel you can tell them things about you? Do you argue much with them? Who else do you see in your family?'

(See Appendix D for interview schedule with all prompts used).

This same interview schedule was used with each child to ensure continuity across interviews. The semi-structured nature of the interview gave the children freedom to discuss other issues of importance to them (i.e., issues that may not have been covered by the interview schedule).

Procedure

The experimenter spent three sessions in the school in the preceding days working with the children. The children were then taken individually to a quiet area away from the classroom, and sat at a table with the experimenter. The study was briefly explained to each child, and verbal assent was gained. The story was read out loud to each child. Each child was then given time to respond verbally to each section with thoughts and feelings about their lives. The children's answers were followed up with the prompts for each section to gain as much information from them as possible. The interviews lasted between 10 and 15 minutes with each child.

Treatment of data and coding

Children's responses were audio taped and transcribed by the author. The interviews served as raw data for the content analysis. To preserve the confidentiality of the children all names mentioned by the children have been changed or omitted. All coding was conducted by the author. The interviews were analysed using a content analysis technique. A coding framework was developed - based on the ten themes set out by the pages of the storybook used to interview the children (see Table 5.1).

Table 5.1: Coding framework for the child interviews

Area	Coding theme	Details of coding
1	Personality	• Quiet/loud person • Reasons for being loud/quiet • Clever person • Reasons for being clever or not
2	About your home	• Big/small house • Garden • Like house • Own room/share • If share, want own room?
3	Activities or hobbies	• Types of games
4	School life	• Like/dislike going to school • Favourite/worst subject and reasons for this

5	Peer relationships	• Lots or a few friends/best friend • Reasons for having 'best friend' • Differences between types of friends • If bullied or teased by friends • Argue with friends reasons for this, and how make up again • Friends come over/ visit theirs, and if not, reasons for this
6	Family relationships	• Mentioned parents, if they play with them, and if not would they like them to do so • If parents estranged/divorced, how they deal with this • If argue with parents, and reasons for this • Whether have siblings, what play with them, and if not, reasons for this • Argue with siblings and reasons for this • Mentioned grandparents or cousins, and what they do with them • Mentioned pets
7	Feelings or thoughts	• If they get: - Scared - Cross or angry - Sad - Happy • If gave examples of what makes them feel this way
8	Earliest memory	• What their earliest memory was and if patterns in types of memories
9	New things learning	• What new things they were learning, and who was teaching them
10	Things want to have/do	• What things want to have or want to be able to do, which do not currently have

A paper copy of each transcript was read through a number of times until it was familiar to the author. Excerpts that related to each of the ten themes were highlighted in the paper copy, and using a word document version of the transcripts each excerpt from each transcript was 'copied and pasted' into a separate word table (see Table 5.1 in Appendix D for example of coding table). These tables were used to produce frequency data to give information about what were the most commonly mentioned aspects of the children's lives. The data was reported in relation to these ten areas in the results (i.e., by theme).

5.3 Results

5.3.1 Interviewing children about their thoughts and feelings on their lives (content analysis)

Personality/self-esteem

Table 5.2: Children's responses relating to their personality

	Total (N=89) % (n)	Nursery (n=22) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
What type of person are you?				
Loud	23 (21)	27 (6)	30 (9)	16 (6)
Quiet	45 (40)	50 (11)	50 (15)	38 (14)
Both	26 (23)	9 (2)	20 (6)	41 (15)
Don't know	6 (5)	14 (3)	0 (0)	5 (2)

The children were asked about what sort of person they thought they were. As shown by Table 5.2, nearly half of all the children reported that they were usually quiet. However the children gave different reasons for their behaviour across the ages:

"*I always, in the night, I always quiet*" (girl; 4.3 years)
"*When I do work I am quiet*" (girl; 6.2 years)
"*I am quiet when my daddy smacks me, when I am naughty*" (girl; 6.4 years)
"*Quiet... I am just a quiet person*" (girl; 8.2 years)
"*Quiet...um coz I do hard work*" (girl; 8.4 years)

A quarter of the children answered that they were sometimes loud and sometimes quiet (n=23, see Table 5.2). The reasons they gave for this included comments like:

"*Think both... (loud) when my brother takes my toys... (quiet) when I'm doing my work*" (boy; 8.1 years)
"*(loud) when I am mad and sometimes when I am playing games... (quiet) when I reading a book, and asleep*" (girl; 8.2 years)
"*Both...I'm loud when I am with my friends and I am not loud when I am on my own, but I am loud with my sister*" (girl; 8.7 years)

Table 5.3: Children's responses relating to academic self-concept

	Total (N=89) % (n)	Nursery (n=22) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
Do you think you are clever?				
Yes	73 (65)	23 (5)	83 (25)	94 (35)
No	4 (4)	-	10 (3)	3 (1)
Don't know	23 (20)	77 (17)	7 (2)	3 (1)

The children were also asked whether they thought they were clever or not, and three quarters of all the children answered positively (see Table 5.3). This may have been due to the nature of the question and the situation of the interview (i.e., it may be hard to admit you are not clever, especially to an adult in a school setting/environment). Of the 65 children that said they were clever they nearly all answered "yes" to the question without giving any justification or explanation for why they thought this was the case (98%). The two children who did give reasons for why they thought they were clever said:

"Yeah coz I am in the top maths group" (girl; 8.2 years)
"(nods) I am top of everything" (girl; 8.5 years)

Three quarters of the youngest children at nursery (3.0-4.5 years) gave no answer to this question (see Table 5.3).

About your house

Table 5.4: Children's responses to questions about their home

	Total (N=89) % (n)	Nursery (n=22) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
What is your house like?				
Big	57 (51)	45 (10)	50 (15)	70 (26)
Middle	9 (8)	5 (1)	13 (4)	8 (3)
Small	27 (24)	45 (10)	27 (8)	16 (6)
Don't know	7 (6)	5 (1)	10 (3)	6 (2)
Do you have garden?				
Yes	89 (79)	73 (16)	90 (27)	97 (36)
No	11 (10)	27 (6)	10 (3)	3 (1)
Do you like your house?				
Yes	44 (39)	5 (1)	43 (13)	68 (25)
Don't know	56 (50)	95 (21)	57 (17)	32 (12)

The children were asked about their home environment. First, they were asked what their house was like, and whether they had a garden in which to play. As shown by Table 5.4, just over half of all the children said that they had a big house. 89% of all the children said that they had a garden. Nearly half of all the children talked positively about their home - saying they liked it (see Table 5.4):

"A big garden with a swing in it" (boy; 3.8 years)
"Yeah I have it (garden) with vegetables growing under the ground... they are not grown up yet" (girl; 4.4 years)

"*Yeah I have a little pond of my own*" (boy; 6.3 years)
"*Quite big (garden), got loads of flowers in it*" (boy; 6.5 years)
"*My house is excellent! I have got a really cool tool set, and I have a real hammer, a real saw*" (boy; 7.9 years)
"*Well, it's kind of nice, it's big, coz it's a four bedroom place*" (girl; 8.7 years)

Table 5.5: Children's responses about sharing a bedroom

	Total (N=89) % (n)	Nursery (n=22) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
Do you have your own room?				
Own room	67 (60)	68 (15)	67 (20)	68 (25)
Share (with siblings)	33 (29)	32 (7)	33 (10)	32 (12)

The children were asked whether they had their own bedroom, or if they had to share a room with brothers or sisters. Two-thirds had their own bedroom (see Table 5.5). One-third of the children had to share (see Table 5.5), and 20 of these said that they wanted their own room. Some of the children's comments included:

"*I share a room... and he (brother) always keeps me awake coz he always, he doesn't go to sleep, that's why he always keeps me awake*" (boy; 6.2 years)
"*I share my room with Denika... yeah... coz she (sister) always wants to have the light on at night and I want it off*" (boy; 6.4 years)
"*Yeah, coz he (brother) always, because he always puts radio on*" (boy; 8.4 years)
"*... would rather have my own room... because my sister always snores and she wakes me up*" (girl; 8.5 years)

Activities or hobbies

Table 5.6: Frequencies of different activities mentioned by children

	Total (N=89) % of total (n)
Activities mentioned:	
Watching television/videos	91 (81)
Playing sports (including football & bike riding)	65 (58)
Chase/tag/hide & seek	45 (40)
Playing pretend/make-believe games	42 (37)
Playing computer games	36 (32)
Playing board games	24 (21)

The children were asked about what sort of games they liked to play at home and with their friends. The children talked about a range of different activities and games. Nearly all the children reported watching television and videos (see Table 5.6):

"*I like watching thomas, thomas the tank engine*" (boy; 3.2 years)

"*I play with my Barbies and I watch TV*" (girl; 6.3 years)
"*I watch telly... when I come home from school*" (girl; 7.8 years)

And for one boy television was very important to him:

"*The telly is my life, I can't stand without my telly...I sit in front of the telly all the time... I can't live without the telly*" (boy; 7.9 years)

Two-thirds of all the children said that they played various sports (see Table 5.6) including football and bike riding:

"*I have two bikes but my old one has got thrown away in the dump*" (girl; 4.1 years)
"*I like to play, I just play on my bike... I just peddle round on my bike*" (girl; 4.6 years)
"*I like football... yeah I have just played it outside for an hour*" (boy; 7.7 years)
"*And we have got a shed with my bike in it that my mum bought me... yeah it was for Christmas, I really wanted it and my old bike got mucked up coz when I was riding it, all the way down to tescos and on the way back, the pedal come off*" (boy; 8.6 years)

Nearly half of all the children talked about playing chase/tag/hide and seek type playground games at school (see Table 5.6):

"*Play kiss chase and power rangers*" (girl; 6.1 years)
"*Duck-duck-goose... You have to say duck, duck, and when you say goose the person chases you... if you get the person, you're the one that says duck-duck-goose, and then goose again*" (boy; 8.4 years)
"*Play chasing the girls at school... yeah they always come up to me and they say you get us now and then we have to go running after them and after a few minutes they come back up and start again*" (boy; 8.6 years)

Nearly half of children also referred to pretend/make-believe games as a past-time (see Table 5.6):

"*Mums and dads, I play*" (girl; 3.2 years)
"*Yeah, I play dressing up as well*" (girl; 3.9 years)
"*Half the games I make up... don't know, I take snowman, my teddy, a toy one, I pretend he's alive and I do his voices*" (boy; 5.9 years)
"*James Bond... like when you have to choose your person, and the people are called James Bond and Odd-Job... yeah you have got to shoot each other even though you are on the same side*" (girl; 7.8 years)

Playing on the computer was mentioned by a third of all the children (see Table 5.6):

"*Yeah and on the computer... it has got games and the internet*" (girl; 5.4 years)
"*I invite my friends round, that just live next door, the second ones, and they come round my house... and we go on the playstation*" (boy; 6.4 years)
"*And I have got some very good computer games...the internet hasn't been installed on my computer yet*" (boy; 7.9 years)
"*Well, I like to play my two player computer games, and my friend has got this game I borrowed and it's up to four players, and it's really good*" (boy; 8.1 years)

The influence of technology was seen in children's lives as 20 of the children who did not have a computer at home to play on still mentioned working on one at school or playing on one of their friend's computers.

One quarter of all the children mentioned playing board games (see Table 5.6):

"My mouse game and with my dad... playing with my dad" (boy; 3.8 years)
"I have got Pokemon monopoly at home, and sometimes my dad plays with me" (girl; 8.2 years)
"Like to board games at home...play cluedo, drafts and snakes and ladders" (girl; 8.4 years)

School life

Table 5.7: Enjoyment of attending school/nursery

	Total (N=89) % (n)	Nursery (n=22) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
Do you like going to school/nursery?				
Yes	70 (63)	55 (12)	73 (22)	78 (29)
No	15 (13)	9 (2)	17 (5)	16 (6)
Don't know	15 (13)	36 (8)	10 (3)	6 (2)
What is your favourite subject/activity?				
Art	48 (43)	45 (10)	50 (15)	49 (18)
Playtime	33 (29)	32 (7)	33 (10)	32 (12)
Other (various)	19 (17)	23 (5)	17 (5)	19 (7)
What is your worst subject/activity?				
Nothing – like everything	46 (41)	36 (8)	50 (15)	49 (18)
Mathematics	23 (20)	0 (0)	27 (8)	32 (12)
Writing	12 (11)	0 (0)	16 (5)	16 (6)
Other (various)	19 (17)	64 (14)	7 (2)	3 (1)

The children were asked about school/nursery life. As shown by Table 5.7, three quarters of all the children said that they liked coming to school or nursery.

The school-age children (5.0-8.5 years) reported a variety of different favourite subjects/times. The most popular subject was art with half of the school-age children mentioning this as their favourite subject (see Table 5.7):

"I like doing my favourite... painting" (girl; 6.5 years)
"Doing art... because you do fun things, and like, and I am good at art and I am not very good at some things" (boy; 7.9 years)

"Well, what I like best about school is art coz it's fun, which we are going to do today" (girl; 8.2 years)

For a third of all the children, playtime was their favourite time of the day (see Table 5.7):

"When it's playtime" (girl; 6.1 years)
"It's having erm it is where you are allowed to what you want, where you are allowed outside at the rest of afternoon til home time" (boy; 8.6 years)

The school-age children mentioned mathematics and writing most as their worst subjects, and below are some examples of why they disliked these subjects:

"Writing... coz I can't do it, it's boring" (boy; 5.9 years)
"Um maths coz it's quite hard and I'm not very good at it" (girl; 6.3 years)
"Writing... coz it's hard" (girl; 6.6 years)
"Hard work... when you have to do really hard writing work like... really hard writing when you have to do hundreds of writing" (girl; 7.8 years)
"I don't much like, um, um, english... coz um I am not very good at writing" (girl; 8.2 years)

However, half of all the children found it harder to talk about what they did not like doing, said that they "liked everything" when asked what they did not like doing (see Table 5.7).

The nursery children gave different answers as to what they liked doing at nursery. The examples of activities the youngest children gave tended to be less detailed:

"I like dressing up" (girl; 4.0 years)
" I do drawing... I like to draw my daddy" (girl; 4.1 years)
"I like playing with... the Barbie telephone, it is new" (girl; 4.4 years)

Although one girl gave a good explanation about not liking 'boys' games:

"I like doing everything, things what's boys games is boring isn't it... er all sorts of games boys do... like star wars" (girl; 4.3 years)

Peer relationships

The children were also asked about their friends. Nearly all the children were able to answer the questions about their friends, however the school-age children gave more detailed comments about their friendships compared to the nursery children.

Table 5.8: Children's responses to questions about their friendships

	Total (N=89) % (n)	Nursery (n=22) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
Mentioned friends during interview	94 (84)	86 (19)	100 (30)	95 (35)
Do you have lots or a few friends?				
Lots	52 (46)	41 (9)	50 (15)	59 (22)
A few	42 (37)	41 (9)	50 (15)	35 (13)
None	1 (1)	5 (1)	-	-
Don't know	5 (5)	13 (3)	-	6 (2)
Do you have a best friend?				
Yes	58 (52)	32 (7)	73 (22)	62 (23)
No	5 (4)	4 (1)	-	8 (3)
Don't know	37 (33)	64 (14)	27 (8)	30 (11)

As shown by Table 5.8, nearly all of the children mentioned their friends in the interviews. When asked whether they had a lot or just a few friends 52% of all the children said they had a lot of friends, and 42% said they had just a few friends (see Table 5.8). It was evident from the children's answers that friendships were an important aspect of their lives:

"*Yeah I am quite good at making friends, I say, oh what's your name and they say whatever their name is and just that we are friends*" (boy; 7.9 years)
"*Well I always play with my friend Helen and she's always saying can we play it*" (girl; 8.3 years)

However nearly two thirds of all the children reported that they had one or two 'best' friends (see Table 5.8). Ten of the oldest children (7.0-8.5 years) talked about why they had some 'best' friends giving some detailed reasons, for example:

"*I have two sticking out friends, it is usually I usually always play with them two... probably the strongest so far would be James*" (boy; 7.9 years)
"*My best friend's Helen, coz she's always playful, and Emma's always funny coz she always makes me laugh coz she runs so slow, but she can run faster, she just likes being on*" (girl; 8.3 years)
"*I got Anna and Amelie, she is the one, she helps me a lot, she keeps lots of secrets that I don't want her to tell*" (boy; 8.6 years)

Sixteen of the children also talked about differences between different types of friends (see Table 5.8), and these were all the oldest children (7.0-8.5 years):

"*It depends which type of friends... well, I have got friends who I like and they come round and help me do cheats ad I give them cheats and they borrow games... then*

there's those who I don't like to come round... they come because my mum likes their mum... not very often but sometimes" (girl; 7.8 years)
"*Yeah I have lots, well, I like my friends, Emma and Tim, Tilly and Lianne, I like my friends because, well, Lianne makes me laugh a lot, and Tilly helps me do my work when erm someone's away by my desk, and Emma, well, she helps me to um find another person if someone's away and she plays with me when I haven't got a game to play*" (girl; 8.2 years)

Eight of the school-age children (5.0-8.5 years) also discussed in detail about not having many friends, or being bullied by older children, and how they felt about this:

"*Coz like you haven't got no-body to play with... no, sometimes no-one*" (boy; 6.5 years)
"*Nuffing, no-one don't like me... yeah I just walk around and that*" (girl; 7.7 years)
"*I told Carl 'go away' and that's the first time anybody in my class or my year has stood up to him... well, but there's one problem with bullies... when somebody stands up to someone, that they go on to a different person and that's what I hate about bullies, they can never get the bullies out of them... Carl is the biggest coward of all, coz he doesn't care about anybody that's older than him, he just leave them alone, he goes to year three's and four's...I sometimes walk away, but sometimes he is beating up someone I know and like... then I tell him to go away*" (boy; 7.9 years)

Table 5.9: Children's answers on how they get on with their friends

	School-age total (n=67) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
Do you get on well with your friends?			
Yes	64 (43)	63 (19)	65 (24)
No	0 (0)	0 (0)	0 (0)
Don't know	36 (24)	37 (11)	35 (13)
Do you argue with your friends?			
Yes	54 (36)	37 (11)	68 (25)
No	18 (12)	33 (10)	5 (2)
Don't know	28 (19)	30 (9)	27 (10)

As shown by Table 5.9, just over two thirds of the school-age children reported that they got on well with their friends. Just over half of the children also talked about when they argued with their friends (see Table 5.9), and some gave reasons why they thought this happened:

"*When we try and put as game but Helen says 'oh you be the mum' and I say 'I'm gonna be the mum' but then Helen says 'I'm gonna be the mum' and then I say 'I'm gonna be the baby' and then she says 'I'm gonna be the mum', then we just argue*" (girl; 6.3 years)
"*A couple of times... like what we are going to play and that, we squabble*" (boy; 7.9 years)
"*Sometimes... um coz if Carl draws some paint over my work like he did today, he slaps me and then I argue with him*" (boy; 8.1 years)

"Sometimes... like when they say you're it and I wasn't, like ellie wasn't and they start arguing about it and then they say I should have been on and then they saying oh you shouldn't be on" (girl; 8.3 years)

Fifteen children also went to elaborate that even though they argue they usually make up again quite quickly showing that arguments are seen as part of friendship:

"Yeah sometimes we do have little arguments...we always get back together, it is mostly about like, it is not much really" (girl; 8.2 years)
"No I get on well with my friends, but today we had a little argument... coz James started hitting Ben... yeah, we made up in the end" (boy; 8.5 years)
"Sometimes, sometimes we break up but then we make up again" (girl. 8.5 years)

Table 5.10: Children's answers on seeing friends

	School-age total (n=67) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
Do your friends come over to your house?			
Yes	79 (53)	77 (23)	82 (30)
No	16 (11)	20 (6)	13 (5)
Don't know	5 (3)	3 (1)	5 (2)
Do you go over to theirs?			
Yes	76 (51)	67 (20)	84 (31)
No	4 (3)	6 (2)	3 (1)
Don't know	20 (13)	27 (8)	13 (5)

As shown by Table 5.10, three quarters of the school-age children talked about their friends' coming over to play at their houses. Three quarters of the older children talked about going round to their friends' houses to play as well (see Table 5.10). Thirty of these children gave actual examples of when and where they saw their friends out of school time:

"They knock on my door and I go out and play with them, and I play with my scooter" (boy; 5.9 years)
"I am going round Carl's house for his party" (boy; 5.9 years)
"Yeah I have been round to Gavin's house and I have been round to Ryan's house and Ricky and Gus' house" (boy; 6.6 years)
"Yeah loads of them... I have already had Denny and Karl to sleep" (boy; 7.7 years)
"Yeah, well I have Emma round my house, coz she's the nearest, so it's easier, coz she's round the block" (girl; 8.2 years)

Of the eleven school-age children that reported that their friends did not come round to play (see Table 5.10) five children gave reasons why they were not able to come over, for example:

"I would like him to come round to my house... well, I don't get much time, he don't get much time too, my mum and his mum don't get much time to chatter coz his mum's quite quick at talking" (boy; 6.3 years)
"Well I don't really play with my friends in my house, they are not really allowed over coz my mum and dad's always busy" (girl; 8.7 years)

The nursery children were less able to communicate about their friends and what they thought about them but two fifths made understandable comments about their friends including comments like:

"Um, they can't play with my new toys coz if they play with my new toys, they get broken" (boy; 3.8 years)
"Erm, only Joanna did, at my house, but I haven't been to anyone's house, except my nanny and granddad's and nanny Rebecca's" (girl; 4.0 years)
"Yeah, I got Cara coming round my house, but she's not coming round yet, we got the ring her mum up" (girl; 4.1 years)

Family relationships

The children were also asked about their family lives and relationships. Both parents appeared to play an important part in the children's lives. When talking about their home lives the children mentioned both parents almost equally (see Table 5.11).

Table 5.11: Percentages of children mentioning parents during interview

	Total (N=89) % (n)	Nursery (n=22) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
Mentioned mum				
Yes	97 (86)	91 (20)	97 (29)	100 (37)
No	3 (3)	9 (2)	3 (1)	0 (0)
Mentioned dad				
Yes	96 (85)	91 (20)	97 (29)	97 (36)
No	4 (4)	9 (2)	3 (1)	3 (1)

As shown by Table 5.11, nearly all the children mentioned their mothers in the interviews:

"Um, she (mum) gives me cuddles and I give her cuddles" (girl; 4.0 years)
"... and I like my mummy giving me a cuddle" (girl; 4.2 years)
"...and make things with my mum... cooking things... yeah sometimes we go to the shopping place and the beach" (boy; 6.2 years)
"My mum mostly buys me things... she buys me lots of toys and things for my bedroom" (girl; 6.3 years)
"Um, my mum, I like playing with my mum... cooking, I like baking cakes" (girl; 8.4 years)

"The bestest thing I like to do is drawing... my mum teaches me... she is kind of because she, when she went to school, everyone kept asking her to draw pictures" (boy; 8.6 years)

Nearly all the children mentioned their fathers (see Table 5.11):

"My daddy takes me to the shops" (boy; 3.2 years)
"My daddy put me on his head and plays rock-a-bye baby with me, and he rocks me in his arms, and then he's ready to throw me on the sofa, he throws me on the sofa" (girl; 4.3 years)
"No he just goes to work... but sometimes he plays, when my mum says 'get daddy, get daddy' and we get daddy down on the floor, and we try to push him, but he gets us in the air so we can't get down" (girl; 6.6 years)
"Going to the park, only my dad does that" (girl; 6.6 years)
"With my dad I like doing sports with him... badminton, football, rugby" (boy; 7.8 years)
"He works, but he's not at home sometimes, but I help him do the cleaning" (girl; 8.4 years)

Nine children reported that they did not have both parents living at home with them, and all nine mentioned the other non-resident parent in some way, for example:

"I got two dad's" (boy; 4.4 years)
"I don't have a dad, I only have a mum... but I used to have a dad when I was a little baby, but my mummy and my daddy, well my dad was called Andrew but now I haven't got any dad's" (girl; 6.3 years)
"I have got two dads... my fake dad is Andy and my real dad is Alan" (girl; 7.7 years)
"My dad doesn't live with us, but my mummy does...see my dad every two weeks... he takes us out sometimes" (girl; 8.3 years)

Table 5.12: Percentages of children playing/arguing with their parents

	Total (N=89) % (n)	Nursery (n=22) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
Do you parents' play with you?				
Yes	91 (81)	95 (21)	97 (29)	84 (31)
No	9 (8)	5 (1)	3 (1)	16 (6)
Do you argue with your parents?				
Yes	54 (48)	23 (5)	57 (17)	70 (26)
No	22 (20)	27 (6)	27 (8)	16 (6)
No answer	24 (21)	50 (11)	16 (5)	14 (5)

Nearly all the children mentioned that their parents actively played with them at home (see Table 5.12). Of the nine children who didn't mention that their parents played with them, all mentioned that they would like to play with their parents more:

"... they don't always play games coz they are always busy tidying up" (girl; 6.2 years)

"I don't seem to play with them, no because my mum, my dad usually in London... they are always busy, my dad is usually at work, and my mum is usually working in the house... so I usually go and play in my own to do something" (boy; 6.3 years)
"She don't do nothing with me! She just cooks" (boy; 8.4 years)
"My mum isn't into games... yeah I tried to get her into doing it this morning, but... she just went out to do the washing up" (girl; 8.7 years)

Despite parents being an important aspect of their lives, half the children reported that they argued quite a lot with their parents (see Table 5.12). However this seemed a common part of family interactions, as thirty of those children who said they argued with their parents could offer good explanations as to why this happened including:

"Sometimes she argues with me...coz sometimes she come up to my room and she says 'your dinner's ready' and I say 'mummy I'll come down in a minute' and she says 'no now'... I am trying to tidy my toys up" (girl; 6.3 years)
"Um sometimes... um that I don't want to go to bed and I want to watch tv, and she says no I have to tidy up now, and all that" (girl; 7.8 years)
"Sometimes I argue... coz my dad's always annoying me... coz always when my mum says 'Freya, can you come here', he asks where I am going and then um I tell him and then he starts shouting... then I start shouting back" (girl; 8.3 years)
"Well, I don't get on with my dad sometimes... coz sometimes he makes me get cross, after he has got cross with me and I get cross but I have to go to my bedroom" (boy; 8.6 years)

Half of the younger (nursery, 3.0-4.5 years) children did not provide any answer the question (see Table 5.12).

Table 5.13: Children's comments about their siblings

	Total (N=89) % (n)	Nursery (n=22) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
Do you have siblings?				
Yes	78 (69)	54 (12)	80 (24)	89 (33)
No	22 (20)	46 (10)	20 (6)	11 (4)
Do you play with your siblings?				
Yes	60 (53)	36 (8)	60 (18)	73 (27)
No	18 (16)	18 (4)	20 (6)	16 (6)
No siblings	22 (20)	46 (10)	20 (6)	11 (4)
Do you argue with your siblings?				
Yes	66 (59)	36 (8)	67 (20)	83 (31)
No	7 (6)	9 (2)	10 (3)	3 (1)
No siblings	22 (20)	46 (10)	20 (6)	11 (4)
No answer	5 (4)	9 (2)	3 (1)	3 (1)

The children were asked whether they had brothers or sisters, and if so how well they got on with them. Three quarters of all the children did have siblings and mentioned them during the interview (see Table 5.13). Siblings did appear to be an important part of the children's lives. Two thirds of all the children talked about playing with their siblings at home (see Table 5.13):

"Then he (brother) tells me jokes, lots of jokes, and he makes me laughing and that" (boy; 4.6 years)
"Football with my sister... and play in my room with my toys" (girl; 6.5 years)
"Um we, with my brother, we play snakes and ladders... and sometimes football and climb trees sometimes" (boy; 6.6 years)
"I play football with my brother" (boy; 8.4 years)

One fifth of all the children who reported not playing with their siblings (see Table 5.13), and gave various reasons why they could not play with them:

"My sister, Alisha, always doesn't want to play, and Denika always has to do her homework, so I play on my own" (boy; 6.3 years)
"Harry's about ten and Shaun is about nineteen... nah, Shaun doesn't like playing with me and Harry doesn't like playing with me" (boy; 6.4 years)

One fifth of all the children did not have siblings at home (see Table 5.13). Further illustration of the importance of siblings was that ten of the children who did not have any brothers or sisters talked about having pretend ones for example:

"Yeah, I have got some pretend ones... other ones but they are not proper ones... they are pretend brothers and sisters" (girl; 8.3 years)

Two thirds of all the children said that they argued with their siblings (see Table 5.13). Their answers revealed why the children argue with their siblings:

"But when she (sister) smacks me I cry, then she says shut up and I don't like that and mum says stop it" (boy; 3.8 years)
"But I argue with Timmy... coz he gets things first and I want it, but I get things first and then he wants it" (girl; 4.2 years)
"She (sister) beats me up... she punches me and kicks me... I do it back" (girl; 5.8 years)
"Sometimes he (brother) screams at me and I can't play with him" (boy; 6.2 years)
"I do argue with Frances quite a lot... but with Julie, my youngest sister... she is still younger-er than me, um I some, I get to play with her, but sometimes I do have arguments with her" (boy; 7.9 years)
"Erm, when he (brother) doesn't let me come in his room, and he does come in my room sometimes" (boy; 8.4 years)
"She (sister) keeps taking my Barbie doll away... yeah, and my bigger sister, she comes, she just comes in my room, I am not allowed to go in her room" (girl; 8.7 years)

Table 5.14: Children's comments about other family members, pets, and trips

	Total (N=89) % (n)	Nursery (n=22) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
Mentioned grandparents				
Yes	79 (70)	64 (14)	80 (24)	86 (32)
No	21 (19)	36 (8)	20 (6)	14 (5)
Mentioned cousins				
Yes	8 (7)	0 (0)	7 (2)	14 (5)
No	92 (82)	100 (22)	93 (28)	86 (32)
Mentioned pets				
Yes	39 (35)	32 (7)	40 (12)	43 (16)
No	61 (54)	68 (15)	60 (18)	57 (21)
Mentioned going to trips				
Yes	72 (64)	64 (14)	73 (22)	76 (28)
No	28 (25)	36 (8)	27 (8)	24 (9)

Three quarters of all the children mentioned their grandparents when talking about their families (see Table 5.14), either in relation to going to see their grandparents, or about their grandparents taking them out on trips. Their answers revealed that their extended family was important to them:

"*Yeah when we go to our Nanny's, nanny is coming for tea tonight*" (boy; 3.8 years)
"*Yeah I like going to my Nan's*" (girl; 4.2 years)
"*Um I like Nanny and granddad babysitting coz they bring me, they bring me, Molly and Alice a picnic with them, and one time when they brang a picnic, they brang jammie dodgers and um lollies and um crisps and a picnic*" (girl; 6.2 years)
"*I have got two... Nanny Smith and Nanny Alexander... I see my Nan's a lot... coz on Wednesdays and Saturdays I go round my Nanny Smith's*" (boy; 6.6 years)
"*I see my Nan every weekend... she lives in Canterbury... um we have a Macdonalds and then we go, to my Nan's and see my granddad*" (girl; 7.9 years)
"*I go and see my granddad sometimes, he likes to sleep and eat a lot of sweets*" (boy; 7.9 years)
"*Yeah coz my granddad and my granny, they come round every Tuesday... well if I see them, granddad normally wants a cuddle and I just watch nanny talk to mummy and daddy sometimes*" (girl; 8.2 years)
"*Yeah I got my Nan and granddad and they have a little dog and they lives round the corner from me*" (boy; 8.6 years)
"*... and my granddad Geoff, he always works and he goes to work Monday and he sits in his lorry til Friday, and then he comes to see us every Saturday*" (girl; 8.5 years)

Other extended family members (e.g., cousins) were mentioned by seven children (see Table 5.14):

"*... and my auntie she gives me five pounds in pocket money*" (girl; 7.7 years)

"I see my auntie and uncle and my cousins Dean, coz they always come round every Friday and every Saturday" (girl; 8.5 years)

Pets were mentioned by two fifths of all the children (see Table 5.14):

"I have got some rabbits... and my rabbits is called Flopsy, and Thomas' guinea pig is called Phillip... it goes 'Eeeek, eeeek'" (girl; 4.2 years)
"... and I have a dog, and her name is called Iden, I dunno how old she is... but she's very friendly" (girl; 5.8 years)
"They are our dogs, we've got three... some guinea pigs and hamsters" (girl; 6.4 years)
"Yeah we have got guinea pigs... they are just eating the grass all the way down like a lawn mower" (boy; 7.9 years)
"I have got quite a lot... they all have been named... I have got three dogs, and I have got three hamsters, two guinea pigs and er one rabbit... I know all their names, the dog's names are Roxy, Tilly and Tessa, and the rabbits names Alice, and the hamsters names Nutter, as he's a bit nutty, and then there's Freda and she bites the bars, and we have got Coco and Millie and Pee-Wee" (boy; 8.6 years)

Three quarters of all the children mentioned going on trips and outings (see Table 5.14):

"My daddy takes me to the swings" (girl; 4.1 years)
"Yeah, my mummy is going to take me to the park with my daddy" (boy; 4.6 years)
"Yeah my dad is England and my mum is german, and we are going to Germany in holidays again and the holiday we are going to go in Spain... with my cousins and all my family" (girl; 7.8 years)
"Yeah they sometimes take me to, they took me to London and they are taking me on holiday soon...Spain" (girl; 8.3 years)

Feelings or thoughts

The children were also asked about their feelings and emotions that they feel in their everyday lives. The children were asked about four emotions (i.e., scared; cross or angry; sad; and happy) and asked to give examples of things, situations, or people that make them feel like this.

Table 5.15: Children's comments about emotions

	Total (N=89) % (n)	Nursery (n=22) % (n)	Year 1 (n=30) % (n)	Year 3 (n=37) % (n)
Do you feel happy?				
Yes	94 (84)	86 (19)	97 (29)	98 (36)
No	0 (0)	0 (0)	0 (0)	0 (0)
No answer	6 (5)	14 (3)	3 (1)	2 (1)

Do you get cross?				
Yes	74 (66)	55 (12)	80 (24)	81 (30)
No	9 (8)	9 (2)	10 (3)	8 (3)
No answer	17 (15)	36 (8)	10 (3)	11 (4)
Do you get sad?				
Yes	61 (54)	50 (11)	53 (16)	73 (27)
No	11 (10)	5 (1)	20 (6)	8 (3)
No answer	28 (25)	45 (10)	27 (8)	19 (7)
Do you get scared?				
Yes	62 (55)	59 (13)	57 (17)	68 (25)
No	22 (20)	0 (0)	33 (10)	27 (10)
No answer	16 (14)	41 (9)	10 (3)	5 (2)

Nearly all the children admitted that they had felt happy at some point (see Table 5.15), and could give examples of what things/situations/people had made them feel like this. There were patterns in the examples that children gave for what made them feel happy. Nearly half of these children's examples were related to people (46%, n=39) - of these 15 children gave family members as the reason, and 24 children mentioned their friends. Two-fifths of these children talked about doing various games or activities that they enjoyed (39%, n=33). Some examples of their comments are given below.

"*Playing with my friends*" (boy; 3.8 years)
"*When go to bed, my mummy sings a song... when you happy and you know it*" (girl; 4.2 years)
"*That I love my mummy and daddy, and I cuddle them*" (girl; 4.3 years)
"*When I cuddle mummy that makes me really happy*" (girl; 4.3 years)
"*When my mummy buys me bubble gum and chocolate and sweets and all that*" (girl; 5.8 years)
"*Playing with my brother, and playing with my mum and dad*" (girl; 6.6 years)
"*Going to play in the park with my dad*" (girl; 6.6 years)
"*When my mummy gives me a cuddle*" (girl; 7.8 years)
"*When my friends make you laugh*" (girl; 7.8 years)
"*I like it when my brother likes to play with me a lot*" (girl; 8.2 years)
"*When I can play with my friends nicely and my brothers and that lot*" (boy; 8.5 years)
"*Well, the things I like doing is playing my computer games... and playing with my friends*" (boy; 8.6 years)
"*I really really like playing with my friends and singing*" (girl; 8.6 years)

Three quarters of the children admitted that they had felt cross or angry at some point (see Table 5.15). Three quarters of children's examples for what made them cross were related to people (71%, n=47). Forty-two children gave family members as the reason (often siblings), and five children mentioned their friends. Some of the children's comments are given below.

"*When I keep fighting my sister coz she gets cross with me, I don't get cross with her but she hits me*" (boy; 3.8 years)
"*My mummy tells me off, I get cross*" (girl; 4.1 years)
"*Cornea (sister) makes me cross and sometimes she bees horrible to me*" (girl; 4.6 years)
"*Angry with my mum shouts at me and I shout to her and my dad*" (girl; 6.2 years)
"*When daddy tells me off*" (boy; 6.3 years)
"*When my sisters want to play with my toys… I don't want her to*" (girl; 6.4 years)
"*I get cross with my brother… coz he doesn't let me in his room and I do let him in my room*" (girl; 6.5 years)
"*When people don't let me play, like sometimes Rebecca plays with Jack and they don't let me play*" (girl; 7.8 years)
"*When my sisters just go into my room without my permission*" (boy; 7.8 years)
"*When my brother's take my stuff and they don't give it back, I fight for it*" (girl; 8.2 years)
"*I know what makes me cross, my sister, she annoys me, I get really angry with her… when I am in bed, she comes in opens the door and says let me in your room, and I said no, and she just comes in*" (boy; 8.6 years)

Two-thirds of the children admitted that they got sad (see Table 5.15). Three quarters of these children's answers about what made them feel sad were related to people (77%, n=42). Twenty-five children gave their family members as the reason, and 17 children mentioned their friends (relating to concerns about being lonely and having no-one to play with). Some of their answers are given below.

"*My mummy coz she tells me off everyday*" (girl; 3.3 years)
"*When you hurt myself, it does*" (girl; 4.0 years)
"*When mummy tells me off*" (girl; 4.4 years)
"*When somebody bees nasty*" (girl; 6.3 years)
"*When mum shouts at me*" (boy; 6.3 years)
"*When I am lonely*" (boy; 6.4 years)
"*When I have got no friends to play with*" (girl; 6.5 years)
"*When my friend comes round and I don't want her to go and she doesn't want to go…*" (boy; 7.7 years)
"*When my great Nan died*" (girl, 7.8 years)
"*When people shout at you and say names*" (girl; 8.2 years)
"*Breaking up with my friends… when people bully me… older kids… they always hit me*" (boy; 8.4 years)
"*When I ain't got no-one to play with*" (boy; 8.4 years)

Two-thirds of the all the children admitted that they had got scared (see Table 5.15):

"*When the lights aren't turned on… yeah, and I can't see where I am going*" (girl; 4.0 years)
"*Bad dreams do… yeah I do have bad dreams sometimes*" (girl; 4.3 years)

"Well sometimes I am scared of the dark, and I'm scared about scary monsters and scary dreams" (girl; 4.3 years)
"Being without my mum... she said she was going into Iceland and she went into the pound shop and I didn't know where she was" (boy; 6.3 years)
"When I left my mum and dad... I lost them... when we in supermarket place, and I was standing, and I thought they were over there, and they wasn't, and I lost them" (boy; 6.3 years)
"I was scared in a aeroplane coz I didn't know where it was gonna land" (boy; 6.6 years)
"When it's dark and like my door is creaking and stuff" (boy; 7.8 years)
"Um I have, but it was when I was little at Christmas time, I thought there was a ghost in the cupboard as they was this little green line, I opened it the next day but it wasn't there" (girl; 7.9 years)
"Yeah we are going on holiday this year on the plane but I am a bit scared... coz I think the plane might crash" (girl; 8.2 years)

The interviews revealed that even the youngest children (3.0-4.5 years) could give examples of things that made them feel various emotions. Their answers showed that young children are able to talk about when they have felt emotions, distinguish between different emotions, and identify what things/situations/people may have made them feel a particular emotion.

Earliest memories

The children were asked about what their earliest memories were. Three quarters (72%, n=48 out of 67) of the school-age children (5.0-8.5 years) were able to report an early memory in an understandable way:

"When I was a baby, I keep scratching myself, and I started to have excema like I have it now, so I had to have puffers" (boy; 5.9 years)
"I do remember when I was four, I want on a train to see Father Christmas" (boy; 6.2 years)
"Um one thing... when I was about three, I fell on the door and hurt my knee and cut it... was bleeding a lot" (girl; 7.7 years)
"Um I went to Majorca with my Nan and granddad, my mum and my dad, my mum was pregnant and I feel into the sea... I just fell in and I couldn't swim and my cousins got her little sea boat and got me in it, coz I couldn't swim" (girl; 8.3 years)

These memories were related to various situations, 42% (n=20 out of 48) were related to illness, 33% (n=16) to their family, and 25% (n=12) to specific activities (e.g., learning to crawl).

36% (n=8 out of 22) of the youngest children (3.0-4.5 years) were able to talk about what they remembered about being little:

"I had a dummy and in a cot" (boy; 3.2 years)

"Er, I sleeped in a nice carrycot, stringy" (girl; 4.3 years)
"I was very teeny, when I was born... I drink milk... but now I drink fizzy and orange juice... and a cup of tea!" (girl; 4.0 years)
"I was a baby, I went in a cot and had baby clothes with teddy bears on" (boy; 4.6 years)

The youngest children's answers about memories were all related to specific events (e.g., drinking, sleeping) and they gave no further explanation of these memories in the interview.

New things learning to do

The children were also asked about what new things they were learning to do. 85% (n=76) of all the children said that they were learning new things, and they were able to give examples of things they were learning. Of these children, two-thirds (66%, n=50) of the children reported that they were being taught new 'skills' at school (e.g., reading or sums). One third (34%, n=26) of the children mentioned that they were learning how to do new activities or sports (e.g., netball or football). Some examples of their answers are given below.

"I am learning to do cartwheels... and I am learning to, I can do handstands" (boy; 3.2 years)
"Um I am learning sums, and I am learning drawing, and maths" (girl; 3.9 years)
"I am learning how to play with my new toys, and to play football" (boy; 3.8 years)
"Well, a little bit forwards and then a little bit backwards the pedals go... I am learning" (girl; 4.0 years)
"I have learnt how to write numbers up to a hundred and there are some hard ones" (girl; 5.8 years)
"I am learning to eat all my dinner on Sundays... yeah coz I waste things what I don't like" (boy; 6.3 years)
"Reading... yeah, my mum said I'm getting there, but I need to practise... coz we take reading books home and when we have finished one, we get another one" (boy; 6.4 years)
"I am learning to swim... my mum is helping, normally I go like this and I nearly swim, and then but I can't stay on the water so I swim under water, I try to stay on the top, but I just sink, but I can swim under water" (boy; 7.7 years)
"Well, playing football... I learning at home, how to kick the ball on my foot and bounce it loads of times... dad is teaching me football" (boy; 7.8 years)
"I am learning how to do maths quite a lot and multiplication, coz I get troubled with that sometimes" (boy; 8.5 years)

Things want to have or want to be able to do

The children were asked about whether there were things that they would like to have that they did not have, or things that they would like to be able to do that they were not able to do. Nearly all (90%, n=60) of the school-age children (5.0-8.5 years) were able to answer these questions, and give examples of what they would like to have or like to be able to do:

"*I wish I could ride my big bike without stabilisers on*" (boy; 5.6 years)
"*Some toys what we haven't got enough money for*" (boy; 5.9 years)
"*I wish I could colour in really neatly*" (girl; 6.5 years)
"*I wanted to learn how to play ball, but I don't know how to do that, I don't know how to do Frisbee, it is really hard*" (girl; 6.5 years)
"*I play on my brother's computer, coz I haven't got one...I would like to have a computer*" (boy; 7.8 years)
"*I would like to be able to swim, I cam swim a little bit, but not along*" (girl; 8.6 years)
"*Well my mum and dad can have as much chocolate as they want and I can't, I only get half a bit a day, I want a whole one.*" (girl; 8.7 years)

Only one third (36%, n=8) of the younger children (3.0-4.5 years) answered this question giving answers like:

"*... I like animals... only my mum lets me have pretend... I want a little mouse!*" (girl; 4.0 years)
"*I want to go to big school*" (girl; 4.2 years)

5.3.2 Using the interview data used to directly inform the content of items in the TedQL measure

The content of the TedQL.1 and TedQL.2 items were originally developed from a review of the literature (see Chapter 2, p. 31-3), and previous experience with children (based on comments children had spontaneously volunteered about what they liked to do, who they played with, and who was important to them, for details of this early development see Chapter 3, p. 71-2). The data from Study 3 built on the previous studies by justifying and providing evidence for the:

i. continued inclusion of any previous items within the TedQL.3
ii. removal of items that were not relevant to young children
iii. addition of other items that were important in young children's lives.

The TedQL.2 measure contained 23 items divided into five areas. Based on the results of Study 3, six items were deleted ("can tie shoes", "good at hopping", "likes to play with friends", "mum talks to them", "is happy/sad", and "gets upset if can't do their work"). Table 5.16 shows how the TedQL.2 measure was expanded to include more items, and some items were removed or altered. Seventeen of the original items were retained - with 13 of these being altered slightly as a result of the interview data to make the wording more appropriate or the items more understandable to young children (see Table 5.16). Thirteen new items were added to the measure, and these items were developed directly from the information provided by children on what was important in their lives (see Table 5.16). Based on the results of Study 3, a new version of the measure was developed containing 30 items (the TedQL.3).

Table 5.16: Domains and items altered due to Bruce's story interview data

Domain	Study 2: Items in existing version (TedQL.2)	Study 3: Items in new version, as altered by interview data (TedQL.3)
Physical Compet-ence	Good at swinging Good at running Good at playing with balls Can climb high things Can tie shoes Good at hopping	Good at swinging *Good at running* *Good at playing ball games* *Good at climbing high things* **Good at bike riding (or scooter/other)** **Good at computer games/video games**
Peer Accept-ance	Has lots of friends Likes to play with friends Likes to tell friends what to do (N) Has friends to play with (others like playing with them)	*Having lots of friends at school* Bossing friends (N) Having lots of friends to play with (others like playing with them) **Friends bossing them(N)** **Friends coming over to their house**
Maternal Accept-ance	Likes to play with their mum Mums talks to them Likes to tell mum what been doing Mum tells them off a lot (N)	*Mum/dad playing with them at home* *Telling mum/dad what been doing at school* *Mum/dad telling them off at home (N)* **Playing with siblings** **Siblings fighting/bossing them around (N)** **Seeing grandparents** **Going on trips with mum/dad/other**
Psycholo-gical Funct-ioning	Is happy/sad Has bad dreams at night (N) Worries about losing their things (N) Some things make them really cross	*Having bad dreams at night (N)* *Worrying about losing things (N)* *Getting cross/angry (N)* **Type of person they are** **Getting scared** **Playing pretend games/dressing up**
Cognitive Funct-ioning	Gets upset if can't do their work Remembers what people tell them to do Good at reading Good at writing Getting better at drawing	Remembering what people tell them to do *Good at reading – read words/look at pictures* *Good at writing – spell name/other words* *Good at drawing – what can draw* **Good at mathematics/numbers** **Good at playing board games**

Note. Key to alterations to items to TedQL measure:
Underlined = item deleted from measure as result of Study 3
Italics = item retained but slightly altered as a result of Study 3
Bold = new item added to measure as a result of Study 3.

5.4 Discussion

This chapter presented the results of content analysis of 89 child interviews with healthy children aged 3-8 years. Key areas of the children's lives were discussed as guided by ten pre-determined themes (i.e., personality; home; activities; school life; peer relationships; family relationships; feelings and thoughts; earliest memories; new things learning to do; and things want to have or be able to do).

The interview data provided information about what was important in young children's everyday lives. The analysis revealed that the school-age children (5.0-8.5 years) were more confident and competent than the nursery age children (3.0-4.5 years) at discussing their lives, feelings, friends, families, and school lives. The frequency data showed that the older children (5.0-8.5 years) provided the more detailed and coherent answers to questions, and gave good examples to support their answers. It was clear that the older children were able to understand what was required of them during the interviews. They all understood the statements read to them on each page of the storybook, and responded with appropriate and detailed answers. However the youngest children (3.0-4.5 years) had more difficulty answering some of the questions. For example, they were less able to talk about what sort of person they were (i.e., quiet or loud; clever or not, see p.142-3) and found it difficult to report memories or early experiences (see p. 159-60), or what they wanted to have or be able to do (see p. 160-1).

The majority of the children across all the ages were positive about their school and home lives. For the most part they were able to talk about what they liked doing at school, and were able to give examples to support their answers (see p. 146-7). However all the children found it harder to talk about what they did not like doing and often simply said they "liked everything" (see p. 146-7). The majority of the children mentioned their friends and families a lot, and it was evident from the interviews that their friends and families were important aspects of their lives (see p. 147-8 & p. 151-2). Arguing with friends, parents, and siblings seemed a typical part of their lives, and many of the children were quick to qualify that they made up with them after arguments (see p. 149). The school-age children (5.0-8.5 years) were able to, and did, compare themselves to others (as illustrated by the constant references to friends, siblings, and parents within their interview answers, see p. 147-154).

Children were presented with four types of emotions (scared; cross or angry; happy; sad), and asked to give examples of things/situations/people that have made them feel this way in the past. Nearly all the children's answers across all ages related to people; with family members being more frequently mentioned than friends (see p. 156-9). The school-age children were more competent than the nursery age children at giving appropriate examples of when they had felt scared, cross, sad or happy. However the majority of the younger children (3.0-4.5 years) were able to provide examples of when they had felt these different emotions. This result showed that even at this young age children are able to talk about when they have felt emotions, distinguish between emotions, and identify things/situations/people that have made them feel a different emotion (see p. 156-9).

A major consideration for the development of our self-report measure for children below eight years (the TedQL) is that the specific content of the items within our measure should be based on information obtained directly from children themselves, rather than assumptions by the researcher as to what is important in children's lives. This study aimed to use information from children themselves to inform the content development of our TedQL measure. Study 3 built on the previous studies by justifying and providing evidence for the removal and/or alteration of existing items in the measure, and the addition of other items. As a result of Study 3 a new version of our measure was developed (the TedQL.3).

The younger children (3.0-4.5 years) found it hard to understand and answer appropriately on some of the key areas of the storybook, and it was difficult to get these children to concentrate on the task. The school-age children (5.0-8.5 years) who were used to the routines and demands of school and classroom life were better able to concentrate. Some of these younger children did not understand some of the questions they were asked, and therefore we felt that the storybook task may have been slightly beyond the youngest children's capabilities. It would have been useful to interview children who have just started school (Year 1, 4.5-5.5 years) to see if they could have answered the interview questions without much difficulty.

Study 3 developed the content for the TedQL measure, but there was also a need to establish what the most appropriate format for this measure should be, as discussed in

Chapter 3 (see p. 65). The reliability and validity of a measure is reliant upon the respondent's ability to understand and manipulate the response scale provided to answer items (Schwartz & Sprangers, 1999). The use of scales as response options assumes the existence of several abilities. As discussed in Chapter 7 (see p. 191-2), Taplin, Goodenough, Webb, and Vogl (1999) argued that researchers cannot assume that children have the cognitive skills involved in rating and seriation tasks. There is little consensus about the most appropriate response format for child self-report measures should be (Rebok et al., 2001). Therefore the development of any instrument must involve testing the appropriateness of a chosen response format, and we will address this issue for our TedQL measure in Study 4 (see Chapters 8 & 9).

The following chapters (Section 3, Chapter 6 & 7) review the literature on children's ability to respond to the content and presentation style of self-report measures (Chapter 6), and whether children can understand the response scales used to answer self-report items (Chapter 7).

Section 3: Further development of the measure

Chapter 6: Young children's ability to respond to items in self-report measures.

Summary

The communication and language barriers that can influence what children report on self-report items and how they behave in interview situations were discussed. There is an imbalance of power implicit in any interactions between adults and children that can impact on children's responses. Due to young children's lack of experience in conversations, children may get confused when researchers ask them questions where the answers seem silly or too obvious, and they may not react well to repeated questioning or probing. Young children may be less able to appreciate the task demands and match their behaviour to meet these demands than older children. They may also get distracted more easily than older children, and be influenced by immediate circumstances or feelings. Children may not understand the meaning of specific vocabulary, and their responses can be influenced by the way items are phrased. Young children are able to recall and sequence past events over a short time period before they are able to do so for future events.

We also discussed the ways to overcome these problems when designing measures. Researchers need to establish rapport with children - to make them feel relaxed and comfortable with the interaction situation. Readability formulas can be useful to help establish the skills necessary to answer items. Pictures or props can be used to present items to children to reduce the task demands on young children. Using age-appropriate language and phrasing can help avoid language problems. The use of specific time recall periods for self-report measures may also help ease the recall task for young children. The implications of this literature for the development of the TedQL measure were also discussed.

6.1 Introduction

Self-report measures have traditionally been presented to adults and older children using written formats (Paulhus, 1991). Such written formats require respondents to read the items themselves and then give their answers on a response sheet (i.e., the questionnaire method). Our review in Chapter 3 revealed that the most common way to present items in measures targeted specifically at children below eight years has been to read items aloud (see p. 49). However other researchers have shown that presenting items with pictures or props as aids can be helpful for younger children by facilitating their understanding of items (Ernst et al., 1994, Harter & Pike, 1984), and therefore leading to more accurate and meaningful responses (Mize & Ladd, 1988, Mueller, 1996).

This chapter was divided into two sections. First, this chapter discussed the ways in which children's ability to respond to measures may be compromised, often inadvertently, by the way that items and measures are presented to children. These difficulties may be related to an imbalance of power inherent in adult-child relationships, or may stem from young children's lack of understanding of basic conversation rules, or young children's misinterpretations of the language used in items. Second, this chapter highlighted the techniques that can be employed to aid children in responding to items in self-report measures - from establishing rapport and giving children active roles in measures, to using readability formulas to assess the necessary skills required to answer items.

6.2 Compromising young children's ability respond to measures

We identified from the literature five ways that children's abilities to respond to items in measures can be compromised, including:

i. the impact of adult-child power relationships
ii. children's communication abilities and understanding of conversation rules
iii. children's inability to attend to the task demands
iv. children's understanding and use of language
v. children's concepts of time.

Impact of the adult-child power relationships

The nature of the relationship between 'adult' researchers and 'child' respondents can cause problems for researchers working with children. There is an imbalance of power that is implicit in any interactions between adults and children (Freeman, Sinha, & Condliffe, 1981). This power relationship can influence what children are prepared to report, and how children behave in the interview situation. First, children may be intimidated by adults, and if children are anxious this can hinder their ability to respond accurately, or to pay attention to the questions they are being asked. Alternatively, young children may give answers aimed to end the interaction as quickly as possible, and return to more attractive activities (Champion, Goodenough, Von Baeyer, & Thomas, 1998, Siegal, 1997). Researchers may ask questions that are not morally neutral, and this could cause children to assume that there is a 'right' answer and try to give the answer they think the researcher wants to hear rather than what they actually think (Goodenough, Champion, Laubreaux, Tabah, & Kampel, 1998).

Second, children are taught to follow commands from adults (e.g., at home - parents), and are expected to recognise that adults have more knowledge than they do (e.g., at school - teachers). This imbalance of power can mean children feel obliged to give answers when they are asked something by an adult researcher in an interview situation. Children may also attempt to answer questions they do not understand, as they are aware of the rules of turn taking in conversation and therefore feel pressured to answer the questions even if they do not understand them (Saywitz, 1990). This could mean children will guess answers or even invent information in their attempts to provide answers to the questions they are asked (Siegal, 1997). This imbalance of power may also cause children to keep silent for fear of criticism or fear of being wrong (Siegal, 1997).

Third, young children's responses may also be hindered by the fact that in other situations adults ask questions that imply indirect commands rather than a genuine interest in children's feelings (Champion et al., 1998, Flavell et al., 2002). For example, if a parent asks a child "Are you going to eat that chocolate bar before your dinner?", what they actually mean is "Don't eat that as you won't have room to eat your dinner".

Fourth, they may distrust the adult researcher or feel like they are being tricked, both of which could lead to children giving incorrect answers on purpose (Siegal, 1997). For example young children may form unfavourable perceptions of researchers if they feel they are being asked silly or obvious questions. If they feel the researcher already knows the answer they may wonder why they are being asked the question (as they may not understand the intentions behind the researcher's question, Siegal, 1997).

Children's communication abilities and understanding of conversation rules

Communication involves not only knowing what to say but also when and how to say it (Stone & Lemanek, 1990). Although by preschool age children have mastered at least a basic understanding of phonology, syntax, and semantics, they may not yet fully understand the pragmatics of language (i.e., the rules of communication which are learnt by experience and observations of parents and other adults). Younger children are less experienced in everyday conversations, and this may mean they do not disclose the full extent of their understanding, feelings, and thoughts when questioned using data collection methods such as questionnaires (Taplin et al., 1999). These difficulties can cause communication barriers that hinder young children's ability to respond to self-report measures (Taplin et al., 1999). Therefore asking children directly about their thoughts may not always be the most effective way of finding out information from children (Wilkinson, 1988).

By school age children understand the three basic maxims of conversations, i.e., that speakers' messages will be relevant, unambiguous and informative (Siegal & Waters, 1988, Siegal, 1997). However they may lack experience of situations where individuals depart from these maxims, for example when using irony, being humorous, or seeing further information (Siegal, 1997). Researchers may need to set aside conversation rules to probe the depth and certainty of children's understanding (Siegal & Peterson, 1996, Taplin et al., 1999). They may not understand the intent behind researchers asking them questions where the answer seems straightforward or obvious, and may therefore interpret this to mean they are required to give an alternative answer (Campbell & Rapee, 1996, Taplin et al., 1999). Therefore if researchers ask children questions that appear irrelevant or deceptive they may answer incorrectly not because they misunderstand the questions, but because they have been misled by the way the questions were asked (Siegal & Peterson, 1996).

Young children also may not respond well to repeated questioning. Instead of appreciating that the researcher is simply seeking affirmation of their previous answer, they may feel they are being told that their first answer was wrong and to change their answer (Ceci, 1991). This links to their experience with adults in other situations, like the school and family environment. For example, at school when children give an answer in class which is wrong, and teachers often use the technique of asking the child the question again, therefore sending an implicit message to the child that they should try again with an alternative answer. The conventions of these teacher-child dialogues can influence children's responses to researchers questioning (Rose & Blank, 1974). Siegal and Waters (1988) have reported evidence of this type of behaviour on number tasks in young children. Siegal and Waters (1988) showed that when children were subjected to repeated questioning they used a type of switching strategy. This strategy involved children changing their answers when they were asked the same question a second time, as they assumed their first answer was incorrect (Siegal & Waters, 1988). Therefore repeated questioning may convey ambiguity and mislead children to give inconsistent answers to questions they do understand. Repeated questioning may also cause children to invent details in an attempt to provide researchers with what they think they want to hear (Ceci, 1991, Ceci & Bruck, 1993). Goodnow (1994) pointed out that children may interpret 'why' questions as a signal that they should change their answers.

Children's inability to attend to the task demands

Children may answer inaccurately to questions or items if the demands of the task are too high (Banerjee, 1997). Younger children may be unable to appreciate fully the task demands of assessment or interview situations, and therefore be less able to match their behaviour to meet these demands (Martin, 1986). Older children may be more able to understand that they need to concentrate during the interview, and therefore try hard to pay attention for the time period required. However younger children may not realise the need for this effort (Dockrell, Lewis, & Lindsay, 2000).

Young children may also be more easily distracted by events outside of the interview, and their behaviour is more likely to be influenced by immediate circumstances or feelings (e.g., if they are tired, bored or anxious, Marsh, 1986). This may effect how much information they are willing to disclose during an interview (Marsh, 1986). Younger children also have shorter attention spans than older children, especially when they are using pencil and paper measures (Irwin, 1985). For these reasons measures need to be brief, relevant, and simple (Eiser & Morse, 2001).

Children's understanding and use of language

Although young children may be able to understand much more than has been previously thought (Flavell et al., 2002, Siegal, 1997), their language abilities may still present problems for researchers designing self-report measures. Most questionnaires designed for adults involve an advanced level of literacy - usually a reading age of thirteen or fourteen years or more (Titman, Smith, & Graham, 1997). Therefore presenting items in a written format may cause problems for children below eight years, as they may have difficulty reading and comprehending the words used (La Greca, 1990).

Children may not be able to understand the meanings of specific vocabulary in questionnaire items (Campbell & Rapee, 1996, Dockrell et al., 2000). Donaldson and Balfour (1968) pointed out that although preschoolers may understand the terms 'more' and 'less', they may still misunderstand these words when they are used in different contexts. Indeed misunderstanding words in the English language is common even with adults (e.g., 'Can germs grow bigger?' could mean either 'Do individual germs grow?' or 'do colonies of germs grow?', Siegal, 1997). The same words are used in different contexts in everyday conversations, and can have different meanings depending on these contexts. Young children have less experience with language, have a smaller vocabulary, and can misunderstand words used in items (Siegal, 1997).

An example of how such misunderstanding can occur when answering self-report items is shown when we consider the effect of negative item phrasing on young children's self-reports. Reverse wording of items has been traditionally used in self-report measures as a technique to avoid acquiescent response styles (e.g., Mehrens & Lehmann, 1983). However balancing item phrasing may cause more problems than it solves (McLaughlin, 1999, Rorer, 1965). Some researchers have shown that in adult measures the use of negative wording can make items more difficult to understand, and can distort the factor structure of a measure (Harrison, McLaughlin, & Coalter, 1996, Schriesheim, Eisenbach, & Hill, 1991).

Researchers such as Benson and Hovecar (1985) and Marsh (1986) have investigated the effect of item phrasing on young children's ability to understand measures. Benson and Hovecar (1985) considered whether item phrasing influenced the validity of attitude surveys used with school children. They showed that the insertion of 'not' had a profound effect on children's responses, in that children were less likely to indicate agreement with negatively worded items when they had agreed with equivalent positively worded items (Benson & Hovecar, 1985). Confirmatory factor analysis (CFA) also indicated that positive and negative items from the same attitude survey loaded on different factors from each other (Benson & Hovecar, 1985). Benson and Hovecar (1985) argued that school-aged children did not understand negation and therefore failed to reveal their true attitudes when faced with negatively worded scales. Marsh (1986) showed that seven to ten year old children reported lower self-concepts on negatively worded items, compared to their reported self-concepts to the same items positively worded. Consistent with Benson and Hovecar (1985), Marsh (1986) also found that negative items loaded differently to positive items in factor analysis. In addition the negative items were less consistent with the other items on the scale, and their removal increased the overall reliability of the measure considerably (Marsh, 1986).

However Chapman and Tunmer (1995) argued that the use of negative items in measures may not be so problematic. They pointed out that there may be subtle, but important, phrasing differences between the items Marsh (1986) and Benson and Hovecar (1985) used in their scales. Benson and Hovecar (1985) suggested that the use of negatively worded items confuses younger children, due to the linguistic complexity involved in disagreeing with a negative item to indicate the opposite is true. For example, disagreeing with that statement 'I do not like riding a bus to school' to report that they do like travelling by bus to school may be a difficult reasoning task for children under eight years. However Marsh (1986) used items that were affirmatively worded, using 'I' in the items to make statements that children had to agree or disagree with (e.g., 'I am dumb at reading').

Chapman and Tunmer (1995) argued that children would have less difficulty answering a question posed as "Are you...?", as opposed to verifying a statement beginning "I am...". They showed that when negative items were worded as interrogative statements using the referential pronoun 'you', children as young as five years could answer these items in the same way as positively worded items (Chapman & Tunmer, 1995). Chapman and Tunmer (1995) reported that the correlation coefficient between children's ratings on the positive and negative items was r=.04 when negative items were worded as 'I' statements, and this increased to r=.39 when negative items were worded as 'you' questions (Chapman & Tunmer, 1995). Indeed Akiyama and Guillory (1983) showed that children's understanding of the verification system (involved in agreeing or disagreeing with statements) develops later than their understanding of the answering system (used to answer questions posed with the referential pronoun 'you'). Chapman and Tunmer (1995) suggested that instead of removing negative items from measures (which actually may provide children with ways to indicate negative as well as positive self-concepts) researchers should word items as interrogative statements.

In summary, young children may not find negatively worded items difficult when items are worded as questions using 'you' (e.g., "Do you find it hard to tidy up your room by yourself?"). However items worded as statements using 'I' that children have to indicate agreement or disagreement with may be difficult for young children to answer (e.g., " I am not good at remembering things that I have to do at school").

Children's concepts of time

A variety of time frames have been used in self-report QOL measures - from a few days to several weeks (Eiser, & Morse, 2001). Some understanding of time and the ability to remember over given time frames is necessary for children to answer items in self-report measures (French & Christie, 1996). Some researchers have argued that children may not have the cognitive capacities to remember over the time frames required for some measures, or may not be able to separate the present from the past (Hinds, 1990, Hinds & Martin, 1988, Valla, 2000). Valla (2000) argued that young children tend to 'live' in the present world, which can mean they over-emphasise current events. Hinds and Martin (1988) argued that young children may have more difficulty recalling past events with accuracy. Palmer (1983) showed that young children's chronology of events often differs from the actual temporal ordering of events in real time.

However there is evidence that young children can accurately recall past events in sequence over short time periods (e.g., Fivush, Haden & Adam, 1995, Fivush, Haden & Reese, 1996, Freidman, 1990, 2000, Hudson & Fivush, 1991). Fivush et al. (1995, 1996) studied interactions between young children and mothers involving reminiscing about past events. Fivush et al. (1995) showed evidence that children's ability to recall past events was influenced by their mother's elaborative style (highly elaborative meant that mothers gave new information about past events during reminiscing). Reminiscing helps children learn how to recall past events, and children of highly elaborative mothers learn to recall events in greater detail than children of less elaborative mothers.

Freidman (1990, 2000) argued that three and four years old have an understanding of the passage of time, and knowledge of the duration of normal activities. Freidman (1990) showed that by three and four years of age children can discriminate between school activities that happened a week ago and events that happened seven weeks ago, and between a birthday one moths ago and a holiday nine months ago. Freidman (1990) argued that basic memory processes that are present in preschoolers permit an intuitive sense of ages of remembered events. Freidman (2000) reported that although four year olds failed to judge differences in the timing of future events they were successful in distinguishing time differences in past events (for example separating events which happened last month from events which happened a longer time ago). Freidman (2000) argued that a differentiated sense of time in past events precedes a differentiated sense of future or anticipated events. Therefore it may be that children under eight years can accurately recall past events, before they can accurately sequence future events (Freidman, 2000).

6.3 Techniques to aid children's understanding and communication abilities

Based on this review we considered how the presentation of child self-report measures that can help children understand and respond to items. We identified five techniques that could be incorporated in measures:

i. establishing rapport and safe environments
ii. assessing lower age boundaries using readability formulas
iii. reducing task demands using pictorial aids or props
iv. using age-appropriate language
v. using specific time recall periods.

Establishing rapport and safe environments

As discussed in the previous section (see p. 169-70), there are various ways that the adult-child relationship can impact on what children report or how they behave in interview situations. The imbalance of power inherent in interactions between adults and children can make children feel intimidated or anxious, which may cause them try to give the answer they think the researcher wants to hear rather than what they actually think. Children are taught to follow commands by adults in other situations, and this may mean they feel obliged to give answer even when they have none. Children may also distrust adult researchers if they feel they are being tricked, for example if researchers ask questions that seem obvious or silly.

Researchers need to think about how to bridge the gap between themselves as 'adult' researchers and the 'child' they are interviewing (Bendelow, France, & Williams, 1998). Children respond better to someone they can trust (Pollard, 1987). Being seen as not a 'proper adult' can help in establishing good rapport with children (Pollard, 1987). Children need to feel safe and happy in order to talk openly, honestly, and freely, and it is necessary to build up an 'ideal discourse' with young children (Boggs & Eyberg, 1990, Wilkinson, 1988). Wilkinson (1988) commented that in such 'ideal discourses' there is shared language and freedom to express thoughts without fear of judgement or criticism. Stone and Lemanek (1990) recommended that researchers should avoid long silences. This will help avoid children thinking they have answered incorrectly or becoming frustrated or defensive.

It takes time, effort, and skill to negotiate such situations, and to reduce young children's anxiety within the interview context. However the effort put into establishing rapport will be beneficial to researchers interviewing children. If tasks are set in a relevant, co-operative, and informative atmosphere, young children will be more willing to communicate their thoughts and feelings and to demonstrate their knowledge (Siegal, 1997). For example, play therapists have shown that children show enhanced verbal comprehension when engaged with a listener who can adjust for, and find meaning in their expressions (Boggs & Eyberg, 1990, Schaefer & O'Connor, 1983).

Assessing lower age boundaries using readability formulas

There are communication barriers facing researchers working with young children, as highlighted previously (see p. 171-72). Young children have less experience with everyday conversations and this can mean they do not disclose all their knowledge in interview situations. By school age children understand the basic maxims of conversations, but they may become confused when researchers set these maxims aside to probe their understanding further. Young children may also not respond well to repeated questioning as they may take this as indication that their first answer was wrong and change their answer the second time.

Stone and Lemanek (1990) argued that researchers should integrate developmental perspectives into their measures to avoid such communication problems. It would be useful for researchers designing measures to determine if the task is within the capabilities of the children that their instrument is targeted at (i.e., whether the complexities of the test procedure match the child's communication abilities and if the questions require information processing or memory skills that children have not yet acquired).

Beitchman and Corradini (1988) recommended that the reading level of new instruments should be assessed using "readability formulas". Such formulas provide information on the reading age required for answering the items in a measure. One example is Flesch's (1951) reading formula which assesses word length and syllable count to provide an average reading level. This formula has been used in the development stages of some self-report measures. For example, Mishoe et al. (1998) used Flesch's (1951) reading formula to assess the reading age required for their asthma

QOL measure (About my Asthma, AMA). On the basis of this, Mishoe et al. (1998) reported that their AMA measure was appropriate for children over six years of age.

Reducing task demands using pictorial aid or props

As discussed previously (see p. 172-3), younger children may answer inaccurately to questions if the task demands of an interview situation are too high, and they may be less able than older children to match their behaviour to meet the demands of the task. Young children can also be distracted easily by events outside of the interview, and their behaviour is easily influenced by immediate circumstances or feelings. Young children also have shorter attention spans than older children, especially when performing verbal or written tasks. Indeed Garbarino and Stott (1992) have suggested that there may be a limit on the amount of information young children can convey with verbal methods alone.

To help overcome there problems, Flanery (1990) recommended piloting measures to ensure that the demands in completing measures do not exceed young children's capabilities. Stone and Lemanek (1990) argued that children should also be encouraged to take an active role in measures. Researchers need to be able to maintain children's attention and interest throughout measures. Some researchers suggested the use of pictures to help make measures more attractive to young children, and to maintain their attention (e.g., Harter & Pike, 1984). Researchers using pictorial support for items in their measures have reported good levels of reliability for self-reports from four year old children on measures of self-concepts (Harter & Pike, 1984); anxiety (Venham & Gaulin-Kremer, 1979); and depression (Martini et al., 1990).

However despite attempts to make self-report measure more appealing using pictures and icons, there is some research showing that pictures are less effective than three-dimensional objects (i.e., puppets, dolls or teddy bears) in facilitating memory and other cognitive processes (e.g., DeLoache, 1986, Hartley, 1976, Steinberg, 1974). Such three-dimensional aids may clarify and concretise items to children (Ernst et al., 1994), and therefore lead to more meaningful responding (Mize & Ladd, 1988). Greenspan & Greenspan (1991) reported that young children themselves prefer toys and life-like props over pictures for self-expression.

Props can increase the length of responses by enabling children to enact their thoughts and feelings (Salmon, 2001). The use of props in child measures has been associated with longer and more detailed responses from children (Bernhart & Prager, 1985, Getz, Goldman & Corsini, 1984, Irwin, 1985, Mize & Ladd, 1988, Mueller, 1996). Getz et al. (1984) reported that the use of dolls not only encouraged preschoolers to enact their responses, but also produced a greater variety of responses to social dilemmas. Mize and Ladd (1988) used puppets to ask four and five year old children about hypothetical social dilemmas (e.g., "You are building a tall tower with blocks. Another kid comes over, and knocks down your tower, crash, and says 'I was playing with those before and you can't play with them"). Mize and Ladd (1988) found that children produced more detailed responses to questions when puppets acted out the items, compared to when items were read aloud without the additional of puppets. Mize and Ladd (1988) also found that the children's enactive responses (i.e., playing out their strategy for a situation using the puppets) were more often significant predictors of their behaviour than responses obtained using a verbal method (i.e., asking children what they would do in the same situation).

Mueller (1996) developed a measure called "Teddy Bears Picnic" using teddy bears to illustrate nine incomplete story stems. Mueller (1996) reported that children's story telling can produce relevant information about their emotional health and well-being, as the children's coded responses on the teddy bear measure distinguished between children with emotional and behavioural problems. Measelle et al. (1998) developed the Berkeley Puppet Interview, a measure which used two identical hand puppets called 'Iggy' and 'Ziggy' to ask four to seven year old children about their academic self-concepts. They found evidence that four year old children could report perceptions of themselves that were reliable and consistent over time (Measelle et al., 1998).

Props may also augment responses by acting as memory aids, as young children often have more difficulty spontaneously retrieving information from memory (Schneider & Bjorkland, 1998). Young children may find the task of recalling information about themselves easier when they have external aids for recall, such as dolls, toys, or models (Salmon, Bidrose, & Pipe, 1995) Props may also decrease the cognitive demands of memory tasks as they remain present while the child complete items (unlike verbal prompts) so reducing the load on the child's memory (Patterson, 1995). The use of props may help to reduce the social and emotional demands of the task by minimising the amount of intrusion into the child's 'world' (Salmon, 2001). Measelle et al. (1998) argued that the use of puppets in their measure meant children could respond non-verbally to items (i.e., by pointing to the puppet which they felt was most similar to themselves), which may be less intimidating than having to verbalise their responses.

DeLoache and Marzolf (1995) argued that the effectiveness of props is influenced by children's ability to detect and respond to the correspondence between the prop and the referent (the item it represents). Children need to understand that props stand for something and then link this back to themselves (Salmon et al., 1995). DeLoache (2000) provided evidence that this understanding develops around three years of age. After three years children can recognise that symbolic objects have both a concrete and an abstract reality (termed dual representation, i.e., a model of a car exists as a model physically, but also represents the term 'car', DeLoache, 2000). DeLoache (2000) suggested that this cognitive development may be related to experience (in that three year olds engage in much symbolic play and have more experience with symbols than two year olds). School-aged children can benefit from the use of dolls or toys in interview situations without this marked cost to accuracy (e.g., Gordon, Ornstein, Nida, Follmer, Crenshaw, & Albert, 1993, Saywitz, Goodman, Nicholas, & Moan, 1991). However children under three years may be unable to understand that a doll (or teddy bear) represents themselves due their lack of understanding of dual representation and also to the salience of the object's identity as a toy (DeLoache, 2000).

Using age-appropriate language

Children may have problems understanding specific vocabulary or wording used in measures (see p. 173). The same words can have different meanings depending on the context that they are placed in, and this can be a source of confusion for young children (see p. 173). Researchers have often used negative phrasing in measures to help avoid acquiescent response styles, however such phrasing used in child measures may cause mis-understanding. As discussed previously (see p. 173-4), both Chapman and Tunmer (1995) and Marsh et al. (1998) have argued that phrasing items as questions will be more natural and easier for young children to understand compared with using first-person declarative statements (i.e., phrasing items as questions beginning with "Are you...?" for children to answer, as opposed to "I am..." statements that they have to agree or disagree with).

The language used in child measures needs to be age-appropriate and relevant to children's everyday experiences (Eiser & Morse, 2000, Marsh et al., 1998). Researchers should try to phrase items using language that is as simple, straightforward, and clear as possible (Eiser et al., 2000). Questions need to be phrased as explicitly as possible, avoiding ambiguous wording or repeated questioning wherever possible (Siegal, 1997). Questions should be phrased in ways that young children can understand, as children may have difficulty if terms are unfamiliar to them or they can not read certain words (La Greca, 1990). Simple vocabulary and short sentences that only communicate one idea at a time, and questions that require concrete answers should be used (Flannery, 1990). This will help avoid inconsistent answers from children due to misunderstanding of the task rather than a lack of knowledge.

Using specific time recall periods

As discussed in the previous section, there is evidence that young children are able to recall and rate the frequency of their behaviours or feelings over a narrow time period (see p. 175-6). Researchers have shown that by three and four years of age children have a differentiated sense of past events, and their ability to recall and sequence past events precedes their ability to do so for future events (see p. 176).

Winkielman, Knuper and Schwartz (1998) recommended that the time periods used should be as short as possible to make it easier for young children to recall their behaviour or feelings. A short time period will also help to avoid recall bias – as the longer the recall period the greater the expected bias in the reporting of episodic information (Mathiowetz, 2000).

Kieckhefer (1987, 1988) reported that although school children could not report reliably on their "usual health", they could provide reliable responses when asked to rate "today's health". Therefore questions in child measures should require children to report on the frequency at which certain behaviours have occurred during a specified period. An example of one such measure using a short time period is the PedsQL (Varni et al., 1998). In the PedQL children are asked: "How much of a problem have you had with ...... *in the last few weeks*". Riley et al. (2001) also used a short time period in their measure (CHIP-CE). In the CHIP-CE children are asked: "*In the past week,* how often did you have a stomach ache?". Ravens-Sieberer, Thomas, Kluth, Teschke, Lilienthal, and Bullinger (2001) also used a one week time period for recall in their Kiddy-KINDL measure, where the children are asked: "*During the last week* I got on well with my parents (never to very often)". Bruil (1999) used the same recall period in the How Are You? measure, with children being asked: "Have you remembered what you learned at school *during the past seven days*?".

6.4 Implications for the development of the TedQL measure

Our review of the literature revealed that children's ability to respond to items within self-report measures can be hampered in various ways (see p. 170-6). While recognising these problems, researchers have also suggested ways to aid children in answering self-report items (see p. 177-83). Implications for the development of our TedQL measure have been summarised in the following sections.

Resolving adult-child power imbalances

Our review highlighted the need to establish trust between the adult researcher and child interviewee, and ensure children feel safe and able to express their opinions (see p. 170 & 177-8). There is an imbalance of power implicit in any interaction between children and adults, and this can cause problems for researchers attempting to gain accurate self-reports from young children. Spending time with children beforehand and getting to know them in their normal environment can help establish rapport and therefore eliminate some of this power imbalance (see p. 178). We recognised the importance of establishing rapport with the children beforehand, and therefore we spent several sessions with the children in their classroom environment before asking them to complete our TedQL measure (see Chapter 4, p. 103).

We also recognised that children may feel the need to answer items regardless of whether they actually understood the questions due to this power imbalance (see p. 170). We attempted to avoid this problem by the use of a "don't know" option in our measure in Study 4 (see Chapter 8, p. 221). The children were trained to use the don't know option (a blue question mark, see Figure 8.2, p. 221) if they didn't know the answer or didn't understand the question. We felt this would help avoid some of the social desirability bias that can be present in young children (i.e., to avoid children answering questions to please the interviewer, Campbell & Rapee, 1996, Hughes, 1984).

Reducing communication problems

Young children may also have problems expressing their thoughts and feelings due to their lack of experience in everyday conversations (see p. 171-72). While young children may understand the basic maxims of conversations, researchers may set aside these rules during interviews and such departures may confuse young children. For example, young children may change their answers or invent information when they are faced with questions that seem too obvious or irrelevant (see p. 172). Young children may not respond well to repeated questioning, as they may assume that they are being told their first answer was wrong and hence may change their answer the second time (see p. 172). We considered this when we attempted to test the reproducibility of children's responses on the TedQL.3 measure in Study 4 (see Chapters 8, p. 223), and therefore used two separate time points to re-test children (spaced by one week to avoid recall bias).

Maintaining children's attention and interest

We also recognised that young children get distracted easily and are unable to concentrate for long periods of time (see p. 172-3). We therefore made our measure as short and as attractive as possible. We presented the items to children in a game format using teddy bears as props. Given that children produce more meaningful responses when using props as memory aids (see p. 179-80), we felt three-dimensional props would be more helpful than pictures for our measure. The research reviewed in this chapter also highlighted that props can be helpful in engaging children's attention, and maintaining it throughout a given task (see p. 179-80). The results of our first study revealed that children themselves reported that they found our teddy bear measure easier and more enjoyable to use than a similar QOL measure using a verbal presentation style – the PedsQLTM4.0 (see Chapter 4, p. 112).

Our review revealed that props help reduce the load on young children's memory, as the props remain present while children recall information to answer items (see p. 180-1). In addition, young children can respond to items non-verbally if they wish, by simply pointing to the props. We designed the TedQL measure so that children could either respond verbally or non-verbally using the teddy bears (see Chapter 3, p. 70-1).

Avoiding language problems and assessing readability

The research reviewed in this chapter also highlighted the importance of the language and phrasing used in items (see p. 173-5). We recognised the need to use simple, clear, and unambiguous language to avoid misinterpretation. We designed Study 3 as an interview study, with the aim to use information obtained directly from children themselves to inform the development of our measure (see Chapter 5, p. 136-7). We used content analysis to identify the areas most important in young children's lives, and we used the results of this study to alter, remove, and add items to our measure to produce a new version (the TedQL.3, see Chapter 5, p. 162-3).

We attempted to phrase items in an appropriate and helpful way for young children. Based on the recommendations made by researchers (see p. 175 & p. 181-2) such as Chapman & Tunmer (1995) and Marsh et al. (1998), we designed our TedQL items using interrogative questions for children to answer (such as "How good are you at running?"), rather than using first person declarative statements that children had to agree or disagree with (e.g., "I am not good at tying my shoes", see Chapter 4, p. 103-4).

We tested the readability of our TedQL measure to ensure children aged between three and eight years would be able to understand the words and phrases used in our self-report items. We used Flesch's readability formula (Flesch-Kincaid Grade Level) to calculate the reading age required for the TedQL.4 in Study 5 (see Chapter 10, p. 288).

Using specific time recall periods

The literature reviewed in this chapter also highlighted the importance of using specific recall periods for items (i.e., giving children a time reference for recalling their behaviours or feelings, see p. 182-3). There is evidence that young children can recall information accurately over short periods of time (e.g., Fivush et al., 1996, Hudson & Fivush, 1991, see p. 176), and that four year old children can accurately recall past events over a specific time period (Friedman, 2000). However researchers have argued that young children are still much more present-oriented than adults (Hinds, 1990), and using a narrow recall period can be helpful in measures aimed at children below eight years (see p. 182). Following from the literature reviewed in this chapter, we designed our TedQL measure to use a recall period of one week (see Chapter 4, p. 101) – as short a time period as possible to avoid recall bias and to aid children in the recall task for items.

The following chapter (Chapter 7) considered more issues relevant to the design and development of our child self-report measure. In this chapter we focused specifically on how young children respond to the content and presentation style of measures, and in Chapter 7 we focused on whether young children can understand the response scales used to answer self-report items.

Chapter 7: Young children's ability to self-report – understanding and use of response scales.

Summary

The measurement concerns that arise when children use response scales to answer self-report items were considered. Children's understanding of the anchors used to differentiate between points on a scale can influence the scaling properties of a measure. Young children may have problems understanding vague quantifiers such as *often* or *sometimes*. Children's cognitive and developmental status impacts on their ability to understand and use response scales. There is evidence that younger children are just as able as older children to recall and rate the frequency of their behaviour, thoughts, and feelings. Children are also prone to response biases that can impact on their self-reports. Research has found that position biases, acquiescence bias, positive response bias, and negative item bias are all more common in younger than older children. The type of scale used and the number of points included on a scale can also influence the quality of children's responses.

We also discussed the suggestions that have been made to help overcome these problems, and to make the rating task more understandable to children. Researchers can test children's understanding of the words used as anchors in their measures, and use specific anchors such as *very true/not true of me* to avoid mis-understanding. Graphic response scales, such as the Poker Chip Tool or the Pain Thermometer, may help ease the rating task for children. Researchers can also assess the relative value of different response scales for use with instruments that will enable informed decisions on choice of scale. Calibration tasks or hypothetical questions can be used to assess children's rating abilities and their understanding of a given response scale. The implications of this literature for the development of the TedQL measure were also discussed.

7.1 Introduction

Quantitative response scales have been used to represent the response choices to respondents for self-report items. These types of scales have been extended to measures targeted specifically at children. A challenge to any child measure is whether children can understand and use the response scales employed in self-report measures.

The research into response styles in adults (Paulhus, 1991, Schwartz & Sprangers, 1999) shows that adults are prone to a variety of biases and that their responses may be affected by scale length, scale type, and the anchors used to represent the response options. However despite the use of response scales in child self-report measures, there have been few attempts to test children's abilities to use these scales. This chapter discussed the measurement concerns that arise when using response scales with young children, and the suggestions that have been made to help researchers overcome these difficulties. We also considered the implications the literature reviewed in this chapter for the development of our TedQL measure, specifically for our choice of response scale type.

7.2 Measurement concerns

Five measurement concerns were identified from the literature that arise when children use response scales to answer self-report items. These were:

i. children's understanding of the anchors used in response scales

ii. the effect of children's cognitive and developmental status (on their ability to respond appropriately using response scales)

iii. the impact of common response biases on children's responses

iv. the effect of the response scale type on children's responses

v. the impact of the number of response options on children's responses.

Children's understanding of the anchors used in response scales

When considering young children's ability to use response scales a relevant issue is whether they understand the anchors used in these scales. Anchors are used to differentiate each of the response options of a given scale (e.g., 0=*not at all*, 1=*sometimes*, 2=*always*). Children's understanding of the anchors can influence on the sensitivity and scaling properties of measures (e.g., what children understand by the words used to describe anchors, like *sometimes* and *often*, Horn & Munafo, 1997, McLaughlin & Bjornson, 1999, Schwartz & Sprangers, 1999).

The use of verbal anchors is based on three assumptions: that respondents i) understand the terms used; ii) interpret and use these terms in a similar way to other respondents who are subsequently given the same measure; and iii) understand the concepts of time involved. It is important for researchers that one person's use of the anchor sometimes is equivalent to other people's use of the same anchor when completing a measure. However, some researchers have argued that the meaning of anchors depends on who is using it and what item is being rated (Stone, 1992).

Schwartz and Sprangers (1999) pointed out that researchers developing self-report measures have often used vague quantifiers (such as *sometimes* and *frequently*) to avoid influencing responses by using too specified response alternatives (such as *three times a week*). Schwartz and Sprangers (1999) argued that this may not be the best solution, as the same expression (such as *occasionally*) can denote different frequencies in different domains and for different behaviours (e.g., frequently drinking beer is not the same as frequently drinking coffee).

The effect of children's cognitive and developmental status (on their ability to respond appropriately using response scales)

Young children may lack the cognitive skills necessary to use response scales to rate their answers to items in self-report measures (Beitchman & Corradini, 1988). Young children's cognitive abilities are an important determinant of whether they are able to successfully use response scales, and perform the tasks involved in classification and seriation (Taplin et al., 1999).

Items in self-report measures require children to process the question, retrieve relevant information, determine the relevance of that information to the question asked, and produce a response (Mathiowetz, 2000). Dockrell et al. (2000) has argued that self-report measures often rely on higher order cognitive abilities which young children may not have developed and therefore they are unable to meet the demands involved in rating tasks. Children below eight years of age may not have the cognitive maturation needed to think abstractly about their current physical or psychological states, relate their current state to other states in the past, order their current experiences, feelings or thoughts into degrees and then project these onto a subjective continuum (Taplin et al., 1999). Some types of questions may require complex information processing skills that young children may find difficult to master (Martin, 1986). For example, the question "Do you like going to school?" requires children to remember positive and negative events, and to weigh those events before giving as response (Martin, 1986).

However other researchers have argued that children as young as three years do have the skills necessary to rate self-report items using response scales (e.g., Champion, Goodenough, von Baeyer, & Thomas, 1998, Zelter, LeBaron, Richie, Reed, Schoolfield, & Prihoda, 1988). Cognitive skills such as classification, seriation, matching, and estimation develop around the age of three and four years (Champion et al., 1998). Zelter et al. (1988) assessed whether five year old children could use a rating scale to rate items in a similar way to adults. Zelter et al. (1988) reported no differences between children's and adult's ratings of somatic symptoms.

It has also been debated as to whether children below eight years of age are reliable reporters of their behaviour for measures which require not only qualitative recall (i.e., type of behaviour) but also frequency recall (e.g., how often eaten, or exercised in a given time period) (Johnson, 1991). For example, Johnson (1991) argued that young children have difficulty recalling the frequency of behaviours that require understanding of more abstract concepts, such as time. However there are a number of studies which support the view that frequency of occurrence is encoded automatically (e.g., Harris, Durso, Mergler, & Jones, 1990). Following from this premise, Harris, Durso, Mergler, and Jones (1990) have argued that younger children should be just as able as older children to judge frequency of their behaviour, thoughts, and feelings.

The impact of common response biases on children's responses

Young children may be prone to various responses biases that can lead to inaccurate responses to items in self-report measures. Children under eight years of age may be more likely than older children to produce extreme response patterns in self-report items (i.e., focus on either end of a given response scale, Goodenough et al., 1997). There is evidence that young children tend to engage in dictonomous thinking (Gelman & Baillargeon, 1983). Children may be more likely to view situations, experiences, and feelings as extremes (e.g., they feel either happy or sad) and this may make it difficult for them to appreciate and use the finer grades between the endpoints of response scales. Indeed children's behaviour is frequently treated as binary in home and school situations (for example, a child either goes to bed or does not). This tendency for dichotomous thinking may cause children to focus on the extreme ends of response scales, and alter scales to a yes/no format (Gelman & Baillargeon, 1983).

Researchers have found evidence for these extreme response patterns in younger children (e.g., Chambers & Craig, 1998, Chambers & Johnston, 2002). Chambers and Craig (1998) reported that when asked to rate their pain experiences, younger children were more likely than older children to chose the extreme points of the scale (i.e., *a lot* or *not at all*). However, other work has shown that choosing extremes may be related more to the question being asked than the age of the children (Chambers & Johnston, 2002). Chambers and Johnston (2002) examined children's response patterns to three different types of questions. These were: a physical characteristic task (e.g., children with different heights); a social objective task (e.g., children winning different positions in a race); and a subjective task (e.g., wanting to go to the movies today and going tomorrow). Regardless of age children responded in a more extreme manner when rating their own and other's emotions and feelings (i.e., the subjective task, Chambers & Johnston, 2002). Chambers and Johnston (2002) concluded that it may be something specific about rating emotions that causes children of all ages to produce more extreme responses.

Younger children may also be more prone to a number of other response biases that can lead to inaccurate responding. For example, researchers have found evidence that position biases, acquiescence bias, positive response bias, and negative item bias are all more common in children than adults (Garbarino & Stott, 1992, McBrien, & Dagenbach, 1998, Pantell & Lewis, 1987, Warren, Hulse-Trotter, & Tubbs, 1991). Warren et al. (1991) studied suggestibility in children, and found that young children were more likely to acquiesce to leading questions (i.e., to agree with the researchers regardless of the content of the question). Additionally McBrien and Dagenbach (1998) reported that when three and four year olds were asked to remember, and questioned about, an imagined event, a strong positive response bias was evident (i.e., tendency to say "yes" to questions).

Another problem which may also lead to inaccurate responding is that younger children are less likely to ask for classification or to indicate uncertainty when asked questions that do not make sense. For example, Hughes and Grieve (1980) found that virtually all the five to six year olds they interviewed would try to give an answer to a variety of nonsense questions (such as "Is red heavier than yellow?"). Hughes and Grieve (1980) reported that the seven year olds more frequently indicated uncertainty, but the children under six years rarely did so. Pratt (1990) asked children similar nonsense questions (e.g., "Is a fork happier than a knife?"), and showed that younger children were more likely to attempt to answer the question than to say that the question was silly.

The effect of the response scale type on children's responses

There are a variety of response scales that have been used by scale developers in their measures (from Likert, graphic, facial expression, and visual analogue, see Chapter 3, p. 46). Researchers need to consider the effect of scale type on children's answers. Respondents may use contextual information provided by the way response scales are presented to make sense of the question they are being asked, and to understand the frequency of behaviours, feelings, or abilities that researchers are interested in (Schwartz & Sprangers, 1999). For example, if numerical anchors range from positive to negative (e.g., -5 to +5) respondents are likely to interpret the dimension as bipolar (i.e., that the two poles of the scale refer to the presence of opposite attributes, Knauper & Turner, 2003). Alternatively, if the numerical anchors are composed of positive numbers (such as 1 to 5) then respondents are likely to interpret the dimension as unipolar (i.e., referring to degrees of the same attribute, Knauper & Turner, 2003). The way response scales are set out (e.g., bipolar or continuous) gives information to the respondent on what the researcher is interested in, and may influence how the respondent views the attribute in question, or how they report the frequency of their behaviours.

Schwartz, Grayson, and Knauper (1998) also showed that the graphical layout of the response scales affected the frequency of behaviours reported by adult respondents. Respondents were found to be more likely to endorse a value in the lower half of a scale when the scale was presented as a pyramid, than as stacked boxes of the same size (Schwartz et al., 1998). Schwartz et al. (1998) argued that respondents used the graphical layout of the scale to determine the intended meaning of the scale points, which in turn influenced the frequency of behaviour reported.

Such research illustrates that respondents use information provided (often unwittingly) by the types of response scales researchers use in their measures to help determine the meaning of the questions that they are asked and the frequency of behaviours, feelings, or abilities on which they should report (Knauper & Turner, 2003). The meaning conveyed by the scales used in a response scale needs to match with the types of items being rated (McLaughlin, 1999, Schwartz et al., 1998). Knauper and Turner (2003) argued that this has been frequently neglected in developing both adult and child QOL measures.

The impact of the number of response options on children's responses

Young children's ability to self-report accurately may also be affected by the length of the response scales employed (i.e., the number of points on a scale). Researchers developing pain self-report measures have argued it is important not to present young children with too many options, as this may cause their responses to gravitate towards the end points of a scale – essentially meaning they are not using the full response options available (Arts et al., 1994, Champion et al., 1993, Goodenough, Champion, von Baeyer, & Ziegler, 1997). Therefore a linear visual analogue scale may not be appropriate for younger children as it offers an indeterminate array of choices (i.e., the child can point to any given point on the whole scale, rather than choosing between four options). Following from this argument, some researchers have suggested that for children under six years response scales should be limited to three or four options (Arts et al., 1994, Hester, Foster, & Kristensen, 1990).

However reducing the number of points does not mean young children can understand and use these response scales. Reducing the number of options on a scale may serve to inflate responses, by forcing a higher choice of category (Goodenough, Champion, McInerney, Taplin, & Ziegler, 1999). Some researchers have suggested an alternative way to help children understand the rating task, and use fewer response options (Eder, 1990, Harter & Pike, 1984). For example, Harter and Piker (1984) in their measure (Pictorial Scale of Perceived Competence and Social Acceptance) used a four-point scale which was presented to children as bipolar. They achieved this by first asking children to choose which of two different pictures they were most like, then further probing them as to whether they were *a lot/always* like this or *a little bit/sometimes* like this. This meant essentially children provided four-point responses using two consecutive bipolar scales. This kind of technique may be useful when working with younger children who may be confused when faced with more than two or three response options.

Johnston (1998) argued that there is no definitive answer to the appropriate number of points, as it is dependent on the construct being measured and the sensitivity required for a given measure. Reducing the number of response points on a response scale may be at the expense of sensitivity, as there is often a trade-off between the number of options available and the sensitivity of the scale (Goodenough et al., 1997).

7.3 Techniques to enhance children's understanding

A number of suggestions have been made to help overcome children's difficulties in using and understanding conventional response scales. These suggestions include:

i. assessment of whether children understand the anchors used

ii. the use of creative ways to represent response scales

iii. comparison of the relative value of different response scales

iv. the use of a calibration task to test children's rating abilities

v. the use of a training period involving practise questions.

Assessment of whether children understand the anchors used

Adjustments to the verbal anchors used in response scales with adults may be needed to ensure children can understand the language and concepts involved (French, Christie, & West, 1994). Young children may not understand some of the fine-graded linguistic differences between verbal anchors used in response scales. For example. a five-year-old may not understand the difference between *sometimes* and *often.* There is a need to test children's understanding of anchors when designing and using response scales in self-report measures (Fantuzzo, McDermott, Holliday, Manz, Hampton, & Burdick, 1996). If researchers want to compare child reports with parents or other proxy reports, then they need to consider whether adults and children interpret and use these common verbal anchors in similar ways to each other (Schwartz, Grayson, & Knauper, 1998). If adults understand and use anchors in different ways to children (e.g., children refer to *a lot* when adults choose *a little bit*) this would have implications for comparisons between child and parent reports. For example, differences in use may mean that researchers conclude a lack of concordance between parent and child QOL reports, when in fact they are both reporting the same levels of functioning but their use and interpretation of the anchors in the scale differs.

Stone and Kennedy-Moore (1992) have suggested using anchors like this item is *very true/not true of me* or *very similar/different to me* rather than vague frequency based anchors. Juniper et al. (1996) used anchors in the Pediatric Asthma Quality of Life Questionnaire (PAQLQ) that were generated from words actually used by children to describe their problems. This technique may avoid problems of misinterpretation or misunderstanding of anchors by younger children, but may also limit the comparability of children's responses with proxy ratings.

Fantuzzo et al. (1996) conducted a study into whether children could understand and differentiate between the verbal anchors used in the PCPCSA, designed by Harter and Pike (1984). Despite Harter and Pike's (1984) assertion that this measure is developmentally appropriate for preschool children, Fantuzzo et al. (1996) argued that no empirical studies have tested the appropriateness of the verbal anchors used in the PSPCSA. Fantuzzo et al. (1996) developed a test to assess whether children could distinguish between the various quantities that were used as verbal anchors in the PSPCSA. This testing involved a procedure with cups and balls, representing various concepts of quantity (e.g., a cup with *a lot* of balls in it compared with a cup with *hardly any* balls in it) and children had to identify cups with these stated quantities in them (e.g., "Which cup has *hardly any* balls in it?"). They found that on average four and five year old children correctly identified only ten out of sixteen quantity concepts, and only one child demonstrated 100% comprehension of all the quantity concepts (Fantuzzo et al., 1996). The results of this study raised questions about the developmental appropriateness of the PSPCSA for preschool children. Their work highlighted the importance of testing whether young children understand verbal anchors. This testing should form part of the development stages of any child self-report measure.

Using creative ways to represent response scales

There are a number of response scales available to scale developers. The most common for work with children include: Likert; graphic; facial expression; and visual analogue (see Chapter 3, p. 46). Different forms of graphic response scales have been developed to try to make scales more concrete and understandable to children. Dockrell et al. (2000) reported that devices such as a rolled/unrolled toothpaste tube and drawing miniature curtains and rolling fabric over a scale (i.e., to represent opening and shutting curtains to help understand the rating task involves considering how much a given behaviour happens) have been used with children.

Another example of a graphic scale used frequently in pain self-report measures is the Poker Chip Tool which has been used with children as young as three years old (Hester, Foster, & Kristensen, 1990). This measure uses a concrete ordinal scale as an alternative to a visual analogue scale (Ready & Edwards, 1992). Children rate their pain by choosing one of four poker chips (each representing a *piece of hurt* with four chips representing *the most pain possible*). This type of scale may help children to appreciate the rating task – by showing how to rate frequency with varying numbers of chips.

Another graphic response scale type is the Pain Thermometer which is a visual analogue scale using a picture of a thermometer numbered from 0-100 where zero represents the *absence of pain* and 100 the *worst pain possible* (Belter et al., 1988, Szyfelbein et al., 1985). A different version of the Pain Thermometer (the Oucher, Beyer & Knapp, 1986) uses both a numerical scale and six photographs of children experiencing increasing amounts of pain.

Another option is the Coloured Analogue Scale that varies in three dimensions (colour, width, and length) to help children to see concretely how each of the scale points would represent different values in pain intensity (McGrath, 1996). There is evidence that this format may be easier to use than a traditional visual analogue scale (Champion et al., 1997), as it does not require children to understand numbers or verbal anchors. Another innovative pain measure used a toy koala on a vertical wooden pole to rate items (Goodenough et al., 1997). Children have to move the koala to the place on the pole to show how much pain they are feeling (with the bottom representing *no pain* and the top representing the *most pain possible*). These different ways to represent response scales to children may help to improve children's understanding of response scales, ease the rating task, and lead to more accurate ratings.

Comparing the relative value of different response scales

Our review of child self-report measures in Chapter 3 discussed the arguments for and against using different types of response scales, and we pointed out that authors should to provide psychometric evidence for their scale choice (p. 47-49). From our conclusions in Chapter 3 it was apparent that there was a need to judge the relative value of different response scale types (p. 65). Such assessments could be made by directly comparing the psychometric properties of measures across different response scales (i.e., to provide evidence of which scale produces the most consistent and reliable responses from children). Wallander, Schmitt, and Koot (2001) have pointed out that researchers rarely report their empirical basis for selecting the type of scale used in their measures, or which scales are suitable for which levels of competence. Studies comparing the relative value of response scales for instruments could help researchers make informed judgements on the most appropriate response scale for their measures.

Some researchers have designed studies which allow these relative judgements to be made (e.g., Juniper, Guyatt, Feeny, Griffith, & Ferrie, 1997, Rebok et al., 2001). Rebok et al. (2001) gave children a visual analogue scale and a graphic scale (consisting of circles gradually increasing in size) to complete items in their CHIP-CE measure. They asked children which scale they preferred to use, and found three-quarters preferred the graphic circular scale (Rebok et al., 2001). Juniper et al. (1997) investigated the minimum age and reading skills required to complete four child QOL measures, each employing a different response scale. The Paediatric Asthma Quality of Life Questionnaire (PAQLQ) (Juniper, 1996) used a traditional seven-point response scale. The Health Utilities Index (HUI) (Feeny, Furlong, Boyle, & Torrance, 1995) scored the level of impairment on ten attributes using dichotomous responses. The Feeling Thermometer (Torrance, Furlon, Feeny, & Boyle, 1986) was a graphic scale on which children used a thermometer scale from zero to 100. The Standard Gamble (Torrance et al., 1986) used a different procedure: children decided whether they wish to stick with the certainty of their current health state or gamble with a treatment that could result in either perfect health or death. Juniper et al. (1997) reported that the PAQLQ and Feeling Thermometer were most appropriate for (and could be used reliably by) children aged seven years and above. Most children could complete the HUI alone, but the validity was poor for this measure. However, children under twelve years old found the Standard Gamble difficult to use (Juniper et al., 1997), which is not surprising considering the complex concepts involved in using this instrument (e.g., concept of death).

Champion et al. (2000), Goodenough et al. (1997), and Chambers et al. (1999) have investigated the relative value of response scales for child pain measures. Champion et al. (2000) compared children's ratings of their pain across six different scales, by considering whether children's scores were correlated across these scales. Champion et al. (2000) found children's pain ratings were related across all six scale types. Goodenough et al. (1997) performed similar comparisons and reported that children's pain ratings on different scales all loaded on the same factor and therefore appeared to be measuring the same construct – i.e., pain. Chambers et al. (1999) compared five different pain measures that all used facial expression scales. Despite finding high correlations among children's pain rating across all five scales, Chambers et al. (1999) reported significantly different mean pain ratings.

Using a calibration task to test rating abilities

It is important to establish whether young children can appreciate, understand, and rate frequencies when they are using response scales to answer self-report items. Siegal (1997) argued that young children can understand relations and graduations if tasks are made explicit and attractive to them. For example, Goswami (1995) found that three and four year olds could choose mum, dad, and baby sized cups from a set of cups after having been read the story of Goldilocks and the three bears. They also chose the same relative sized cups from a different set of cups, showing that they could ignore the absolute size of the original set and choose relatively. In addition, they also chose the same relative sizes of chocolate and pizza.

However few researchers have considered whether children can use the different graduations within the response scales employed in their measures. Quittner et al. (1998, 2000) required children to choose their response on a Likert response scale and then choose the equivalent on an additional thermometer scale for their measure (Cystic Fibrosis Questionnaire, CFQ). Such testing could provide evidence that children can understand, interpret, and use a given response scale in the way researchers intend it to be used.

McGrath, Siefert, Speechley, Booth, Stitt, and Gibson (1996) employed a calibration task to check children's ability to use response scales to rate frequencies. They developed a task where children had to rate the size of seven circles which varied in area to each other. This assessed the relationship between children's perceptions of the sizes of circles to the actual size of circles, and provided a check of whether young children could make proportional judgements (McGrath, 1990). Cummins (1997) used a similar approach to test respondents' rating abilities in a QOL scale for individuals with intellectual disabilities. This assessment was done in three stages. First, wooden cubes of unequal sizes had to be arranged in size order. Second, the appropriate blocks had to be transferred to a Likert scale anchored by *largest* to *smallest*. Third, individuals had to indicate the 'correct' response on a Likert scale using items of known importance to them. Such testing could be beneficially incorporated in the development of self-report measures for young children.

Using hypothetical questions in measures could also help researchers to determine whether children can understand and use response scales accurately. A check of children's understanding can be performed with the use of such questions, where researchers can reasonably assume the 'correct' answers (Von Baeyer, personal communication, September 10, 2000). For example, for pain measures the question "How much would it hurt if you: a) opened a birthday present, b) burnt your hand on a hot stove, or c) bumped your toe on the curb?". Researchers could assume from this example that a) would be rated as the least pain, b) the most pain, and c) would be rated somewhere in between the two others.

Using a training period involving practise questions

The use of training periods has been suggested by some researchers (e.g., Collier et al., 2000, Harris et al., 1985, Jirojanakul & Skevington, 2000, Measelle et al., 1998, Stein et al., 1998). Marsh et al. (1991, 1998) make the point that although it was originally thought that measures for young children should be a short as possible (to avoid fatigue or boredom), it may be counter-productive to use very short measures with fewer items. Marsh et al. (1998) argued that the use of short measures could explain why some researchers have failed to develop measures with good psychometric properties. Training children beforehand may impact on the reliability of children's responses to the actual items in measures. For example, Harris et al. (1985) reported that a training period enhanced four year olds understanding of the items and response categories. Their results indicated that when four year olds were trained beforehand, their responses were comparable to those of six year olds (Harris et al., 1985). Harris et al. (1985) argued that four year olds may not lack understanding of items, but may be less able to communicate their answers due to unfamiliarity with the design of self-report measures.

Some researchers have used training periods to help children understand the rating task before completing the main items in their measures (e.g., Collier et al., 1997, 2000, Hughes, 1983, Jirojanakul & Skevington, 2000, Stager & Young, 1982, Stein et al., 1998). For example, Collier et al. (1997, 2000) used a training period with children.

Children were trained using a question about how much they watch television, giving them practise in the presentation style of the items in the measure (questionnaire format using cartoons as visual aids), and also in using the five-point response scale to rate their answers. Jirojanakul and Skevington (2000) also used a training period. Children were trained in the rating task using sets of pictures developed to illustrate different types of items to children – smiley faces to show evaluation; fingers to show intensity, capacity, and importance; and clocks to show frequency (Jirojanakul & Skevington, 2000). The use of a training period at the start of measures may help children understand the rating task, and become familiar with how to use the response scale to answer items (Davis-Kean & Sandler, 2001).

7.4 Implications for the development of the TedQL measure

Reliability and validity of any measure is partly dependent upon the respondent's ability to understand the response scales provided. The use of response scales in child self-report measures assumes the existence of a variety of necessary abilities – including the ability to abstract their abilities, project them onto a subjective continuum and fraction them into degrees. Our review of the literature in this chapter identified the potential difficulties young children can have when using response scales (see p. 190-6). While recognising these problems, researchers have also suggested ways to aid children in answering self-report items using response scales (see p. 197-203). We have considered these recommendations for the development of our TedQL measure in the following sections.

Ensuring children's understanding of anchors

The research reviewed in this chapter showed that children's understanding of the anchors used in response scales can influence their ability to provide accurate self-reports (see p. 197-8). Following the recommendations of Stone and Kennedy-Moore (1992) and Schwartz and Sprangers (1999), we chose specific anchors for the TedQL measure rather than using vague quantifiers such as *occasionally* which may be difficult for children below eight years to understand. The TedQL was designed with children being asked which of two teddy bears were *most like them*, and then whether they were *really like this* or *just a little bit* (see Chapter 4, p. 101). We felt that this wording would avoid the problems that may occur if we used words like *sometimes*, *often* and *nearly always* as verbal anchors in the TedQL measure.

Choosing the most appropriate response scale type

Researchers need a clear basis for their choice of response scale for their child self-report measures. Both Juniper et al. (1997) and Chambers et al. (1999) have shown that the response scale type can impact on children's responses, and that children's responses may not be equivalent across different scale types (see p. 199-201). In addition, respondents may use the response scale to make judgements about what information they should report. We recognised that the type of response scale used could impact on children's responses, and therefore we compared children's responses across different scale types (see Study 2, Chapter 3; & Study 4, Chapters 8 & 9). Study 1 compared children's responses to the TedQL.2 items using linear and circles scales.

We found that the reliability of children's responses was affected by scale type (see Chapter 4, p. 128).

Researchers have suggested creative ways to represent response scales to children to help ease the rating task, and help children understand the differences between each point on a chosen response scale (see p. 198-9). We identified the four most common response scales in our review in Chapter 3 (i.e., Likert, visual analogue, facial expression, and graphic, see p. 46). We provided an empirical basis for selecting the most appropriate response scale for our measure, by comparing the psychometric properties of children's responses across three different response scales in Study 4 (i.e., circles, faces, and thermometer response scales, see Chapter 8, p. 210-11). Study 4 enabled assessment of the relative value of different scale types, by comparing the reliability and reproducibility of children's responses across these three different response scales. The results of Study 4 provided clear evidence for our choice of response scale type for the TedQL.4 measure (see Chapter 9, p. 234-8, & p. 258-61).

Testing children's understanding of the rating task

We also considered whether young children could understand the rating task and appreciate the different graduations used in response scales. Calibration tasks or hypothetical questions have been used to assess children's use of a given response scale (see p. 201-2). We incorporated hypothetical questions in the TedQL measure to provide a test of whether the children understood the task and could use the response options appropriately. We used two hypothetical questions at the beginning of our TedQL measure – one where children were expected to choose the high end of the scale (how much they liked their favourite sweet, see Chapter 4, p. 122, & Chapter 8, p. 221), and one for the low end of the scale (how they would feel if their favourite toy was lost, see Chapter 8, p. 221, & Chapter 10, p. 275).

Training children in using response scales

Young children may also benefit from the use of a training period to help them practise using the response scale to answer items (see p. 202-3). Training children beforehand may help to enhance the accuracy and reliability of children's responses to the actual items in measures. We developed a training period for use with the TedQL measure, which included items such as: "How good are you at singing?"; "How much do you like doing colouring?"; "How good are you at hopping?" (see Chapter 4, p. 104, Chapter 8, p. 221, & Chapter 10, p. 275).

The following chapters (Chapter 8 & 9) report the methodology and results of Study 4, which was designed to further develop our measure by: assessing the relative value of three different response scales for the TedQL.3 measure; investigating the cognitive strategies children use to answer TedQL.3 items; and highlighting items that could be removed from the TedQL.3 measure.

Chapter 8: Further investigation of the TedQL.3 measure – investigating response scales, cognitive strategies used, and items for removal (Study 4).

Introduction and Methodology.

Summary

Aims

Study 4 aimed to: i) establish the most appropriate response scale for the TedQL.3 measure, ii) investigate the strategies children used to answer the TedQL.3 items, and iii) identify items that could be removed from the TedQL.3 measure.

Sample

Two hundred and seventy-seven children (5.0-6.5 years, n=139; 7.0-8.5 years, n=138) completed a QOL measure (TedQL.3) at Time 1. Two hundred and sixty-six children (5.0-6.5 years, n=130; 7.0-8.5 years, n=136) were re-tested at Time 2 (one week later).

Design

Children completed the measure using one of three different response scales – circles, faces, or thermometer. Children were either given the same response scale across both time points or different response scales at Time 1 and Time 2.

Method

Interviewing techniques (a combination of probes and the 'think aloud' methodology) were used to provide information on the strategies children used when answering TedQL.3 items.

8.1 Introduction

Studies 1 and 2 focused on the preliminary validation of a new child QOL self-report measure (TedQL.1, TedQL.2). Study 3 focused on developing the content of this measure (TedQL.2) by providing interview data from children about what was important in their lives. On the basis of the results from Study 3, a new version of the TedQL was developed (TedQL.3).

8.1.1 Establishing the most appropriate response scale for the TedQL.3 measure

As shown in our review of self-report measures in Chapter 3 (see Table 3.2, p. 75-90), some researchers have developed QOL measures for children below eight years (e.g., Collier et al., 2000, Eiser et al., 2000, French et al., 1994, Harter & Pike, 1984, Varni et al., 1998). These measures have employed a number of response scale types with children of different ages (see p. 46). However as discussed in Study 2 (see p. 118) and Chapter 7 (see p. 199-201), research from the pain literature has highlighted that children's ratings may not actually be equivalent across different types of response scales (i.e., their responses may vary depending on the type of scale used, Chambers & Craig, 1998).

Our review in Chapter 3 showed that few authors provide justification for their choice of scale type (only 24 authors out of 105 measures reported evidence for their choice, see p. 47-9). When authors did provide evidence, the most common method was to compare the relative value of different response scales for a given measure (see p. 47). Researchers such as Rebok et al. (2001) have argued that empirical evidence should be used to determine the format for self-report measures, and this would provide guidance on designing measures to maximise the reliability and validity of responses.

These issues have been considered in relation to child self-reported health below eight years (e.g., Chambers & Johnston, 2002, Rebok et al., 2001, Riley et al., 2004). As discussed in Study 2 (see p. 118), Rebok et al. (2001) argued that concrete response scales (such as circles of varying sizes) were better understood and preferred over linear response scales by children below eight years. It may be that such graphic response scales (differing in size or colour) offer children more information on how to grade their answers to items in measures.

However there are some issues and gaps in the literature that warrant further investigation. As discussed in Chapter 7 (see p. 193), Chambers and Johnston's (2002) work revealed developmental differences in children's use of response scales, however there were methodological problems with their work. For example, Chambers and Johnston (2002) used only one type of scale (Likert). Further work is needed to compare children's use of various response scales (e.g., graphic scales such as circles and thermometer scales). Further work is also needed to explore how the reliability, reproducibility and validity of children's responses on a specific self-report measure may be compromised or enhanced by a chosen scale. While Rebok et al. (2001) did compare three different response scales, they repeated items to children with different scales which could have meant the children's responses were affected by practise or order effects. Future work could test children at more than one time point to avoid such biases.

We argued that researchers need a clearer basis and justification for the scale type they choose for their child measures. The first aim of this study was to establish the most appropriate response scale for use in our measure (TedQL.3). This was achieved by comparing the psychometric properties of the TedQL.3 measure across three different scales that have been commonly employed in child self-report measures: a) circles; b) faces; and c) thermometer scales. Our review in Chapter 3 showed that these scale types have been frequently used by researchers for child pain, QOL, mental health and self-concept/esteem self-report measures (see p. 46 & Table 3.2, p. 75-90).

The literature reviewed in Chapters 3 and 4 (see Table 3.2, p. 75-90 & p. 118) showed that circular scales have been used successfully with children under eight years in previous measures (e.g., Harter & Pike, 1984, Rebok et al., 2001). Some researchers have argued this type of scale is more understandable to children (e.g., Harter & Pike, 1984). Harter and Pike (1984) reported internal reliability values above the .70 standard when using their circles scale for children aged four to seven years. Rebok et al. (2001) have provided evidence that their graduated circles scale helped children to rate frequency, intensity, capacity, and importance by providing concrete scales that emphasise the differences between points on a scale. Rebok et al. (2001) also reported that the majority of children preferred a circles response scale over a visual analogue scale. In addition, the results of Study 2 (p. 128) showed that children produced more reliable responses when using a circles (graphic) scale, than children using a linear (visual analogue) scale.

Predicting the most appropriate response scale for the TedQL.3 measure

Based on the arguments above, we predicted that the circles response scale would be the most appropriate scale for the TedQL.3 measure. We predicted that children using the circles scale would produce: 1) fewer 'don't know' answers to TedQL.3 items; 2) a higher amount of responses to items with item-total correlations above .20; 3) responses with higher internal consistency (or internal reliability); and 4) responses with higher reproducibility over time (or test-retest reliability), than children using the faces or thermometer scales. We also predicted that children would prefer the circles scale over the faces and thermometer scales.

Predicting the comparability of children's responses across different response scales

We predicted that children's responses would not be comparable across different responses scales (i.e., that their responses to items using one scale would not be correlated to their responses to the same items using a different scale). This prediction was based on the argument that different scale types are used differently by children, and therefore their responses to the same items will not be equivalent across different scales (e.g., Champion et al., 2000, see Chapter 7, p.199-201).

8.1.2 Investigating the strategies used to answer items in the TedQL.3 measure

Another issue for the development of child self-report measures is whether children can fully understand the concepts behind many items in measures (see Chapter 2, p. 20 & p. 31). Researchers need more information on how children are interpreting and answering self-report items. Dockrell et al. (2000) argued that researchers cannot assume that young children understand items in the same way as adults, and has advocated the use of ad hoc testing for child self-report measures. McColl, Meadows, and Baronsky (2003) have pointed out that the focus of much research has been on establishing the psychometric properties of newly developed measures, such as reliability, validity, and responsiveness to change, with little regard for the cognitive processes that respondents use when answering such self-report measures. Examining these processes can provide researchers with information on how respondents are interpreting and answering items in their measures, and can help identify the cause of problems with items (e.g., comprehension, processing, or communication problems, Collins, 2003). The sequential use of qualitative and quantitative methods can be useful when developing and testing measures (Goering & Streiner, 1996). Therefore we drew on some qualitative methods to investigate the processes children engaged in answering the items in our measure (i.e., what strategies children use to rate their abilities, thoughts and behaviours).

There are various techniques which can be used to test the cognitive processes respondents are engaged in when answering items. These include cognitive task analysis, focus groups, and cognitive interviewing (Jobe, 2003). Different cognitive methods may be more or less appropriate and helpful, depending on the type of measure and the target population (Jobe, 2003).

One technique consists of the use of the "think aloud" method (TAM) and/or follow-up probes (Jobe, 2003). TAM is a cognitive technique for understanding the processes individuals engage in while attempting to answer questionnaire items (Harrison et al., 1996). TAM requires subjects to verbalise their self-talk aloud while performing a pre-determined task (Ericsson & Simon, 1993). Follow-up probes can be used to understand cognitive processes when answering specific questions, for example asking respondents "How did you go about answering that question?" or "How did you remember that?" (Collins, 2003). TAM was designed to assess the ongoing internal dialogue that individuals engage in when performing a given task (Glass & Merluzzi, 1982).

Such interviewing has been used successfully with adults in examining processes involved in a variety of different tasks (Schwartz & Sudman, 1995). Whitney & Budd (1996) argued that TAM can be used to reveal the contents of an individual's working memory and allows the evaluation of respondent's processing strategies when answering items. Harrison et al. (1996) have used this technique to examine the processes behind answering questions in survey research. Their work examined the effect of contextual cues on subjects' answers. Harrison et al. (1996) argued that this method could be applied to research into the cognitive mechanisms underlying self and social report responses to questionnaires and surveys.

This type of interviewing has been applied to the development of self-report measures (Rebok et al., 2001). Children have been shown to produce self-talk that is codeable when used in such research contexts (Fox, Houston, & Pittner 1983, Kendall & Chansky, 1991, Lodge, Harte, & Tripp, 1998, Lodge, Tripp, & Harte, 2000, Prins, 1985, 1986). For example, Lodge et al. (1998) used this technique with nine year old children to examine the relationship between children's positive and negative proportions of self-talk and anxiousness. In applying this method to children's understanding of self-report items, Valla et al. (1994) used TAM to examine children's comprehension of items from a measure assessing the presence of mental disorders. Rebok et al. (2001) used TAM to investigate whether children aged 5-11 years understood some of the complex concepts in their health statue measure (CHIP-CE, such as nervous, healthy, proud, & temper). Rebok et al. (2001) achieved this by coding children's "think aloud" answers for the level of understanding they showed (i.e., poor, some, or clear understanding). They reported that children's understanding of these concepts substantially increased with age. Rebok et al. (2001) identified several key terms in their wording of items that were problematic for the five, six, and seven year old children. For example, 72% of five year old children showed poor understanding of "proud", and 43% showed poor understanding of the words "healthy", "nervous", and "energy".

This study aimed to investigate the strategies that children below eight years used to answer the TedQL.3 items. This aim was achieved through the use of two interviewing techniques - TAM as developed by Ericsson and Simon (1993), and follow-up probes (Willis, Royston, & Bercini, 1991). As young children may be more reluctant to verbalise their thoughts than older children, perhaps due to shyness or inexperience with interview situations, we combined TAM with follow-up probes to provide more opportunity for children to report on the strategies they used to answer items. This methodology provided data from children on the strategies that they used when answering a selection of TedQL.3 items. The strategies children used (coded into five categories from their responses during interviewing) were then compared across different question types.

We hypothesised that children may use different strategies when answering different items. The way certain things/activities are presented and talked about to young children (i.e., in the home and in school) may differ which may in turn influence how children evaluate their levels of ability and functioning. For example, children are often put into teams for races during physical education classes and this could make children more prone to make social comparisons when asked about their physical abilities. Social comparisons (as a way to judge ability) are encouraged in the classroom and are often implicit in the curriculum. As Eccles et al. (1993) argued that school helps children appreciate what they are good at and what they are not good at. We argued that when children are told off at home, parents often refer to stable character attributes in the child when talking about their child's behaviour. This may lead children to make references to stable attributes when asked about their own psychological functioning or behaviour at home.

Following this argument, we predicted that children would use and report different strategies when answering different question types. We predicted that children would be more likely to use social comparisons when answering physical competence items (e.g., "How good are you at running?") compared to other types of items. We also predicted that children would be more likely to refer to stable character attributes when answering psychological functioning items (e.g., "How much of the time do you get scared?", "How much of the time do you get cross or angry?").

8.1.3 Identifying items that could be removed from the TedQL.3 measure

Child self-report measures need to be quick, easy, and straightforward to administer (Eiser & Morse, 2001). Some researchers report reducing measures in length when adapting for a younger age group (Juniper, Guyatt, Feeny, Ferrie, Griffith, & Townsend, 1996, Juniper, Howland, Roberts, Thompson, & King, 1998, Varni et al., 2002). Lengthy measures can take too long to administer, and thus either cause a loss in the child's attention or become too tiring for the child to sit through and answer. Long and Dixon (1996) pointed out that the length of measures is often directly related to the amount of missing data. Landgraf (1996) argued that while establishing the basic psychometric properties of instruments is important, if measures are going to be supported and used by health professionals they need to be easy to administer, practical, and provide information that has clinical relevance. Indeed if a measure is to be adopted, supported, and used by clinicians and other professionals working with children on an everyday basis, it is necessary for that instrument to be straightforward, simple and quick to use. Professionals working in paediatrics have many time constraints, and therefore will prefer instruments that are not too time consuming (Eiser & Morse, 2001). Therefore usefulness, practicality, and feasibility are also important issues that we considered within the development of our QOL measure (the TedQL.3).

Although we discussed how longer measures can aid children in answering items by giving them time to learn how to understand and respond to the items in Chapter 4 (see p. 114-115), on balance we argued here that the TedQL.3 measure would benefit from the deletion of items that were not contributing reliability to the whole measure, or items that were not synonymous with the rest of the items. Therefore, the third aim of this study was to identify items that could be removed from the TedQL.3 measure to make it shorter and quicker to administer. This was achieved by examining the distribution of children's scores on items, the amount of missing data, the item-total correlations, and the conceptual similarity of items to each other to provide information on which items might need removing from the measure. Such analyses would produce data to justify the deletion of specific TedQL.3 items. No predictions were made in relation to this aim as this section was exploratory in nature, and therefore we were unable to predict which items could be deleted prior to conducting this analysis.

8.1.4 Summary of the aims and predictions of Study 4

i. Establishing the most appropriate response scale for the TedQL.3 measure

The first aim of Study 4 was to establish the most appropriate response scale for use with the TedQL.3 measure. We predicted that the circles response scale would be the most appropriate scale for our measure. Therefore children's responses using the circles scale would produce better psychometric properties for the measure, than when they were using the faces or thermometer scales. Additionally, we predicted that children's responses would not be comparable across the three different responses scale types.

ii. Investigating the strategies used to answer items in the TedQL.3 measure

The second aim was to investigate the strategies children used to answer TedQL.3 items. We predicted that children would use and report different strategies when answering different types of items. We predicted that children would be more likely to use social comparisons when answering physical competence items, and more likely to refer to stable character attributes when answering on psychological functioning items.

iii. Identifying items that could be removed from the TedQL.3 measure

The third aim was to identify items that could be removed from the TedQL.3 measure. No predictions were made in relation to this aim.

8.2 Methodology

Sample

Ethics approval was obtained as in Study 1 (p. 100). Three hundred participants aged 6-8 years were identified from three primary schools in Sheffield, South Yorkshire, and two primary schools in Faversham and Bridge, Kent. These children were from two age groups: Year 1 (aged 5.0-6.5 years, n=150), and Year 3 (aged 7.0-8.5 years, n=150). As in Study 1, their parents were given a letter explaining the study. Their parents were asked to complete a permission slip for their child to take part. Two hundred and eighty children (144 females and 136 males) completed the study, as the parents of 20 children did not return the permission slips for their children. The responses of three children (3 males) were excluded from further analyses due to teacher ratings that major changes had occurred in these children's lives between Time 1 and Time 2 (one week later).

Therefore two hundred and seventy-seven children (144 females & 133 males) were included at Time 1. Eleven children were away from school and could not be re-tested at Time 2, therefore two hundred and sixty-six children (138 females and 128 males) were included at Time 2. The children were taken from two age groups, Year 1 (n=139 at Time 1, n=130 at Time 2) and Year 3 (n=138 at Time 1, n=136 at Time 2). The mean age of the Year 1 children was 6.10 years (SD= 0.31 years). The mean age of the Year 3 children was 8.15 years (SD= 0.34 years). Two hundred and fifty-seven (93%; 247 at Time 2) of the children were Caucasian, six were of Afro-Caribbean origin (2%; 6 at Time 2) and 14 were of Asian origin (5%; 13 at Time 2).

Design

The children were randomly assigned to one of 3 response scales at Time 1 and Time 2 (see Table 8.1). Across the two ages half of the children were given either the *same* response scale at Time 1 and Time 2 (e.g., circles response scale at Time 1 and Time 2). The remaining children were given *different* response scales at Time 1 and Time 2 (e.g., circles response scale at Time 1, and faces response scale at Time 2). For those children assigned to different response scales at Time 1 and Time 2, the order of response scales was counterbalanced to avoid order effects (e.g., one half used the faces then circles scales, and the other half used the circles then the faces scales, see Table 8.1).

Table 8.1: Summary of study design (assignment of children to response scale types)

Year 1 (5-6 years), response scale (n)		Year 3 (7-8 years), response scale (n)	
Time 1 (n=139)	**Time 2 (n=130)**	**Time 1 (n=138)**	**Time 2 (n=136)**
Children assigned to the same response scale at Time 1 and Time 2:			
Circles scale (n=26)	Circles scale (n=23)	Circles scale (n=26)	Circles scale (n=26)
Faces scale (n=23)	Faces scale (n=21)	Faces scale (n=26)	Faces scale (n=26)
Thermometer scale (n=22)	Thermometer scale (n=22)	Thermometer scale (n=26)	Thermometer scale (n=26)
Children assigned to different response scales at Time 1 and Time 2:			
Circles scale (n=12)	Faces scale (n=11)	Circles scale (n=13)	Faces scale (n=13)
Faces scale (n=12)	Circles scale (n=12)	Faces scale (n=12)	Circles scale (n=10)
Circles scale (n=12)	Thermometer scale (n=10)	Circles scale (n=10)	Thermometer scale (n=10)
Thermometer scale (n=10)	Circles scale (n=10)	Thermometer scale (n=8)	Circles scale (n=8)
Faces scale (n=14)	Thermometer scale (n=13)	Faces scale (n=8)	Thermometer scale (n=8)
Thermometer scale (n=8)	Faces scale (n=8)	Thermometer scale (n=9)	Faces scale (n=9)

Note. ☆ *= counterbalancing.*

8.2.1 Establishing the most appropriate response scale for the TedQL.3 measure

Measures (child)

TedQL.3 measure

The children were interviewed using two identical teddy bears (40 cm high), which were only differentiated by their name badges, and referred to as either female or male depending on the sex of the child (these can be seen in Figure 3.3, p. 71).

The specific content of items within the TedQL.3 measure was developed from the results of Study 3 as previously reported (p. 161-2). As a result of Study 3, the measure consisted of 30 items, in 5 domains or areas. These are listed in Table 8.2.

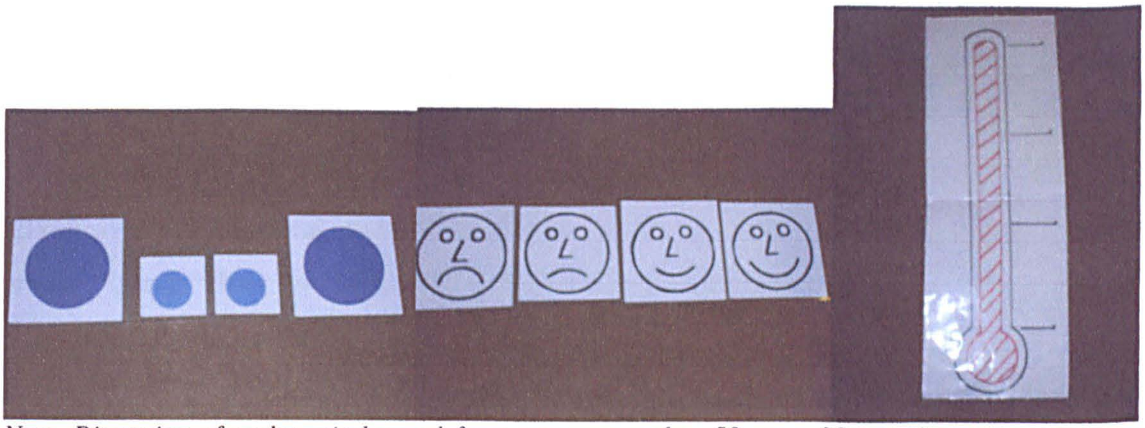
Table 8.2: Domains and items in the TedQL.3 measure

Domain	Items
Physical Competence	PC1: good at swinging PC2: good at playing ball games PC3: good at climbing high things PC4: good at playing computer games PC5: good at running PC6: good at bike riding (or scooter/other)
Peer Acceptance	PA1: having friends to play with PA2: bossing friends (N) PA3: friends coming over to their house PA4: having lots of friends at school PA5: friends bossing them (N)
Family Acceptance	FA1: mum/dad playing with them at home FA2: telling mum/dad what been doing at school FA3: playing with siblings FA4: seeing grandparents FA5: going on trips with mum/dad FA6 : mum/dad telling them off at home (N) FA7: siblings fighting/bossing them around (N)
Psychological Functioning	PF1: type of person they are PF2: getting scared (N) PF3: having bad dreams at night (N) PF4: playing pretend games PF5: getting cross/angry (N) PF6: worrying about losing things (N)
Cognitive Functioning	CF1: good at playing board games CF2: good at mathematics/numbers CF3: good at writing – spell name/other words CF4: good at drawing – what can draw CF5: remembering what people tell them to do CF6: good at reading – read words/look at pictures

Note. (N)= negatively scored item, scores were reversed.

As in Studies 1 and 2, a forced recognition task was used where the bears were first described and then children were to choose which bear was *most like them.* The children were then probed for whether they were *really* like this or *just a little bit.* The children were asked to think about how they had been during the last week when answering the items. Children were assigned to one of three 4-point response scales to answer items. Children used one response scale at Time 1, and the same or a different response scale at Time 2 (one week later). Examples of the three response scales can be seen in Figure 8.1.

Figure 8.1: Photographs of the three response scales (TedQL.3)



Note. Dimension of scales: circles and faces response scales- 50 cm x 20 cm; thermometer response scale- 30cm x 60 cm.

In this version of our measure (TedQL.3) the children were also asked a second question for each item. They were asked about what *they would like to be like* – i.e., their 'ideal' self. The children were trained in answering the second question by being told: "There is also another sort of question that we may answer on. See Iggy is really good at hopping, and goes here. Ziggy is not good at hopping, and goes here. Now I am not good at hopping, so I would point here. But I would really like to be better at hopping. So if I am asked would I like to change how I am at hopping, I would say yes, I would like to change how I am hopping and be much better like Iggy here".

Measures (teacher)

Teachers were asked to judge whether any major changes had occurred in these children's lives between Time 1 and Time 2 (1 week later); which may have altered their perceptions of their lives (hence their QOL) in a dramatic way. The teachers were asked to indicate any children who had experienced significant family or health problems during the period of one week (see Appendix E for example of teacher questionnaire).

Procedure

The interviewer spent a 2-hour long session in the classroom in the preceding days as in Study 1 (p. 103). The children were interviewed individually in a room separate to the main classroom. The same set-up was used as in Studies 1 and 2; where the bears were placed opposite the child and they were asked to take part in a game (see Study 1, p. 103-4 for full protocol).

The study was explained to the children, and they were asked for their verbal assent to the task. Next the children were shown how to use the response scales during a training period. As in Study 2 (p. 122), during this training period children were given one practise item ("How much do you like doing colouring?"), and the children were also given two hypothetical questions ("How much do you like eating your favourite sweet?", and "How would you feel if your favourite toy was lost?"). All the children responded accurately to these practise items.

During the training period the children were also taught how to use a 'don't know' option (a blue question mark at the end of the response scale, see Figure 8.2), if they didn't know the answer or didn't understand the question.

Figure 8.2: Photograph of the 'don't know' option on the response scale (TedQL.3)

This was to help children learn how to respond to questions they did not understand, to help minimise bias (i.e., to avoid children answering questions simply to please the interviewer). The children were given a nonsense question, and shown that they were to point to the 'don't know' option if there was a question that was silly or they didn't understand. The children were told: "Now there may be some questions that the teddies answer on, but we know they are being silly and their answers don't make sense. Or they may answer questions that we don't understand, and we can then use the 'don't know' answer here (point to 'question mark' sign) to show we don't know the answer. See this teddy bear has looked outside today and decided that he/she thinks the sky is orange, and he/she sits here on this point. This teddy bear has looked outside and he/she thinks the sky is purple, and he/she sits here on this point. But I don't know the answer to this question, as it is silly, so I am going to point to the 'don't know' sign over here. See? The sky can't be orange or purple, what colour do you think it is?".

The children were then given the TedQL.3 items. During completion of the TedQL.3 items at Time 1, children were interviewed using TAM and follow-up probes. Ten items were chosen at random from the 30 TedQL.3 items to provide more detailed information on the processes children engaged in while answering items. Table 8.3 details the items included in this selection.

Table 8.3: Table of the items included in the 'think aloud' selection

Domain	Items
Physical Competence	PC5 – Good at running PC6 – Good at bike riding
Peer Acceptance	PA4 - Has lots of friends at school PA5 – Friends boss them (N)
Family Acceptance	FA6 – Mum/dad tells them at home (N) FA7 – Siblings fight/boss them around (N)
Psychological Functioning	PF5 – Gets cross/angry (N) PF6 – Worries about losing their things (N)
Cognitive Functioning	CF5 – Remembers what people tell them to do CF6 – Good at reading – read words/look at pictures

Note. (N) = negatively scored item, scores were reversed.

The children were asked to 'think aloud' on the above selection of ten items from the TedQL.3 measure at Time 1, while they were considering their responses to these items. They were asked to report aloud everything they were thinking and saying to themselves. A practise question was used to train children in using this procedure. The children were told: "We are going to add something else to this game. When you are answering the questions that Iggy and Ziggy answer, I would like you to tell me all the things that come into your head when you are choosing your answer. See I am answering the question about whether I am good at tidying my bedroom… Now what do I think? Well I don't like to tidy my bedroom, but I do tidy it when my mum tells me to… and I make sure that all my things are put away and my clothes are in the wardrobe… so yes I think I am good at tidying my bedroom, I am going to point here. Now we are going to answer some more questions like that one, and I want you to remember to talk aloud to me, and tell me what you are thinking as you answer the questions."

When answering the selected items if a child was silent for more than 10 seconds they were given a maximum of two prompts: "Remember to say out loud all the things that come into your head", and "What are you thinking and saying to yourself right now?".

The children's responses were recorded verbatim on the response sheet (see Appendix E for example of response sheet used).

As in Studies 1 and 2 the two teddy bears were counterbalanced for whether they represented a positive or negative statement (p. 104). As in Studies 1 and 2 (p. 104), the children were encouraged to restate their choice following selection (e.g., "I am like Ziggy, I have lots of friends to play with"). Children chose a sticker at the end of the session to acknowledge their participation. The children were re-tested after a 1-week time period in the same setting and with the same conditions, and again received a sticker.

Scoring

Children's responses to the items on the TedQL.3 measure were recorded as numerical scores on a response sheet by the interviewer (0=*really bad/not at all good/dislike*; 1=*just a little bit bad/not good/dislike*; 2=*just a little bit good/like*, 3=*really good/a lot good/always good/like*). Appendix E gives an example of the response sheet used. Negatively scored items were reversed so that higher scores represented higher levels of QOL.

The children's responses were entered into SPSS to give scores for their 'actual' QOL (i.e., *how they actually are*) and 'ideal' QOL (i.e., *how they would like to be*). The actual and ideal QOL scores ranged from 0 to +3. These scores were then used to compute discrepancy scores for each item in the measure. The scores were calculated in the same way as previous researchers have done in their discrepancy-based measures (i.e., ideal QOL minus actual QOL scores, Collier et al., 2000, Eiser et al., 2000). These discrepancy scores ranged from -3 to +3. The actual QOL scores were reported in the results as these scores were equivalent with how other researchers have calculated QOL scores for their child measures (e.g., Rebok et al., 2001, Varni et al., 2002). The discrepancy scores were also reported as these reflect the children's individual preferences for their levels of ability and functioning.

Treatment of data and analyses

The distributional properties of the children's actual QOL and discrepancy TedQL.3 scores were examined. Assessment of skew and kurtosis were made using the same criteria as Study 1 (see p. 106-7). Normality testing was also carried out using Kolmogorov-Smirnov tests (see p. 106-7). Where data appeared to be significantly skewed, curved, and/or different from a normal distribution, non-parametric statistics and tests were used throughout the analysis (as with Study 1, p. 107). Analysis was conducted to assess whether children's actual QOL and discrepancy TedQL.3 scores differed systematically by two independent variables (age and gender).

The following analyses were conducted to address the specific hypotheses made:

1) Predicting the most appropriate response scale for the TedQL.3 measure

a. The percentage of 'don't know' answers children gave to items was calculated, and compared across the three response scales.

b. The percentage of items with item-total correlations over a standard of .20 was calculated, and compared across the three response scales. Every item in a measure should contribute to the overall score produced by a measure (Chase, 1978). Calculating item-total correlations provide a check of whether items were functioning as planned. Items can be judged as 'good' items if the item-total correlation coefficients were above a standard of .20 (Streiner & Norman, 1995).

c. The internal reliability of the children's responses was assessed using Cronbach's alpha statistics (α), and compared across the three response scales. (see Study 1 for explanation of Cronbach's alpha statistics, p. 107).

d. The reproducibility of children's responses was assessed using Intra-class correlation coefficients (ICC; $\rho\pm$), and compared across the three response scales. Approximately half the children were given the same response scale at Time 1 and Time 2. ICC's were used to assess whether the mean scores were correlated over the two time points. ICC's are a measure of the relationship between the between-subject and the total variance (Landis & Kock, 1977), and can provide an estimate of the reproducibility of responses over time on a given measure. The higher the correlations between the scores at one time point and the scores at a second time point, the greater the ICC value. ICC estimates must reach above the recommended criterion standard of .60 to provide evidence for reproducibility over time (Juniper et al., 1997).

e. The children's individual preferences for response scales to answer TedQL.3 items were gained, by asking children which response scale they preferred to use. The percentages of children preferring each response scale were calculated.

2) Predicting the comparability of children's responses across different response scales
The comparability of children's responses on the TedQL.3 items across different response scales was also assessed. Approximately half the children were given different response scales at Time 1 and Time 2. Wilcoxon significance tests were used to assess whether the children's scores on the TedQL.3 differed significantly over the different response scales. Spearman's rank order correlation coefficients (ρ) were used to assess whether their scores were correlated over the different response scales.

8.2.2 Investigating the strategies used to answer items in the TedQL.3 measure

Measures (child)

Selection of ten TedQL.3 items

Ten items were chosen from the 30 TedQL.3 items to use to interview children using TAM and follow-up probes (refer to procedure section 8.2.1 for details of these items, see p. 222-3).

Pilot work (n=10 children) with this selection of the TedQL.3 items using the TAM identified three main strategies used by children when answering items:

a. the use of social comparisons, i.e., reference to, and comparison with, friends, family or other people (e.g., "I am good at running because I always beat my friends in a race")

b. reference to stable character attributes within themselves or others (e.g., "I do not get told off at home because I am always well behaved")

c. reference to specific instance or concrete example of a situation or behaviour (e.g., "I worry about losing my things, as I lost my favourite teddy yesterday at school and I can't find it now").

The results of the pilot work were used to develop the coding framework used to analyse the think aloud data from Study 4.

Procedure

(Refer to section 8.2.1 for procedure details, see p. 222-3)

Treatment of data and analyses

The children's responses to the ten TedQL.3 items (using TAM and follow-up probes) were recorded verbatim on the response sheets, and then typed up into word documents. To preserve the confidentiality of the children all names mentioned by the children were changed or omitted. These verbatim interviews served as raw data for the interview analysis.

The think aloud data was analysed using a content analysis technique. The coding framework developed from pilot work was used (see p. 222-3) providing three categories for coding children's answers. An additional category was added to include children's answers that showed some understanding of the item, but contained information too varied to fit into the three original categories. A further category was added in order to allow for children who did not offer any reason or refer to any strategy within their responses to items (i.e., children answered 'don't know' or repeated response anchors with no further elaboration). This coding framework allowed comparison of strategies used by children across different question types.

A paper copy of the children's answers was read through a number of times until it was familiar to the author. After this (using a word document version) each answer was 'copied and pasted' into word tables. (See Appendix E for example of answers pasted into coding table). These tables were used to produce frequency data to provide information on the most common answers given by children for each of the 10 items. The data was analysed by age group (Year 1 and Year 3) as well as across both age groups.

8.2.3 Identifying items that could be removed from the TedQL.3 measure

Measures

The children's answers to all the TedQL.3 items at Time 1 were used from the data collected above in relation to the first aim (refer to section 8.2.1 for details of the measure and items, see p. 218-20)

Procedure

(Refer to section 8.2.1 for procedure details, see p. 220-1)

Treatment of data and analyses

The children's actual QOL and discrepancy TedQL.3 scores were used in this analysis, separately for both age groups and combined overall.

The following analyses were conducted on the complete data set:

a. the distribution of children's scores on all the items were examined to identify items that were too similar to each other or redundant;

b. the percentage of 'don't know' answers children gave to items was calculated for all of the individual items to show items that children consistently misunderstood or were unable to answer;

c. the percentage of items with item-total correlations over a standard of 0.20 was calculated for each individual item in the measure. Every item in a measure should contribute to the overall score produced by a measure (Chase, 1978). This enabled a check of whether there were any items that were not contributing reliability to the measure as whole;

d. the conceptual similarity of individual items within each domain was considered. This enabled decisions to be made on items that did not 'fit' well with the other items in any given domain.

The results of this analysis provided information on items that could be removed from the measure. In Chapter 9, the results from both the quantitative and 'think aloud' analyses will be reported and discussed.

Chapter 9: Further investigation of the TedQL.3 measure – investigating response scales, cognitive strategies used, and items for removal (Study 4).

Results and Discussion.

Summary

Results

Children's responses using the thermometer response scale were associated with the 'best' psychometric properties for the TedQL.3 measure for both age groups (i.e., internal consistency and reproducibility). Children's responses to the TedQL.3 items were not related across the circles and thermometer scales, or across the faces and thermometer scales. Children's responses to TedQL.3 items were correlated to each other when children used the circles and faces scales.

The majority of the children gave concrete examples of specific situations or instances that happened to them as reasons for their response choices during interviewing regardless of item type.

Examining the distribution of children's scores on individual items, calculating the percentage of 'don't know' answers to individual items, examining item-total correlations for individual items, and considering the conceptual similarity of individual items within each domain, identified eight items that could be removed from the TedQL.3 measure.

Implications

Response scale can impact on the psychometric properties of self-report measures, and ratings may not be comparable across different response scale types.

The interview data may explain why younger children's self-reports are often less stable over time compared to older children's self-reports. This result may not mean that young children's self-reports are unreliable over time, but that their answers fluctuate more as they use different examples when answering items.

As a result of Study 4, a new 22-item version of the measure was produced (the TedQL.4), using a thermometer response scale, and presenting items to children using four teddy bears (instead of two).

9.1 Results

9.1.1 Data screening analyses

Item descriptives: means, range and assessment of skew/kurtosis

Children's mean scores on the TedQL.3 measure were significantly skewed towards higher actual QOL and lower discrepancy reports. This pattern of skew was consistent across both age groups (Year 1 and Year 3) and both time points (Time 1 and Time 2), with the exception of Year 1 actual QOL scores (Time 1) and Year 3 discrepancy scores (Time 2) (see Tables 9.1 and 9.2). Year 3 children's scores showed high levels of kurtosis across both age groups and both time points (see Tables 9.1 and 9.2). Kolmogorov-Smirnov tests showed that both Year 1 and Year 3 children's discrepancy scores were significantly different from normal at both time points (Year 1: D= 0.11 & 0.08, p<.05, Year 3: D= 0.08 & 0.10, p<.05, see Appendix E for normal Q-Q plots). As in previous studies non-parametric tests were used for analyses because of these distributions (see Appendix E for full details of skew and kurtosis calculations).

Table 9.1: Descriptives for children's scores on the TedQL.3 measure (Time 1)

	Time 1					
Age group	n	Mean (SD)	Range	Skew	Kurtosis	Normality test (D)
Year 1:						
Actual QOL	139	2.09 (0.35)	1.14-2.90	-0.17	0.03	0.05
Discrepancy	139	0.49 (0.30)	0.07-1.33	0.61*	-0.12	0.11*
Year 3:						
Actual QOL	138	1.99 (0.30)	0.87-2.71	-0.42**	1.04**	0.05
Discrepancy	138	0.53 (0.29)	0.03-1.48	0.81**	0.96**	0.08*

*Note. Significance level of skew, kurtosis, and normality: ** p< .01; * p< .05*

Table 9.2: Descriptives for children's scores on the TedQL.3 measure (Time 2)

	Time 2					
Age group	n	Mean (SD)	Range	Skew	Kurtosis	Normality test (D)
Year 1:						
Actual QOL	130	2.07 (0.39)	1.08-2.90	-0.41*	-0.40	0.07
Discrepancy	130	0.52 (0.31)	0.07-1.32	0.34*	-0.69**	0.08*
Year 3:						
Actual QOL	136	1.96 (0.32)	1.00-2.64	-0.27*	0.10	0.06
Discrepancy	136	0.54 (0.28)	0.00-1.40	0.72	0.26	0.10*

*Note. Significance level of skew, kurtosis, and normality: ** p< .01; * p< .05*

Potential item bias: age

Spearman's correlation coefficients revealed that age was not correlated with the children's actual QOL and discrepancy scores in either age group, at either time point (see Table 9.3).

Table 9.3: Relationship between chronological age and mean scores on the TedQL.3 for Year 1 and Year 3 children

	Correlation coefficient (ρ)			
	Year 1		Year 3	
Time period	n	Age	n	Age
Time 1:				
Actual QOL	139	-0.43	138	-0.11
Discrepancy	139	0.21	138	0.01
Time 2:				
Actual QOL	130	0.02	136	-0.03
Discrepancy	130	-0.03	136	-0.08

Potential item bias: gender

Mann-Whitney U-tests revealed that there was no effect of gender on Year 1 children's actual QOL and discrepancy scores at either time point (see Table 9.4).

Table 9.4: Effect of gender on mean scores on the TedQL.3 for Year 1 children

	Male		Female	
Time period	**n**	**Median (Mean)**	**n**	**Median (Mean)**
Time 1:				
Actual QOL	69	2.15 (2.10)	70	2.08 (2.08)
Discrepancy	69	0.43 (0.46)	70	0.47 (0.53)
Time 2:				
Actual QOL	64	2.07 (2.05)	66	2.12 (2.08)
Discrepancy	64	0.47 (0.49)	66	0.52 (0.55)

Note. (Means are reported in brackets for comparison)

Mann-Whitney U-tests showed that Year 3 children's actual QOL and discrepancy scores differed by gender at both time points (see Table 9.5). Boys reported higher actual QOL scores, and lower discrepancy scores than girls.

Table 9.5: Effect of gender on mean scores on the TedQL.3 for Year 3 children

	Male		Female	
Time period	**n**	**Median (Mean)**	**n**	**Median (Mean)**
Time 1:				
Actual QOL	64	2.04 (2.07)	74	1.92 ** (1.92)
Discrepancy	64	0.45 (0.46)	74	0.60 ** (0.60)
Time 2:				
Actual QOL	64	2.03 (2.04)	72	1.91 ** (1.88)
Discrepancy	64	0.44 (0.47)	72	0.58 ** (0.61)

*Note. **p<.01 (Means are reported in brackets for comparison)*

9.1.2 Establishing the most appropriate response scale for the TedQL.3 measure

Percentage of 'don't know' answers to TedQL.3 items

The percentage of children using the 'don't know' option to answer the TedQL.3 items was calculated. This was to assess the amount of missing data, and compare these values across the three response scale types. The percentage of 'don't know' answers was low across both age groups, both time points, and all three response scales (see Table 9.6). Year 1 children produced the lowest percentage of 'don't know' responses at Time 1 when using the faces response scale (2.11%) and at Time 2 using the thermometer response scale (2.15%). Year 3 children produced the lowest percentage of 'don't know' responses when using the faces response scale at both time points (Time 1: 1.67%, Time 2: 2.08%).

Table 9.6: Percentage of 'don't know' answers across three response scales

	Percentage of 'don't know' answers (%)			
	Year 1		**Year 3**	
Response scale	**Time 1**	**Time 2**	**Time 1**	**Time 2**
Circles	2.27	2.67	1.77	2.58
Faces	2.11	2.46	1.67	2.08
Thermometer	2.25	2.15	2.56	3.18

Item-total correlations (homogeneity of the TedQL.3)

The correlation of each item in the TedQL.3 measure to the total score was calculated to assess whether items were contributing reliability to the measure (see Appendix E for table of all item-total correlations). At both age groups and both time points 70% or more of the items were above the .20 standard across all three response scales (see Table 9.7).

Table 9.7: Percentage of item-total correlations falling above .20 standard across three response scales

	Percentage of items falling above 0.20 standard (%)			
	Year 1		Year 3	
Response scale	Time 1	Time 2	Time 1	Time 2
Circles	82	78	85	72
Faces	83	73	77	75
Thermometer	72	78	70	77

Internal reliability (internal consistency)

The internal consistency of children's responses was calculated using Cronbach's alpha statistics, and compared across the three response scales. Year 1 children's actual QOL ratings when using the faces and circles scales were over the .70 standard for consistency levels, however their discrepancy ratings were over .70 when using the faces and thermometer scales (see Table 9.8). Year 3 children's actual QOL ratings were over .70 for internal consistency when using the circles scale, however their discrepancy ratings were over the .70 standard when using the thermometer scale (see Table 9.8).

Table 9.8: Internal consistency (α) of children's responses on the TedQL.3 measure across three response scales (for all 30 items)

	Response scale								
	Circles			Faces			Thermometer		
Age group	n	No. of items	α	n	No. of items	α	n	No. of items	α
Year 1:									
Actual QOL	50	30	0.72	49	30	0.73	40	30	0.66
Discrepancy	50	30	0.62	49	30	0.75	40	30	0.73
Year 3:									
Actual QOL	49	30	0.74	46	30	0.56	43	30	0.53
Discrepancy	49	30	0.57	46	30	0.53	43	30	0.76

Reproducibility (median testing and test-retest reliability)

A sub-group of the children were given the *same* response scale at both time points (see section 8.2.1, p. 205-6) i.e., children answered items using the circles response scale at Time 1 and Time 2, or the faces or the thermometer scale at both time points.

Wilcoxon tests revealed that both the Year 1 and Year 3 children's actual QOL and discrepancy scores did not differ over time when using the circles or the faces response scales (see Table 9.9). The same result was found for Year 1 children's actual QOL and discrepancy scores using the thermometer response scale (see Table 9.9). Year 3 children's actual QOL scores did not differ over time when children were using the thermometer scale (see Table 9.9). However, Year 3 children's discrepancy scores did differ over time when using the thermometer scale, with these children reporting higher discrepancies at Time 1 compared with Time 2 (see Table 9.9).

Table 9.9: Difference between children's scores on the TedQL.3 using the same response scale over two time points

	Year 1			**Year 3**		
		Median (Mean)			**Median (Mean)**	
	n	Time 1	**Time 2**	**n**	**Time 1**	**Time 2**
Circles response scale						
Actual QOL	23	2.17 (2.14)	2.07 (1.99)	26	2.02 (1.99)	1.97 (1.97)
Discrepancy	23	0.43 (0.42)	0.50 (0.51)	26	0.58 (0.59)	0.73 (0.74)
Faces response scale						
Actual QOL	21	2.20 (2.15)	2.27 (2.19)	26	1.95 (2.02)	1.98 (1.97)
Discrepancy	21	0.47 (0.54)	0.33 (0.48)	26	0.50 (0.47)	0.48 (0.52)
Thermometer response scale						
Actual QOL	22	2.08 (2.11)	2.05 (2.07)	26	2.03 (1.98)	2.07 (1.99)
Discrepancy	22	0.55 (0.57)	0.65 (0.65)	26	0.61 (0.60)	0.41 ** (0.47)

Note. ** *p< .01 (Means are reported in brackets for comparison)*

In addition the reproducibility of the children's responses was compared across response scales using the Intra-class correlation coefficients (ICC, ρ±) of their actual QOL and discrepancy scores between Time 1 and Time 2. These figures give a measure of the test-retest reliability of a measure, i.e., the reproducibility of responses over time. Year 1 children's responses to TedQL.3 items using the thermometer response scale showed the highest level of reproducibility (see Table 9.10). The reproducibility of Year 3 children's responses were more mixed than Year 1 children's responses across the scale types. Year 3 children's responses using the faces response scale about their actual QOL showed the highest reproducibility (ρ± = .77). Their discrepancy scores using the thermometer response scale showed the highest reproducibility (ρ± = .62). The reproducibility of the children's responses for Year 1 children was above the .60 standard when children used the faces and thermometer response scale (for actual QOL and discrepancy scores, see Table 9.10). The reproducibility of Year 3 children's responses were above .60 across all three scales when rating their actual QOL, and above .60 for their discrepancy scores when using the thermometer scale (see Table 9.10).

Table 9.10: Test-retest reliability (ρ±) of children's responses using the same response scale across two time points

	Response scale					
	Circles		**Faces**		**Thermometer**	
Age group	**n**	**ρ±**	**n**	**ρ±**	**n**	**ρ±**
Year 1:						
Actual QOL	50	0.57***	49	0.69***	40	0.77***
Discrepancy	50	0.39*	49	0.68**	40	0.78***
Year 3:						
Actual QOL	49	0.61***	46	0.77***	43	0.61***
Discrepancy	49	0.55***	46	0.50***	43	0.62***

Note. *** *p< .001;* ** *p<.01;* * *p<.05*

Individual preferences for response scale type

A sub-group of the children were given *different* response scales at Time 1 and Time 2 to answer items from the TedQL.3 measure (see section 8.2.1, p. 215-6). These children asked which scale they preferred to use to answer items and why. Year 1 children preferred the thermometer scale over the circles and faces scales. Year 3 children expressed an almost equal amount of liking for both the faces and the thermometer scale. Table 9.11 details these preferences.

Table 9.11: Children's individual preferences for response scales

	Percentage preferences: % (n/total n)	
Response scale	**Year 1**	**Year 3**
Circles	44 (19/43)	24 (10/41)
Faces	36 (16/44)	50 (20/40)
Thermometer	63 (26/41)	48 (17/35)

The children's spontaneous answers across both age groups provided reasons for these preferences. Some examples are given below.

"Thermometer easier, it was quite easy you know if you were doing bad, at bottom, understand this better coz can see high and low"

"Thermometer, coz circles were a bit confusing, coz thermometer just lines up, on circles keep changing them, get confused which end is which"

"Faces, coz you can see that if not good, got sad face and medium and that, on thermometer can't see really which type you got"

"Faces, coz had patterns of smiley and sad on them, circles are just big and little"

"Circles, coz two smaller circles and two bigger and faces all the same size"

"Circles, coz happy little and big, and sad little and big on faces look almost the same, but circles don't"

"Circles, coz faces if something really bad happened, it might be on the smiley face, when it meant to be on sad face"

Comparability of children's responses across different response scales

The responses of the children completing the TedQL.3 measure using two different response scales (over the two time points) were analysed to assess whether children's responses were affected by response scale type, i.e., if children gave comparable answers to items when using one type of response scale compared with their answers to the same items with another type of response scale.

Wilcoxon tests revealed no differences between both Year 1 and Year 3 children's actual QOL and discrepancy scores on the TedQL.3 items using one response scale compared with another (e.g., using circles at Time 1 and faces at Time 2, see Table 9.12).

Table 9.12: Difference between children's scores on the TedQL.3 using different response scales

	Year 1			Year 3		
		Median (Mean)			Median (Mean)	
	n	Time 1	Time 2	n	Time 1	Time 2
Circles (T1) and faces (T2) response scales						
Actual QOL	23	1.93 (1.97)	2.07 (1.98)	23	2.00 (1.91)	2.00 (1.86)
Discrepancy	23	0.47 (0.51)	0.50 (0.54)	23	0.57 (0.50)	0.50 (0.51)
Faces (T1) and thermometer (T2) response scales						
Actual QOL	21	2.18 (2.04)	2.21 (2.07)	17	1.97 (2.05)	1.93 (2.01)
Discrepancy	21	0.40 (0.48)	0.37 (0.51)	17	0.37 (0.54)	0.43 (0.49)
Circles (T1) and thermometer (T2) response scales						
Actual QOL	20	2.00 (2.18)	2.03 (2.06)	18	1.88 (1.99)	1.88 (1.94)
Discrepancy	20	0.43 (0.43)	0.47 (0.45)	18	0.54 (0.39)	0.47 (0.41)

Note. (Means are reported in brackets for comparison)

In addition the correlations between children's actual QOL and discrepancy mean scores using one type of response scale were compared to their mean scores when using a different response scale. Across both age groups, the children's responses to items using the circles and faces scales were correlated to each other (see Table 9.13). The children's responses to items using the faces and thermometer scales were not correlated to each other, and this result was consistent across both age groups (see Table 9.13). The data for children using the circles and thermometer scales to answer the same items were more mixed. Year 1 children's responses using these scales were correlated to each other across these two scales, however Year 3 children's responses were not correlated when using these two scales (see Table 9.13).

Table 9.13: Relationship between children's scores on the TedQL.3 when using different response scales to answer items

	Correlation coefficient (ρ)					
Age group	n	Circles and faces scales	n	Faces and thermometer scales	n	Thermometer and circles scales
Year 1:						
Actual QOL	23	0.95 ***	21	0.40	20	0.57***
Discrepancy	23	0.95***	21	0.41	20	0.55**
Year 3:						
Actual QOL	23	0.79***	17	0.39	18	0.19
Discrepancy	23	0.56*	17	0.40	18	0.39

Note. *** p<.001; ** p<.01; * p<.05

9.1.3 Investigating the strategies used to answer items in the TedQL.3 measure

The answers children gave to a sample of the items were compared when interviewed using a combination of probes and the think aloud methodology.

The reasons they gave were coded into one of five categories for each item (see Table 9.14).

Table 9.14: Coding categories for answers to TedQL.3 items

Coding category	Explanation of category
Social comparison used	Child included reference to, and comparison with, friends, family or other people
Stable character reference used	Child included reference to what they are like as a person, e.g., 'am good at reading as got good memory'
Concrete example given	Child included reference to a specific instance or situation, e.g., 'do worry about losing things as lost something before'
Other reasons	Child gave answer that showed some understanding of the item, but did not fit into the categories above
No further reason given	Child answered 'don't' know' or repeated the response anchors with no further elaboration, i.e., 'really good at it' or 'not very good at it'

Physical Competence items

The children were asked whether they were good at running and bike riding or not. The reasons they gave for their answers using the 'think aloud' method were coded into four categories as shown in Table 9.15.

Table 9.15: Percentages of children's answers to physical competence items

	% (n) of children	
Coding category	Running	Bike riding
Social comparison used	38 (105)	7 (20)
Stable character reference used	6 (16)	12 (33)
Concrete example given	13 (36)	42 (116)
Other reasons	13 (36)	27 (75)
No further reason given	30 (84)	12 (33)

a. Running ability

As shown in Table 9.15, two-fifths of all the children used social comparisons in their answers to the item on running ability (Year 1: 32%, Year 3: 46%). Of the children who used social comparisons, 91% made these in reference to their friends or peers, and 9% were made to siblings or other family members, e.g.;

"Very good, coz my friends race and I keep winning them" (girl; 7.6 years)
"Little bit good, coz my friends are really fast and make me think am not as good as them" (boy; 8.1 years)
"Faster than my friends, I'm fastest, except for Jack, he's really fast, Robert's the same speed as me" (boy; 6.0 years)
"Very good, coz quite a lot of races with year 6, with boys too, and I won them" (girl; 8.5 years)
"Very good, coz can run faster than year 2's, even though smaller than friends can still run faster" (boy; 6.2 years)
"Really good, coz faster than my sisters, the older and younger ones" (girl; 7.5 years)
"Not so good, coz... my sister she can run and she beat me" (girl; 6.6 years)

13% of the children made reference to concrete examples in their answers (Year 1: 13%, Year 3: 13%):

"Not very good, not run fast, one time I broken my toe, stubbed it really badly" (girl; 7.9 years)
"Very good coz today I did exercises, very good today" (girl; 6.0 years)

6% of the children made reference to stable character attributes in their answers, e.g.;

"Quite good, coz got long legs and am tall, coz mum's tall too" (girl; 8.3 years)
"Very good coz my legs are long" (girl; 5.9 years)

However more Year 3 children mentioned stable character references (34%), than Year 1 children (7%).

13% of children gave answers that showed some understanding of the item but their responses were too varied to be coded into any given category, e.g.;

"Very good coz keep running, right fast, speedily" *(girl; 8.2 years)*
"Good coz you play tiggy and that a lot" (boy; 7.9 years)
"Really not good, coz haven't learnt yet" (girl; 6.3 years)

However more Year 3 children gave different answers that could not be coded into a category (23%), than Year 1 children (3%).

Nearly one-third children gave no reasons for their response choice on the item (i.e., either said 'don't know' or repeated the response anchors, see Table 9.15). However more Year 1 children gave no further reason (45%), than Year 3 children (14%).

b. Bike riding ability

As shown by Table 9.15, two-fifths of the children used concrete examples in their answers (Year 1: 55%, Year 3: 29%). Of these children 40% referred to the fact that they did or did not use stabilisers when riding their bike, 55% mentioned that they practised a lot on their bikes, and 5% reported that they could do tricks on their bikes. Some examples of answers are given below.

"Really good, coz I don't need stabilisers, just on two wheels, good up and down hills" (boy; 7.8 years)
"Good coz can ride without trainer wheels" (boy; 6.6 years)
"In the middle, not been on bike for a long time, forgot what to do on it" (boy; 8.3 years)
"Very good, coz have massive garden and I get lots of practise and play on bike rides on own" (girl; 8.3 years)
"I'm real good, practising for long, long time, go proper balance and stand up" (boy; 6.6 years)
"Very good, can do bunny hops and stuff" (boy; 7.8 years)
"Real good, I can do wheelies and do ramps, bunny hop and that" (boy; 5.5 years)

12% of children referred to stable character attributes in their answers, e.g.;

"Very good, coz my family is a very sport family, don't have a lot of sweets, coz mum like in olden day, so eat good food and not get fat" (boy; 8.6 years)

However more Year 3 children (21%) mentioned stable attributes than Year 1 children (3%).

7% of all the children made social comparisons when answering whether they were good at bike riding or not, and these were related either to their friends or to their siblings (Year 1: 6%, Year 3: 8%):

"Really good, coz we, my brother and me, always on our bikes, we have races and I win sometimes, my brother said I'm really good" (girl; 8.5 years)
"Don't have any stabilisers, my brother falls off without them, I don't" (boy; 6.2 years)

One quarter of the children gave answers that showed some understanding of the item but their reasons did not fit into any category (Year 1: 24%, Year 3: 31%), e.g.;

"Not very good, coz keep wobbling" (girl; 7.9 years)
"Not very good, can't get balance on bike and scooter" (boy; 8.3 years)
"Little bit good, sometimes have, when big bike, roads thin and is hard to turn, so have to put foot down" (girl; 8.2 years)
"Good at riding, my bike is a Tom and Jerry one" (boy; 6.3 years)
"Not good, the bike angles are loose" (boy; 5.7 years)

"Good, even though it's up high, my seat" *(boy; 6.8 years)*

12% of children provided no reasons for their response choice on the item (Year 1: 12%, Year 3: 11%).

Cognitive Functioning items

The children were asked whether they were good at reading and remembering things or not, and the reasons they gave were coded into the four categories shown in Table 9.16.

Table 9.16: Percentages of children's answers to cognitive functioning items

	% of children	
	Reading	**Remembering**
Coding category		
Social comparison used	14 (39)	0 (0)
Stable character reference used	32 (89)	3 (8)
Concrete example given	14 (39)	70 (194)
Other reasons	7 (19)	15 (42)
No further reason given	33 (91)	12 (33)

a. Reading ability

As shown in Table 9.16, one third of the children made reference to stable character attributes when asked whether they were good at reading, e.g., they can read on their own, they remember a lot of words (Year 1: 30%, Year 3: 34%). Some examples are given below.

"Difficult coz I struggle reading" (girl; 7.8 years)
"Not so good, coz don't have a lot of impression and don't spell words out well" (girl; 8.1 years)
"Very good, coz can say words, can think then say them, and get it wrong, then say again and get it right, am reading a book at home by myself" (boy; 6.1 years)
"Not very good at reading, just get the words mixed, sometimes don't get lot of words right" (boy; 6.6 years)

14% of all the children used social comparisons in their answers (Year 1: 12%, Year 3: 17%). These were related to what level reading book they were on compared to their friends and peers, e.g.;

"Abit, coz some friends on brown, higher than me, I'm only on white" (boy; 7.8 years)
"Good at reading, coz when friends don't know a word, I say what word is and it's right" (girl; 8.2 years)
"Little bit good at reading, coz when I hear my friends' reading, coz sit next to them, it makes me feel am not so good as hear them" (boy; 8.1 years)

"I'm on last two books, after one I'm on then I'll be on green, one of high levels" (boy; 5.8 years)
"Real good, I'm top of the class, coz I'm the best reader, I learn" (boy; 6.5 years)

14% of all the children mentioned a concrete example in their answer (Year 1: 15%, Year 3: 12%). Of these children using specific instances 81% mentioned whether they practised reading or not, and 19% mentioned the amount and difficulty of books they read, for example:

"Used to be not that good at reading, but now I am coz I try to read everything" (girl, 8.5 years)
"Quite good, coz get to stay up late and read so practise a lot" (girl, 8.3 years).
"Very good, coz I always practise at home, when it's the weekend and I read to both of my brother's, they are younger" (boy; 6.1 years)
"Little bit good, coz not like really good, but do read stuff that is quite hard" (girl; 8.8 years)
"Not good at home, am good at reading at school, coz school books have big writing, at home they have little writing" (girl; 5.9 years)

7% of all the children gave reasons that did not fit into any given category (Year 1: 6%, Year 3: 7%).

One third of the children provided no reasons for why they were good or not at reading (Year 1: 37%, Year 3: 30%). Of these children 48% gave the name of a book they had read or what their favourite book was, however these answers did not offer any real justification for their response choices, e.g.;

"Very good, power rangers books I read" (boy; 8.3 years)
"Little bit, read Micheal Jackson and Westlife books" (girl; 8.1 years)
"Really good, like reading goosebumps" (boy; 6.4 years)
"Very good, read the little ginger bread man" (girl; 6.3 years)

b. Remembering things

As shown in Table 9.16, nearly three quarters of the children gave specific examples of things they had to remember (Year 1: 67%, Year 3: 72%). Of these examples 57% related to things that they had to remember at home for their parents or family members, 39% related to things they had to remember for school, and 4% related to things they had to remember for their friends, for example:

"Don't forget, remember to feed geco, and tidy my room" (girl; 7.7 years)
"Forget, once asked me to clean room but I watching my TV and I forgot" (girl; 8.0 years)

"I forget sometimes I forgot to brush my teeth, sometimes I forget to have breakfast, sometimes I forget to wake my brother up" (girl, 6.4 years)
"Forget, like to take my plant to school, I forgot for last two days" (girl; 8.5 years)
"Forget a little bit, once Mr. Cameron told us what to do with homework and forget it when got home" (girl; 7.9 years)
"Always remember, have to remember things like coz had this letter about going early on Thursday and remember to take it" (girl; 5.6 years)
"Not so good, coz I got to round for my friends and I forget to go" (boy; 8.6 years)
"Remember, must remember to go to call people, all my friends and that" (boy; 5.9 years)

3% of children mentioned stable character attributes in their answers (Year 1: 2%, Year 3: 4%), e.g.;

"Sometimes... don't have a very good mind, I forget what to do, at school, I don't get it" (girl; 8.11 years)
"Remember a lot, always being told that I've got a good memory, so really good" (girl; 8.6 years)
"Sometimes forget, got a short memory and that" (boy; 6.3 years)

15% of children gave other answers that showed some understanding of the item (Year 1: 14%, Year 3: 16%), for example:

"Forget stuff, so ask other people" (boy; 7.9 years)
"Forget a little, like hard things, but easy things I don't forget" (girl; 8.5 years)
"Both coz sometimes I forget but this morning I remember" (boy; 6.2 years)
"Remember what to do, remember a lot of things, can't think of any now!" (boy; 6.4 years)

12% of children gave no reasons for their ability to remember things (Year 1: 17%, Year 3: 8%).

Psychological Functioning items

The children were asked about whether they get cross or angry and whether they worry about losing their things (e.g., their toys or clothes), and the reasons they gave were coded into one of four categories shown in Table 9.17.

Table 9.17: Percentages of children's answers to psychological functioning items

Coding category	% of children Feeling cross	% of children Worrying about losing things
Social comparison used	0 (0)	0 (0)
Stable character reference used	6 (17)	3 (8)
Concrete example given	72 (199)	70 (194)
Other reasons	3 (8)	15 (42)
No further reason given	19 (53)	12 (33)

a. Feeling cross

As shown by Table 9.17, nearly three quarters of the children referred to a concrete example of someone that had made them cross recently (Year 1: 69%, Year 3: 75%). Of the answers 55% related to when their siblings had made them feel cross, 32% related to their friends, 13% related to their parents. Some examples of these types of answers are given below.

"Little bit cross, when brother touches my stuff when he's not meant to" (girl; 8.6 years)
"Lots of things, like when my sister doesn't listen, she mimes and that really annoys me" (girl; 8.5 years)
"A lot cross, my brother and sister fight all the time, I have to shout to shut them up" (boy; 6.7 years)
"Little bit cross, when friend talks about me in school and I tell teacher" (boy; 7.7 years)
"Little bit cross, when friend say stuff and lie to me" (boy; 5.9 years)
"Very cross, when I know somethings right and mum keeps arguing with me" (boy; 8.6 years)
"Really cross, when mummy shouts at me" (girl; 6.2 years)

6% of the children referred to stable characteristics in their answers (Year 1: 4%, Year 3: 9%), e.g.;

"Not so many times cross, I'm a happy person, nothing makes me cross" (girl; 8.4 years)
"Not very many times get cross, partially coz I've been brought up not to be bad tempered" (girl; 8.6 years)
"Not really cross, I'm happy every day" (girl; 6.7 years)

3% of children gave other reasons that could not be coded (Year 1: 5%, Year 3: 2%):

"Little bit cross… I want to be a normal girl" (girl; 8.2 years)
"Not cross ever, if cross, I lose my voice and I don't like to lose it" (girl; 6.3 years)
"Not much coz don't really get angry or fight" (boy; 6.6 years)

19% gave no explanation of why they had felt cross or not, past indicating their choice on the response scale (Year 1: 22%, Year 3: 14%).

b. Worrying about losing things

As shown by Table 9.17, two thirds of children referred to specific examples in their answers (Year 1: 51%, Year 3: 67%). Of these children 62% of the examples related to whether what they lost was special or not, 27% related to whether they had lost something before and how it had made them feel. 11% reported that they did not worry as if they lost something they would either find it again or buy a new one. Some examples are shown below.

"Little bit worry, like my jumper that have lost already, it's my new one, lost yesterday" (girl; 8.6 years)
"Not so much worry, the other day I left my jumper at school, but found it the next day" (girl; 8.5 years)
"Little bit worry, about losing some of my things, that are special to me, things that someone gave to me" (girl; 8.5 years)
"Little bit, coz if get new stuff, am worried, but if old, not bothered" (girl; 8.3 years)
"Not really coz everytime I lose something, mum and me search for it and we find it" (girl; 7.5 years)
"I worry a lot, I got a new pencil that's already lost and I'm sad" (girl; 6.6 years)
"Depends if fave thing, don't worry if just a sock, don't mind then" (boy; 6.4 years)
"Don't coz my mum says don't worry if you lose something, coz we'll find it" (girl; 5.9 years)
"Never worry coz don't matter coz mum buys me another" (girl; 5.5 years)

10% of the children mentioned stable attributes in their answers (Year 1: 12%, Year 3: 8%), e.g.,:

"Not very much, don't worry coz don't really lose my stuff, look after it" (girl; 8.6 years)
"Not at all, always look after my toys, put them in my toy box" (boy; 5.9 years)
"Never lose mine, coz just don't, put them in a safe place" (boy; 5.4 years)

13% of the children gave answers that could not be coded (Year 1: 15%, Year 3: 11%), e.g.;

"Don't coz don't like it when worry, when worry keep crying" (girl; 7.5 years)
"Little bit, like when go to brownies, don't want to lose anything" (girl; 8.4 years)
"Do worry might lose it, mummy smack you" (boy; 6.1 years)
"Don't worry about my things, coz I play with them every time" (boy; 5.8 years)

18% did not give any explanation for their response choice on the item (Year 1: 22%, Year 3: 14%).

Family Acceptance items

The children were asked whether they got told off a lot at home or not, and whether their brothers or sisters bossed them at home. The reasons they gave during cognitive interviewing were coded into one of the four categories shown in Table 9.18.

Table 9.18: Percentages of children's answers to family acceptance items

	% of children	
Coding category	**Told off at home**	**Siblings bossing***
Social comparison used	6 (17)	0 (0)
Stable character reference used	33 (91)	20 (56)
Concrete example given	39 (108)	35 (97)
Other reasons	6 (17)	3 (8)
No further reason given	16 (44)	26 (72)

*Note. *16% (n=44) of children reported having no siblings at home during 'think aloud' method (Year 1: 23%, Year 3: 10%).*

a. Getting told off at home

As shown in Table 9.18, two fifths of the children mentioned a specific instance when they had been told off before (Year 1: 33%, Year 3: 45%). Of these reasons 57% related to when they fought with their siblings or when their siblings did something to annoy them, and the other 43% related to a wide variety of other things they had done wrong at home. Some examples are given below:

"Little bit told off, like sister slaps me and dad doesn't see, then slap back and get caught" (girl; 8.3 years)
"A lot told off, when my friends there and sister wants to play, no coz it's my friends, but mum says let sister play so get told off then" (girl; 8.3 years)
"Just once, coz once my brother pushed me and I fell into door, he told on me, and it wasn't my fault" (girl; 5.9 years)
"Little bit, um, like playing on computer too much and for staying up late" (boy; 8.3 years)
"Little bit told off, for swearing and that, if swear mum takes money from my money box" (boy; 6.4 years)

One third of the children referred to stable character attributes of themselves as reason why they did not get told off at home (Year 1: 37%, Year 3: 29%), for example reporting that they were well behaved at home or did what they were asked to do, e.g.;

"I can get told off sometimes, am bit cheeky, they way I speak to mum isn't that good" (girl; 8.3 years)
"Not much told off, coz normally I'm quite good at home" (boy; 8.1 years)
"Don't get told off coz I'm quite helpful at home" (boy; 6.2 years)

"Not at all, coz am good girl at home" (boy; 6.1 years)

6% of the children made social comparisons in their answers (Year 1: 4%, Year 3: 8%), and these related to how much they got told off compared to their brothers and sisters, e.g.;

"Don't get told off coz I don't do anything bad, but my brother does, and I used to, but no he's the youngest s he does, that's fair" (girl; 8.1 years)
"Don't get told off, coz I'm gooder than my sisters" (girl; 7.5 years)
"Not at all, I'm never naughty at home, my brother is naughty instead, he swears and he gets sent off to nans" (girl; 6.4 years)

6% gave reasons that showed some understanding of the item, but could not be coded(Year 1: 7%, Year 3: 5%), e.g.;

"Don't really as like when come home from school, out of breath and don't want to do anything else" (girl; 7.9 years)
"Don't really coz mummy loves me" (girl; 5.5 years)

16% of children gave no explanation for why they did or did not get told off at home (Year 1: 19%, Year 3: 13%).

b. Siblings bossing them around

As shown in Table 9.18, one third of the children gave concrete examples of when or how their siblings had bossed them before (Year 1: 27%, Year 3: 43%), for example:

"Yeah they do, every time I go in the tree house, they push me back out" (boy; 7.9 years)
"Big brother, yeah, he says go and get me a drink, hurry up, he never does anything for me" (girl; 8.3 years)
"He boss me around a lot, he boss me around for his gameboy charger, when his gameboy is going out charge, he says go and get it" (boy; 6.6 years)

One fifth of children referred to stable character attributes when answering the item (Year 1: 22%, Year 3: 17%). Of these answers 65% referred to their age compared to their siblings when giving reasons for being bossed or not, and 35% reported that they got on well with their siblings as a reason for not being bossed around. Examples of these answers are given below.

"Not at all, coz my brother is younger so can't boss me" (girl; 8.6 years)
"A lot they boss me, coz I'm the youngest and they always be horrible to me" (boy; 8.6 years)
"She doesn't know how to talk yet so she can't" (girl; 6.4 years)
"Not at all boss me, coz she's only little, she's three months old" (boy; 6.1 years)

"They don't boss me, coz I like them and they're nice, sister shares sweet with me, lets me have her biscuit" (boy; 8.2 years)
"Don't coz they're good boys" (boy; 6.1 years)
"Really don't boss me, I'm nice to them and after I been nice to them, they are nice to me" (girl; 6.6 years)

3% gave answers that coded not be coded (Year 1: 0%, Year 3: 7%), e.g.;

"She's bossy, she used to slap me" (girl; 8.8 years)
"Does boss me, he makes me cross, he teases me and calls me fat" (girl; 7.9 years)
"She doesn't boss me, she likes to play her own toys" (girl; 6.2 years)
"She bosses me all the time... I don't like her, I don't like girlie whirlies" (boy; 5.8 years)

A quarter of the children did not give any explanation of why they did or did not get bossed or teased by their siblings (Year 1: 28%, Year 3: 23%).

Peer Acceptance items

The children were asked how many friends they had at school, and whether their friends bossed them around. The reasons they gave to support their response choices were coded into the categories shown in Table 9.19.

Table 9.19: Percentages of children's answers to peer acceptance items

	% of children	
Coding category	**Having friends**	**Friends bossing***
Social comparison used	0 (0)	0 (0)
Stable character reference used	20 (56)	25 (70)
Concrete example given	31 (86)	29 (80)
Other reasons	10 (27)	10 (27)
No further reason given	39 (108)	23 (64)

*Note. * 13% (n=36) of children reported that their friends did not boss them around at all (Year 1: 11%, Year 3: 15%).*

a. Having friends at school

When answering about whether they felt that they had a lot of friends or a few, a third of the children gave specific examples of why they had a lot or just a few friends (Year 1: 28%, Year 3: 35%). These answers fell into two types - with children either reporting that they had lots of friends to play with (82%) or not enough friends to play with (18%), for example:

"Lots, coz every night and day, people always call, more than 17! Got a list" (girl; 8.1 years)

"Lots of friends coz like today, I didn't play with same people as yesterday, play with lots of different people" (girl; 8.5 years)
"Lots, coz everyone plays with me" (boy; 5.9 years)
"Not got many, coz sometimes they get ill, so got no one to play with" (girl; 8.3 years)
"Just some, three friends, two friends don't play with me very much, get lonely some" (boy; 5.7 years)

One fifth of children mentioned stable character attributes in their answers (Year 1: 18%, Year 3: 23%), e.g.;

"Lots of friends coz I'm quite popular and all know me a lot" (girl; 8.2 years)
"Lots of friends, I get on with them and be kind to them" (girl; 8.3 years)
"Lots coz I always be nice to them" (girl; 6.1 years)
"Lots coz I think I be nice and play good games" (boy; 6.7 years)

10% of children gave answers that could not be coded (Year 1: 6%, Year 3: 12%), e.g.;

"A few friends, they sometimes play" (girl; 8.7 years)
"Quite a few friends, like more coz they play same as what I do" (boy; 8.0 years)
"Lots of friends coz made friend with them" (boy, 5.5 years)

Just over a third of the children gave no further explanation for why they had a lot or a few friends (Year 1: 48%, Year 3: 30%). Of these children, 46% of children listed their friends names, or how many friends they had (e.g., 'I have 100 friends') which gave no further information on why they had lots or a few friends, for example:

"Lots, coz I got like Leon, Bradley, Jon, Aaron and Brett" (boy; 7.8 years)
"Lots, coz I got about 50 or 60" (boy; 7.8 years)
"Lots, coz got 100 and 100 friends, some in Dundee" (girl; 5.6 years)
"Lots coz tell all their names, Anna, Charlie, Sarah, seven by now" (girl, 6.4 years)

b. Friends bossing them around

As shown in Table 9.19, a third of children made reference to specific examples of when their friends had bossed them around before (Year 1: 27%, Year 3: 32%), e.g.;

"Little bit, coz my friends, when in class she always chooses the game and she bosses our whole table" (girl; 8.5 years)
"Well, yeah, abit boss me, when I want to play my own game, tell me to play their game, I don't like that" (girl; 6.2 years)

A quarter of the children gave example of a stable character attribute when answering the item (Year 1: 27%, Year 3: 21%). Of these answers 72% reported that their friends did not boss them as they were nice to them, 16% related whether they got bossed or not to their age, and 12% related their reasons to whether they bossed their friends around themselves. Some examples of these answers are given below:

"Don't boss me, coz don't be nasty to me, all friendly to me" (boy; 8.2 years)
"Not at all, I'm really nice to them and they're nice to me" (boy; 8.4 years)
"Don't coz like all very kind to me, got nice friends" (girl; 5.6 years)
"Don't like to boss me, coz I'm the oldest of Year 3 friends" (boy; 8.6 years)
"Don't coz well seeing as I'm the oldest, I'm sort of the leader of the gang" (boy; 5.8 years)
"Not so much boss me, coz I don't really boss them much so they don't boss me" (boy; 8.1 years)
"Don't boss me, coz I'm not bossing them around" (girl, 6.2 years)

10% of children gave answers that showed some understanding of the item, but could no be coded (Year 1: 11%, Year 3: 9%), e.g.;

"Don't boss me, coz they've been my friends for a long time" (girl; 8.5 years)
"No, coz I tell Miss and then they tell them off" (boy; 7.6 years)
"Boss me, they sometimes get angry and I get angry with them for copying" (boy; 6.7 years)
"Boss me coz when I play, boss me in the street" (girl; 6.0 years)

A quarter of the children gave no explanation for their response choice to this item (Year 1: 24%, Year 3: 23%).

9.1.4 Identifying items that could be removed from the TedQL.3 measure

Distribution of children's scores on individual items

The distribution of children's actual QOL and discrepancy scores was examined across all of the 30 items within the TedQL.3 measure using histograms and calculations of descriptives (i.e., means, standard deviations, range, skew, and kurtosis). As reported in section 9.1.1 (p. 231-2) the majority of children's mean TedQL.3 scores were significantly skewed towards higher QOL and lower discrepancy reports, and showed high levels of kurtosis. These patterns were also evident when examining children's scores to individual items within the TedQL.3. Examining the distribution of children's actual QOL scores revealed two items that had almost identical distribution patterns, with the same mean scores and standard deviations (M=2.42, SD=0.89). The items were PA1 (having friends to play with at school) and PA4 (having lots of friends, versus none). Figures 9.1 and 9.2 show the similarity between the distributions of children's answers to these two items.

Figure 9.1: Histogram of children's actual QOL scores for item PA1

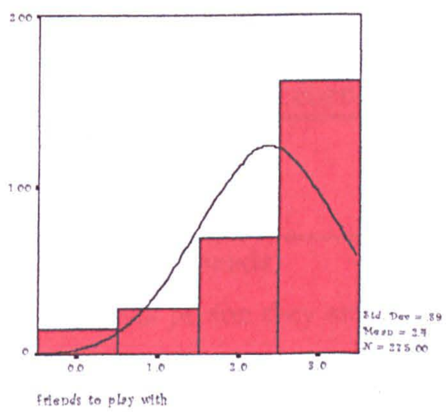
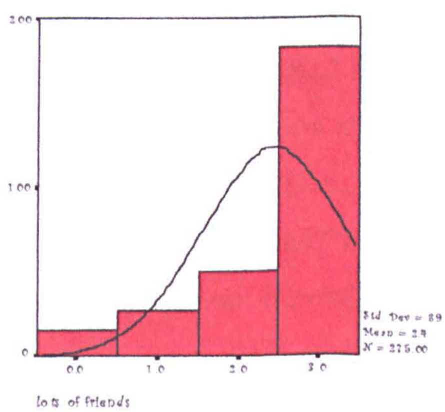


Figure 9.2: Histogram of children's actual QOL scores for item PA4



Percentage of 'don't know' answers on individual items

The percentage of children using the 'don't know' option to answer the items from the TedQL.3 measure was calculated for each item in the TedQL.3 measure. For the majority of the items the percentage of 'don't know' answers was generally low (see Appendix E for full list of percentage for all thirty items).

There were three items that were found to be most frequently answered with the 'don't know' option. These items were: PC4 ('playing computer games', 23%), FA3 ('siblings playing with them', 11%), and FA7 ('siblings bossing them around', 12%).

Item-total correlations for individual items

Item-total correlations were calculated to assess whether items were contributing reliability to the measure (see Appendix E). There were three items on which children's scores were not highly correlated to their total actual QOL and discrepancy scores with correlation values falling below the .20 standard. Table 9.20 shows the three items with the item-total correlation values.

Table 9.20: Item-total correlations for individual items on TedQL.3

	Spearman's correlation coefficient (ρ)	
	Actual QOL	**Discrepancy**
Item		
PA2 – bossing friends	0.14*	0.17*
PF1 – type of person they are	- 0.04	0.05
PF4 – playing pretend games	0.13*	0.14*

Note. * $p<.05$

Conceptual similarity of individual items within each domain

There were seven TedQL.3 items that were related to children's family lives (see Table 8.1, p. 218). These items all broadly referred more to relationships in the family, or emotional experiences (e.g., 'does your mum/dad play with you at home?' or 'do your siblings boss you around?'). However, considering the conceptual similarity between items in this domain showed one item that was somewhat different (FA5, whether they went on trips with their parents).

Of the six items concerned with children's cognitive functioning (see Table 8.1, p. 218), two items did not 'fit' as well with the concepts asked about within this domain (CF1, 'playing board games', and CF5, 'remembering what they are told to do'). The other four items ask children about cognitive skills that are more concrete and visible (e.g., reading and writing abilities).

Items removed from the TedQL.3 measure

Based on our analyses, we removed eight items from the TedQL.3 measure to produce a new version of the TedQL measure (the TedQL.4). Table 9.21 shows the changes that were made to the item content of the TedQL.3 as a result of Study 4.

Table 9.21: Items removed due to results of Study 4

Items in existing version (TedQL.3)	**Items in new version (TedQL.4)**
Physical Competence:	T1: good at swinging
PC1: good at swinging	T2: playing ball games
PC2: good at playing ball games	T3: good at climbing high things
PC3: good at climbing high things	T4: good at running
<u>PC4: good at playing computer games</u>	T5: good at bike riding
PC5: good at running	T6: having friends to play with
PC6: good at bike riding	T7: friends coming over to their house
	T8: friends teasing them (N)
Peer Acceptance:	T9: mum/dad playing with them at home
PA1: having friends to play with	T10: telling mum/dad what been doing at school
<u>PA2: bossing friends (N)</u>	T11: playing with siblings
PA3: friends coming over to their house	T12: siblings bossing them (N)
<u>PA4: having lots of friends at school</u>	T13: seeing grandparents
PA5: friends bossing them (N)	T14: getting told off at home (N)
	T15: feeling scared (N)
Family Acceptance:	T16: having bad dreams at night (N)
FA1: mum/dad playing with them at home	T17: getting cross/angry (N)
FA2: telling mum/dad what been doing at school	T18: worrying about losing things (N)
FA3: playing with siblings	T19: doing mathematics/numbers
FA4: seeing grandparents	T20: writing
<u>FA5: going on trips with mum/dad</u>	T21: drawing
FA6: getting told off at home (N)	T22: reading
FA7: siblings fighting/bossing them around (N)	
Psychological Functioning:	
<u>PF 1: type of person they are</u>	
PF2: getting scared (N)	
PF3: having bad dreams at night (N)	
<u>PF 4: playing pretend games</u>	
PF5: getting cross/angry (N)	
PF6: worrying about losing things (N)	
Cognitive Functioning:	
<u>CF1: good at playing board games</u>	
CF2: good at mathematics/numbers	
CF3: good at writing – spell name/other words	
CF4: good at drawing – what can draw	
<u>CF5: remembering what people tell them to do</u>	
CF6: good at reading – read words/look at pictures	

Note. (N)= negatively scored item, scores were reversed
Key to alterations to items for TedQL measure:
<u>Underlined = item deleted from measure as result of Study 4</u>

9.2 Discussion

9.2.1 Establishing the most appropriate response scale for the TedQL.3 measure

Our review of child self-report measures in Chapter 3 revealed the need for a clearer basis and justification for a chosen response scale, due to the wide variety of different scales that have been previously employed in child self-report measures (see p. 66). The first aim of this study was to establish the most appropriate response scale for use in the TedQL.3 measure. This aim was achieved by comparing the psychometric properties of the TedQL.3 measure across three different scales (circles, faces, and thermometer scales). We predicted that the circles response scale would be the most appropriate scale for our measure, and therefore children's responses using this scale would produce the 'best' psychometric properties for the TedQL.3. Four predictions were made in relation to the superiority of the children's responses using circles scale, and these are discussed below.

First, we predicted that children using the circles scale would produce fewer 'don't know' answers to the TedQL.3 items. This prediction was not supported (see Table 9.6, p. 234). Year 1 children produced the lowest amount of 'don't know' responses at Time 1 when using the *faces* scale (2.11%) and at Time 2 using the *thermometer* scale (2.15%). Year 3 children produced the least amount of 'don't know' answers when using the *faces* scale at both time points (Time 1: 1.67%, Time 2: 2.08%). However the amount of 'don't know' responses was very low across all three scales (ranging from 1.67% to 3.18%), suggesting that the majority of children were able to answer the majority of the TedQL.3 items using any of the three response scales.

Our second hypothesis was not supported; i.e., that children's responses to the TedQL.3 items using the circles scale would show a higher number of item-total correlations falling above the .20 standard. Across all three responses scales and over both age groups, the majority of the item-total correlations were above the .20 standard (ranging from 70% to 85% being above .20, see Table 9.7, p. 234-5). This result means that the majority of the TedQL.3 items were contributing reliability to the measure, regardless of the type of scale used to represent response choices.

Third, the prediction that children's responses using the circles scale would show higher internal consistency was partly confirmed (see Table 9.8, p. 235). It is important to establish the internal reliability (or consistency) of any new measure. Overall children's actual QOL scores using the *circles* scale produced internal consistency ratings for the TedQL.3 measure that were above the 0.70 standard for both age groups (Year 1: α= .72; Year 3: α= .74). Whereas, children's discrepancy scores using the *thermometer* scale produced internal consistency values that were above this standard for both ages (Year 1: α= .73; Year 3: α= .76).

The fourth prediction, that children's responses using the circles response scale would show higher reproducibility over time, was not confirmed (see p. 236-7). Another important aspect of reliability for any measure is whether the responses to items are reproducible over time (i.e., test-retest reliability). The responses of Year 1 children showed the highest reproducibility when using the *thermometer* scale to answer items (actual QOL: $\rho\pm$ = .77; discrepancy: $\rho\pm$ = .78, see Table 9.10, p. 237). Year 3 children's responses showed the highest reproducibility when using the *faces* (actual QOL: $\rho\pm$ = .77) and *thermometer* (discrepancy: $\rho\pm$ = .62) scales to answer the TedQL.3 items (see Table 9.10). Overall, across both age groups, children using the *thermometer* scale produced responses with ICC ($\rho\pm$) values over the .60 standard for both their actual QOL and discrepancy scores (see Table 9.10, p. 237). These results suggest that the thermometer response scale would help children of this age produce responses that are more consistent over time.

We also predicted that children would prefer the circles scale over the faces and thermometer scales. This hypothesis was not supported. Year 1 children preferred the *thermometer* scale the most (26 out of 41), and Year 3 children expressed an almost equal amount of liking for the *faces* (20 out of 40) and *thermometer* (17 out of 35) scales (see Table 9.11, p. 238).

An additional prediction was made concerning the comparability of children's responses to the *same* items when using *different* scales to answer them (answering using one scale at Time 1, and a different scale at Time 2). We predicted that children's responses would not be comparable across different scales, i.e., that their responses to items using one scale would not be correlated to their responses to the same items using a different scale. This prediction was partly supported (see p. 239-40). Across both age groups, the children's responses using the *circles* and *faces* scales were correlated with each other (Year 1: actual QOL ρ= .95; discrepancy ρ= .95; Year 3: actual QOL ρ= .79; discrepancy ρ= .56, see Table 9.13, p. 240). This result means children were interpreting and using the circles and faces scales in similar ways when answering TedQL.3 items.

Year 1 and Year 3 children's responses using the *faces* and *thermometer* scales were not correlated with each other (Year 1: actual QOL ρ= .40; discrepancy ρ= .41; Year 3: actual QOL ρ= .39; discrepancy ρ= .40, see Table 9.13, p. 240). This means children were not producing comparable responses to TedQL.3 items when using the faces and thermometer scales.

Year 3 children's responses using the *circles* and *thermometer* scales were not correlated to each other (actual QOL: ρ= .19; discrepancy: ρ= .39, see Table 9.13, p. 240). However Year 1 children's responses using these two scales were correlated to each other (actual QOL: ρ= .57; discrepancy: ρ= .55, see Table 9.13, p. 240). We are unsure exactly why these age differences occurred, i.e., why Year 1 children's responses were comparable across the circles and thermometer scales, but Year 3 children's responses were not.

Although our results are somewhat contradictory to the predictions we made concerning the comparability of responses across different scales, we can make sense of these results by considering the differences between these types of scales. The thermometer scale is a linear response scale, where all the points lay on a continuum, and the child has to consider the whole thermometer when using this scale to rate items. However the circles and faces scales represent bipolar response scales. These differ from linear scales as the scales are divided into one side or the other, and the child has to consider which side their choice lies on when rating items. Therefore we argued that children's responses when using essentially two forms of a bipolar scale (circles and faces) would be correlated to each other, and that children's responses when using a linear (thermometer) and a bipolar scale (circles or faces) would not be correlated.

The results of this study have highlighted that choice of response scale can impact directly on the psychometric properties of child self-report measures. The results have also shown that different types of scales may not produce comparable ratings from children on the same items. Our results have highlighted the value of providing a justification for a chosen scale, and ensure that children can understand and use the scale in the ways intended. Our study built on the work of Chambers and Johnston (2002) by using three different response scales (i.e., circles, faces and thermometer) as opposed to one type (i.e., Likert). Our work extended Rebok et al.'s (2001) findings by providing assessment of the impact of scale type on the reliability and reproducibility of children's self-reports.

9.2.2 Investigating the strategies used to answer items in the TedQL.3 measure

We needed to gain information on how children below eight years attempt to answer self-report items (i.e., the strategies they engage in to come up with their answers). The second aim of this study was to investigate the strategies children used to answer some of the TedQL.3 items. We achieved this using a combination of two interviewing techniques (probes and the think aloud method). Children were asked to 'think aloud' while answering ten TedQL.3 items, and were probed for the reasons for their response choice. Study 4 extended the findings of Rebok et al. (2001) and Valla et al. (1994), by developing a coding system which allowed examination of the strategies children use when answering questions as opposed to assessing their understanding of items. The results from the content analysis of the answers given by the children offered some interesting insights into how children answer self-report items, and the processes they engage in when rating their abilities, functioning, feelings, and relationships (see p. 241-53).

We predicted that children would report different types of strategies when answering different types of items. Specifically we predicted they would be more likely to use social comparisons when answering physical competence items, and be more likely to refer to stable character attributes on psychological functioning items. Our hypotheses were not confirmed by the results of this study. Although 38% of the children did make social comparisons when referring to their running ability (see Table 9.15, p. 241), the use of social comparisons was not very common amongst many of the children for any other items. The majority of the children used concrete examples of specific situations or instances that had happened to them as reasons for their response choices, regardless of item type (see p. 241-253). For eight out of the ten TedQL.3 items that were used during interviewing, concrete examples were the most common strategy used by children (with the exception of two items; 'running' and 'reading', see p. 241, & p. 244-5)

The fact that children below eight years tended to use specific instances or concrete examples when rating their lives may explain why younger children's self-reports are often less stable over time and may fluctuate more than older children's ratings (i.e., children over eight years). Our results led us to the conclusion that younger children's personal and remembered examples of when they last got told off, or when they got cross or worried about something could easily change over a short time period, which could in turn mean their responses were be less reproducible. Older children may be more likely to refer to stable character attributes, and would therefore provide more stable self-reports (Ruble et al., 1980). We argued that this result does not mean young children's self-reports are unreliable over time, but that their answers may fluctuate more as they use different strategies when answering items. It may be that researchers could improve reliability over time on self-report items by prompting young children with the concrete example they used at the previous time point.

9.2.3 Identifying items that could be removed from the TedQL.3 measure

The third aim of this study was to investigate items that could be removed from the TedQL.3 measure, to produce a quicker measure that was easy to administer to children. We conducted analyses with the complete data set at Time 1, and the results of our analysis revealed eight items that we felt should be removed from the TedQL.3 measure (see Table 9.21, p. 257). Our reasons for deleting these eight items have been explained in the sections that follow.

The distributions of children's scores on two of the peer acceptance items (PA1 and PA4) were found to be almost identical (see Figures 9.1 & 9.2, p. 254). When we examined the wording of these items ('having friends to play with' versus 'having lots of friends at school'), we realised that due to the similarity of wording, children may have been treating these items as the same question. This meant that one of the items was redundant, and therefore we made the decision to remove PA4 ('having lots of friends at school'), and retain PA1 ('having lots of friends to play with').

Three items were most frequently answered with the 'don't know' option (PC4, FA3, & FA7, see p. 255). 23% of the children answered 'don't know' to the item PC4 ('how good they were at playing on computers'), and it was felt this item could be deleted due its lack of relevance to nearly one quarter of the children's lives. FA3 and FA7 asked children about their relations with their siblings, and it was evident from children's additional comments when answering these items that they chose the 'don't know' option because they did not have siblings at home to play or fight with. However as the interview data in Study 3 (see Chapter 5, p. 153-4) had clearly shown that siblings were a big part of most children's lives we felt the continued inclusion of these items in the TedQL.4 was justified.

There were three items that were not correlated with the children's total scores on the TedQL.3 measure (PF1, PF4, & PA2, see Table 9.20, p. 255). Therefore these items were not contributing reliability to the overall scores produced by our measure. The lack of correlation may be explained by considering whether these items were actually relevant to children's *QOL*, as well as being important in children's *lives*. For example whether a child is loud or quiet may not be relevant to their QOL (PF1), although whether they are loud or quiet would still be important in describing a child's personality. Therefore we felt that these items should be removed from the TedQL.3 measure.

We felt it would also be useful to consider whether the items in each domain 'fitted' well together, i.e., asked children to rate similar concepts. We found three items that should be removed from our measure, as a result of this analysis (see p. 256-7). First, the items in the family acceptance domain all broadly referred more to relationships in the family, or emotional experiences (e.g., 'does your mum/dad play with you at home?' or 'do your siblings boss you around?'). However, one item that was somewhat different to the others – FA5 (whether they went on trips with their parents). We argued that this item did not 'fit' as well with the other items in this domain, and therefore should be deleted.

Second, among the six cognitive functioning domain items, there were two items that also did not 'fit' with the concepts asked about in this domain – CF1 ('playing board games') and CF5 ('remembering what they are told to do', see p. 255-6). The other four items ask children about cognitive skills that are more concrete and visible (e.g., reading and writing abilities). These other items are all focused around skills that children are taught and evaluated on during their education (i.e., reading, writing, drawing and mathematics). Both the school environment and routine testing as part of the National Curriculum encourage comparisons with peers on these abilities. This may in turn make it easier for young children to judge how good they are at reading, writing, drawing or maths, than when they are asked to judge how good they are 'remembering things' which is a more abstract skill. Therefore these two items (CF1, CF5) were removed from our measure.

9.2.4 Alterations made to our measure – new TedQL.4 version

Three changes were made to the TedQL.3 instrument on the basis of the results of Study 4. These changes resulted in a new version of our measure (TedQL.4). This version differed from the earlier versions of our measure in that it: used a thermometer response scale to represent the response choices (as opposed to circles), contained twenty-two items for children to answer (as opposed to thirty items), and presented the items to children using four teddy bears (as opposed to two). Each of the alterations made to our measure have been explained below.

In Study 4 we compared the psychometric properties of children's responses when answering TedQL.3 items across three different scales. Contrary to our predictions, children using the *thermometer* response scale produced the 'best' psychometric properties for the measure, although the results were somewhat mixed. Children using the thermometer scale showed some of the highest internal consistency ratings, and the highest reproducibility of responses over time. The children themselves also preferred this scale to the circles and faces scales. Therefore we decided that the thermometer scale would be used to represent the response choices to children for the new version of our measure (TedQL.4).

Researchers in the social attitude field have suggested that the response scale should be matched to the type of question, i.e., that bipolar scales are more suited to questions rating attitudes which fall clearly onto two sides; and linear scales are more suited to questions rating abilities, e.g., from good to not good (Schwartz & Oyserman, 2001, Schwartz et al., 1998, Schwartz & Sudman, 1995). As discussed in Chapter 7 (p. 194-5), the type of response scale (e.g., bipolar or continuous) gives information about the focus of the items. As such scale type may influence how the respondent views the attribute in question, or how they report the frequency of their behaviours (Knauper & Turner, 2003). The meaning conveyed by the scales used in a response scale needs to match the types of items being rated (McLaughlin, 1999). Following this rationale, the thermometer response scale would be more suited to the TedQL.3 items which ask children to rate how 'good' they are at given skills or how 'much' they feel something. We felt that using the thermometer scale for these question types could help younger children to make sense of the question and make the rating task easier for them.

In this study we also investigated items that could be removed from the TedQL.3 measure to produce a shorter and more focused version. The removal of eight items from the TedQL.3 resulted in a 22-item version of the measure (the TedQL.4, see Table 9.21, p. 257). We felt that this shortened version would be easier to administer to children, and meet the need for a short, quick QOL measure for clinicians and other child health professionals.

Originally our measure was administered using two teddy bears to illustrate either side of the item to children (see Figure 3.3, p. 70), and children were required to choose which bear was most like them (i.e., either 'Iggy' or 'Ziggy'), and then they had to rate whether they were a lot like this or just a little bit (see p. 120). We felt that children would benefit from having teddy bears to represent all four response choices on the TedQL.4 measure (i.e., four teddy bears as opposed to two). We chose four identical teddy bears, which could only be identified as different by their badges – four different shapes: circle, square, triangle, and diamond (see Figure 9.3). We introduced this change into the new version of our measure (TedQL.4).

Figure 9.3: Photograph of the four teddy bears (TedQL.4)

9.2.5 Re-evaluating existing measures in light of our results

In light of the findings of this study, we should considered the response scales chosen by scale developers for existing child self-report QOL measures. We found that the thermometer scale was the most appropriate scale type for use with the TedQL measure with children aged below eight years. In addition, the results of Study 4 provided evidence that bipolar scales (like the circles and faces scales) may be unhelpful for representing response choices to children in measures asking for responses to items such a ability (where the responses lie on a continuum from *not at all* to *a lot*). Our results could mean that response scales used in QOL measures such as the Childhood Asthma Questionnaire (Christie et al., 1993, French et al., 1994) which uses a bipolar faces scale, and self-esteem measures such as the Pictorial Scale of Perceived Competence and Social Acceptance (Harter & Pike, 1984) which uses a bipolar circles scale, need to be re-considered. By comparison, measures such as the Generic Children's Quality of Life Questionnaire (Collier et al., 1997, 2000) which uses a continuous linear scale, the PedsQL (Varni et al., 1999, 2002, 2002) which uses a numerical, linear scale, and the Exeter Health-related Quality of Life measure (Eiser et al., 1999, 2000) which uses a visual analogue scale, employ response scales better matched to the types of items they require children to rate.

9.3.6 Issues for further research

Researchers have expressed concern with the use of interviewing techniques, advocating that "thinking out loud" and probing may break the question flow and the relationship between items (Czaja, 1998). It may be that the interviewing process interfered with how children were answering the questions in our measure, causing children to answer items differently to how they would have if they had been asked the items without probing or having to "think aloud". Another limitation with such methods is that they rely on individual's verbal reports of cognitive processes, and it may be that some of these processes occur beyond conscious awareness, and therefore may not be accessible for retrieval from working memory (Tourangeau, Rips, & Rasiniski, 2000). Bearing these criticisms in mind, it may be useful to give children the new version of the TedQL.4 measure without the additional mental tasks required when using TAM and follow-up probes alongside answering these self-report items.

As discussed in Chapter 4 (see p. 96-7), we need to compare the psychometric properties of the TedQL.4 measure to existing self-report measures for children below eight years. In Studies 1 and 2 we compared the PedsQLTM4.0 to the TedQL.1 and TedQL.2 (see Chapter 4, p. 112 & p. 129). The PedsQLTM4.0 (Varni et al., 1999, 2002) has been developed for gaining self-reports from children as young as six years. This measure is questionnaire-based and uses a 3-point scale to measure generic QOL. Normative data for this measure has been collected in the U.S.A. (Varni et al., 1999, 2002). The PedsQL TM4.0 items have been worded in negative phrasing (e.g., "how much of a problem is running for you?"), whereas the TedsQL.4 has been worded from a positive standpoint (e.g., "how good are you at running?"). Due to the alterations made to the response scale, item content, and presentation style of the TedQL measure as a result of Study 4, we designed Study 5 to compare the new version of our measure (TedQL.4) to the PedsQLTM4.0 (see Chapter 10).

The TedQL.4 measure does not have a parent report version to allow comparisons between proxy and child self-reports. Although researchers have highlighted that child and parent reports may not always be correlated (e.g., Guyatt et al., 1997), and there may be valid reasons for this lack of concordance, it is still important for researchers to gain information from both sources wherever possible (Vance et al., 1998). In Study 5 we developed a parent version of the TedQL.4 measure to allow proxy comparisons, and Study 5 compared parent-child agreement across the TedQL.4 and the PedsQLTM4.0 measures (see Chapter 10, p. 277 & p. 209-2).

The following section (Section 4) reports the results of further validation of the TedQL.4 measure (Chapter 10), and provides a general discussion of the whole thesis (Chapter 11).

Section 4: Refinement of the measure and reflections on obtaining self-reports

Chapter 10: Further validation of the TedQL.4 measure – development of a parent report version and comparison to an existing measure (PedsQL$_{TM}$4.0) (Study 5).

Summary

Aims

Study 5 aimed to: i) compare the psychometric properties of the TedQL.4 to an established measure (PedsQLTM4.0), ii) investigate the relationship between these two measures, and iii) compare agreement between child and parent ratings across these two measures.

Sample

One hundred and forty-nine children (5.0-8.5 years) completed two QOL measures (TedQL.4, PedsQLTM4.0). One hundred and three of their parents completed the TedQL.4 and the PedsQLTM4.0 for their child.

Results

Children's and parents' responses were more consistent for the PedsQLTM4.0 and had fewer item-total correlations falling below the .20 standard compared with the TedQL.4. The TedQL.4 measure had fewer ceiling and floor effects and a lower reading age requirement for comprehension of items compared with the PedsQLTM4.0. There was a positive correlation between the children's actual TedQL.4 scores and their PedsQLTM4.0 scores. There was a negative correlation between children's discrepancy TedQL.4 scores and their PedsQLTM4.0 scores. The same pattern of correlations were found for parent-reported child QOL. There were no significant relations between child and parent rated child QOL on the PedsQLTM4.0, however child and parent rated child QOL were correlated across some of their scores on the TedQL.4.

Implications

The PedsQLTM4.0 measure had better psychometric properties than our TedQL.4 measure, and it may be that we need to develop the item content of our instrument further. Children's scores on the TedQL.4 measure were related to their scores on the PedsQLTM4.0 measure, which confirms both measures are tapping into a similar construct. The agreement between child and parent reported child QOL was higher for the TedQL.4 compared to the PedsQLTM4.0. This may have been related to how the instructions and items were worded for the TedQL.4 measure.

10.1 Introduction

Study 5 continued and built on the results of the previous four studies. As a direct result of analysis performed during the previous study (see Chapter 9, see p. 262-4), the TedQL.3 measure was reduced in length and altered considerably to produce a new version (TedQL.4). This new version contained 22-items, used four teddy bears to present items to children, and a used thermometer scale to represent the response choices.

10.1.1 Comparing the psychometric properties of the TedQL.4 measure to an established measure (PedsQLTM4.0 measure):

The first aim of this study was to compare the psychometric properties of the new TedQL.4 measure (child and parent report versions) with an established measure (PedsQLTM4.0). The PedsQLTM4.0 is a well-validated measure that has been used successfully with healthy children and those with chronic diseases, and their parents (see Chapter 4 for details on development of PedsQLTM4.0, p. 96 & p. 101-3).

Researchers have argued that new measures should be compared to existing measures as part of collecting evidence for psychometric validation (e.g., Graham et al., 1997, Langeveld et al., 1996). Differences between our new TedQL.4 and the PedsQLTM4.0 led to specific predictions about which measure would have the 'best' psychometric properties. The PedsQLTM4.0 measure was originally developed for older children and has been downwardly extended to younger children. Children (below eight years) are given the measure verbally and required to give their answers using a 3-point Likert response scale. The items are negatively worded, asking how much of a problem various activities have been for them in the last few weeks (see Chapter 4 for full details, p. 101-2). In comparison, the TedQL.4 measure has been developed specifically for use with children below eight years with the content of the items derived directly from child interview data (see Study 3, Chapter 5, Table 5.16, p. 162-3). This measure is administered in the form of a game in an interview style using three-dimensional visual aids (i.e., teddy bears). Children use a graphic thermometer response scale to answer the items that are acted out in front of them. The items are positively worded asking the children to rate how they are at various activities, and how they have been feeling over the last one week (see Chapter 8 for full details, p. 218-19).

The previous studies in this thesis have been specifically focused on developing both the content (Chapter 5, p. 136-7 & 162-3) and format (Chapters 8 & 9, p. 209-11 & 258-61) of the TedQL.4 measure. In addition, the literature reviewed in Chapter 6 highlighted the importance of ensuring that measures are child-centred, and specific to the age at which they are aimed (see p. 168-76). Therefore based on the extensive adaptations and changes that have been made to ensure the suitability of the TedQL.4 measure for children below eight years, we predicted that children and parents using the TedQL.4 measure would produce: 1) responses with higher internal consistency (or internal reliability); 2) a higher number of responses to items with item-total correlations above .20; and 3) fewer floor and ceiling level responses, compared to when they were using the PedsQLTM4.0 measure. We also predicted that the reading age required for children to be able to understand the TedQL.4 items would be lower than the reading age for the PedsQLTM4.0.

10.1.2 Investigating the relationship between the TedQL.4 and PedsQLTM4.0 measures

Part of the validation process for new measures is to establish construct validity (i.e., whether a measure actually assesses the underlying construct that it is intended to measure, see Chapter 3, p. 43-4). One way to assess the convergent validity of a new measure is by comparisons to similar existing measures, which are hypothesised to be measuring a similar construct (Bryant, 2000). Therefore, the second aim of this study was to explore the relationship between the TedQL.4 and PedsQLTM4.0 measures. This aim was achieved by investigating whether children's and parents' TedQL.4 scores were correlated with their PedsQLTM4.0 scores.

The TedQL.4 measure is based on the argument that an individual's QOL is related to whether their current experiences and abilities match their expectations (Calman, 1987), i.e., how much their 'actual' selves differ from their 'ideal' selves. The TedQL.4 measure produces two types of scores: actual QOL and discrepancy scores (which are a measure of how much the children's ideal selves differ from children's actual selves). The actual QOL scores are a measure of the child's actual functioning, abilities and feelings (i.e., actual self), and therefore hypothetically should be related to their PedsQLTM4.0 scores. Following from this premise and the results of Study 1 (see Chapter 4, p. 112), we predicted that children's and parents' actual TedQL.4 scores would be positively correlated to their PedsQLTM4.0 scores. Although the discrepancy scores take individual preferences into account, the TedQL.4 items are in essence still measuring QOL and therefore should still be related to scores for the PedsQLTM4.0 items. Therefore we also predicted that children's and parents' discrepancy TedQL.4 scores would be negatively correlated to their PedsQLTM4.0 scores (i.e., as low discrepancy scores mean high QOL, children's and parents' discrepancy scores should go down as their PedsQLTM4.0 scores go up).

10.1.3 Comparing agreement between child and parent ratings, across the TedQL.4 and PedsQLTM4.0 measures

As discussed in Chapter 4 (see p. 98-9), researchers have advocated that a necessary requirement for the validation of new child measures is moderate agreement between proxy and child reports (e.g. Graham et al., 1997, Langeveld et al., 1996, Theunissen et al., 1998, Varni, Seid, & Rode, 1999). Riley, Forrest, Starfield, Rebok, Robertson and Green (2004) also argued that children and parents provide different perspectives on child health, and developing parallel parent and child versions of instruments allows researchers to obtain the most accurate picture of child health. The results of Studies 1 and 2 (see Chapter 4, p. 113 & p. 129) showed that when parents and children were using the same measure to rate child QOL their scores were more likely to be related to each other, than when using two different measures to rate child QOL. Therefore we developed a parent report version of our TedQL.4 measure to use in Study 5. The third aim of this study was to explore child-parent agreement on the TedQL.4 measure, and compare this to child-parent agreement when using a similar measure (PedsQLTM4.0).

The TedQL.4 measure differed from the PedsQLTM4.0 in the wording of instructions given at the start of the questionnaires given to parents to rate their child's QOL. On the TedQL.4, parents are asked to: *answer the questions how you think your child would answer them*, compared to the PedsQLTM4.0 where parents are asked: *tell us how much of a problem each one has been for your child.* These instructions at the start of the parent-report TedQL.4 measure were added to improve child-parent agreement, by asking parents to think about how *their child would answer* rather than how *they think their child actually is*. Based on the difference in emphasis of the instructions for parents across the two measures, we predicted that children's and parents' scores would show greater agreement when using the TedQL.4 measure than when using the PedsQLTM4.0 measure.

10.1.4 Summary of the aims and predictions of Study 5

The first aim of Study 5 was to compare the psychometric properties of the child and parent report TedQL.4 measure to an established measure (PedsQLTM4.0). We predicted that the TedQL.4 measure would have 'better' psychometric properties than the PedsQLTM4.0 measure.

The second aim was to investigate the relationship between the TedQL.4 and the PedsQLTM4.0 measures. We predicted that children's and parents' actual TedQL.4 scores would be positively correlated to their PedsQLTM4.0 scores, and their discrepancy TedQL.4 scores would be negatively correlated to their PedsQLTM4.0 scores.

The third aim was to compare agreement between child and parent ratings across the TedQL.4 and PedsQLTM4.0 measures. We predicted that children's and parents' scores would show greater agreement when using the TedQL.4 measure than when using the PedsQLTM4.0 measure.

10.2 Methodology

Sample

Ethics approval was obtained as in Study 1 (see p. 100). One hundred and sixty participants aged 4-8 years were identified from a primary school in Kent. Their parents were given a letter explaining the study. Their parents were asked to complete a permission slip for their child to take part. Eleven children were excluded from the study as their parents did not return the permission slips for their children. Therefore one hundred and forty-nine children (67 females & 82 males) completed the study. The children were taken from three age groups, Year 1 (5.0-6.5 years; n=41), Year 2 (6.0-7.5 years; n=53), and Year 3 (7.0-8.5 years; n=55). The mean age of the Year 1 children was 6.20 years (SD= .29 years). The mean age of the Year 2 children was 7.31 years (SD= 0.33 years). The mean age of the Year 3 children was 8.22 years (SD= 0.29 years). One hundred and forty-five (97%) of the children were Caucasian, three were of Afro-Caribbean origin (2%) and one was of Asian origin (1%).

Questionnaires for parent completion were sent home with all children who had participated in the study (n=149). One hundred and three parents returned their questionnaires giving 103 parent-child dyads in Study 5 (Year 1, n=29; Year 2, n=34; Year 3, n=40).

10.2.1 Child data

Measures

TedQL.4 measure

The children were interviewed using four identical teddy bears as described in Chapter 9, which were referred to as either female or male depending on the sex of the child (see Figure 9.3, p. 266).

As in Study 4, a forced recognition task was used where the bears were first described and then children chose which bear was *most like them*. The children were then probed for whether they were *really* like this or *just a little bit.* Responses were made using a four-point thermometer response scale (see Figure 8.2, p. 221). As in Study 4, the children were also asked a second question for each item. They were asked about what *they would like to be like*, their 'ideal' self. The children were trained in answering the second question, in the same way as Study 4 (see Chapter 8, p. 220). This new version of the measure contained 22 items divided into 5 domains. The full details of the items in the TedQL.4 measure are given in Table 9.21 in Chapter 9 (p. 257).

PedsQLTM4.0 measure

The children also completed the PedsQLTM4.0 measure (Varni et al., 1998). This measure was presented in questionnaire format, where items were read aloud to children, and their responses recorded by the interviewer. The young children version (5-7 years) was used in this study (see Appendix 3 for full measure). This version had been adapted for younger children (Varni et al., 1998, see Chapter 4 p. 101-3 for full details).

This version of the PedsQLTM4.0 measure consisted of 23 items, which divided into 4 domains of functioning. These items and domains are listed in Chapter 4 (see Table 4.2, p.102), as well as the full details of how this measure was developed (see p. 101-3).

Procedure

As with the previous studies, the interviewer spent one 2-hour long session in the classroom in the preceding days (see Chapter 4, p. 103). The children were interviewed individually in a separate room. The measure was administered as described in Chapter 3 (see p.70-1), where the bears used in the TedQL.4 measure were placed opposite the child, and the children were asked to take part in a game.

The children were first asked for their verbal assent, and then shown how to use the response scale for the TedQL.4 measure during a training period. As described in Chapter 8 (see p. 221), children were given one practise item (being good at hopping or not) and one hypothetical question (how they would feel if their favourite toy was lost). All the children responded accurately to these practise items. The children were also taught how to use a 'don't know' option (a blue question mark, see Figure 8.2 p. 221), if they did not know the answer or did not understand the question. The full details of this training procedure are given in Study 4 (see p. 220-1).

The children were then given all the items in the TedQL.4 measure. As with all the previous studies, the four teddy bears were counterbalanced for whether they represented a positive or negative statement (see Chapter 3, p. 74). The children were encouraged to restate their choice following selection (see Chapter 4, p. 104).

The children were given the PedsQLTM4.0 measure (which was administered as directed by Varni et al., 1998). As with previous studies (see Chapter 4, p. 104), the children also chose a sticker at the end of the session.

Scoring

TedQL.4 measure

Children's responses to the items on the TedQL.4 measure were recorded as numerical scores on a response sheet by the interviewer in the same way as Study 4 (see p. 223). Appendix F gives an example section of the response sheet used in this study. The TedQL.4 data was entered into SPSS. Negatively scored items were reversed, so that higher scores represented higher levels of QOL. These scores were used to calculate mean 'actual' QOL (i.e., *how they actually are*) and 'ideal' QOL scores (i.e., *how they would like to be*). The actual and ideal QOL scores ranged from 0 to +3. These were then used to compute discrepancy scores for children for each item (see Chapter 8, p. 223). These discrepancy scores ranged from -3 to +3. As in Study 4, the mean actual QOL and discrepancy scores were reported in the results.

PedsQLTM4.0 measure

Children's responses to the items on the PedsQLTM4.0 measure were recorded as numerical scores on the response sheet by the interviewer in the same way as Study 1 (see p. 105). Appendix 3 gives the response sheet used in this study. The PedsQLTM4.0 data was entered into SPSS. The scores ranged from 0 to +3. As with Study 1 (see p. 105) the scores for the items were transformed to a 0-100 scale, and the same total scores were calculated (total QOL; PH sub-scale; PS sub-scale).

10.2.2 Parent data

Measures

TedQL.4 measure

Parents were asked to complete the proxy version of the TedQL.4 measure in relation to their child. This proxy version of the TedQL.4 was developed specifically for use in this study. The measure was set up in the same way as the child measure, consisting of 23 items to be answered using the thermometer response scale (see Appendix F for full parent questionnaire). This version of the TedQL.4 measure was piloted with 10 parents for ease of administration, understanding, and clarity. Some wording and layout changes were made as a result of this pilot work. Piloting revealed that parents could understand and answer the items, and generally found it straightforward to complete.

Parents were given written instructions on the front page of the questionnaire as follows: "We are interested in what you think your child is like... Please answer the questions how you think your child would answer them by marking the correct place on the thermometer scale. Think about how your child has been in the last week." They were given an example question to ensure they understood the task, and asked to answer all the items in the measure (see Appendix F). As with the children, parents were asked to first rate their child's actual QOL, and then to rate their child's ideal QOL.

PedsQLTM4.0 measure

Parents were also asked to complete the proxy version of the PedsQLTM4.0 measure in relation to their child. As with the children, the young child (5-7 years) version was used. This measure consisted of the same 23 items as the children answered, however, parents answered using a 5-point Likert response scale (see Chapter 4 for full details of items and response scale, p. 102-3, & Appendix 3 for full parent questionnaire).

Procedure

Parents were the TedQL.4 and the PedsQLTM4.0 questionnaires and a letter explaining the study in more detail (see Appendix F for the letter sent to parents). They were asked to complete the questionnaires, and return them either to the school or in the stamped addressed envelopes that were provided.

Scoring

TedQL.4 measure

Parents' recorded their answers to the TedQL.4 items on the response sheet provided. Appendix F shows the response sheet given to parents. The TedQL.4 data was entered into SPSS, and negatively scored items were reversed, in the same way as the child scores above (see p. 278). These scores were used to calculate mean scores for their child's *actual* QOL and discrepancy scores for each item in the measure, in the same way as the children's responses (see p. 278). As with the child data, the actual QOL and discrepancy scores were reported in the results.

PedsQLTM4.0 measure

Parents recorded their answers to the items on the PedsQLTM4.0 measure as numerical scores on the response sheet provided. Appendix B shows the response sheet given to parents. These data were entered into SPSS. As with the child data, the scores for the items transformed to a 0-100 scale, and the same total scores were calculated (total QOL; PH sub-scale; PS sub-scale).

10.2.3 Overall treatment of data and statistical analyses

The distributional properties of the children's and the parent's scores on the TedQL.4 measure and the PedsQLTM4.0 measure were examined. Assessment of skew and kurtosis were made, using the same criteria as previous studies (see Study 1, p. 106-7). Normality testing was also carried out using a Kolmogorov-Smirnov test, as with previous studies (see Study 1, p. 106-7). Where data appeared to be significantly skewed, curved and/or different from a normal distribution, non-parametric statistics and tests were used throughout the analyses (see Study 1, p. 107). Analysis was conducted to assess whether the children's and parents' scores differed systematically by two independent variables (age and gender).

The following analyses were conducted to address the specific hypotheses made, in relation to each of the three aims of this study:

1) Comparing the psychometric properties of the TedQL.4 measure to an established measure (PedsQLTM4.0 measure)

a. The internal reliability (or consistency) of both the children's and the parents' responses were assessed using Cronbach's alpha statistics (α), and compared across the two QOL measures (TedQL.4 and PedsQLTM4.0) (see Study 1, p. 107).

b. The percentage of items with item-total correlations over a standard of .20 was calculated, and compared across the two QOL measures (see Study 4, p. 224-5).

c. The percentage of floor and ceiling effects of both children's and parents' responses were calculated, and compared across the two QOL measures.

d. The reading age (Flesch-Kincaid Grade Level) was calculated for both the child and parent versions, and compared across the two QOL measures.

2) Investigating the relationship between the TedQL.4 and PedsQLTM4.0 measures

a. The correlations between children's *actual* QOL scores on the TedQL.4 and their total QOL scores on the PedsQLTM4.0 were assessed, using Spearman's rank order correlation coefficients (ρ). This was also calculated in the same way for parents' proxy ratings.

b. The correlations between children's *discrepancy* scores on the TedQL.4 and their total QOL scores on the PedsQLTM4.0 were assessed, using Spearman's correlation coefficients (ρ). This was also calculated for parents' proxy ratings.

3) Comparing agreement between child and parent ratings, across the TedQL.4 and PedsQLTM4.0 measures

The agreement between child-rated and parent-rated child QOL was assessed. First, Wilcoxon significance tests were used to assess whether QOL scores differed between children's and parents' ratings. Second, Spearman's correlation coefficients (ρ) were used to assess whether the QOL mean scores were correlated across children's and parents' ratings. This analysis was carried out for both QOL measures (TedQL.4 & PedsQLTM4.0).

10.3 Results

10.3.1 Data screening analyses: child and parent TedQL.4 and PedsQLTM4.0 scores

Item descriptives: means, range and assessment of skew/kurtosis

Children's mean scores on the TedQL.4 measure were skewed towards higher actual QOL and lower discrepancy scores (see Table 10.1). This pattern of skew was consistent across all age groups (i.e., Years 1-3). Parents' scores, when rating their child's QOL using the TedQL.4 measure, showed a similar distribution pattern (see Table 10.2). Both children's and parents' scores also showed fairly high levels of kurtosis, across all age groups (see Tables 10.1 and 10.2). Additionally, Kolmogorov-Smirnov tests showed that Year 2 children's discrepancy scores, and Year 3 actual QOL scores, were significantly different from normal (Year 2: *D*= -0.73, p<.01, Year 3: *D*= 0.82, p<.01, see Appendix F for normal Q-Q plots). As in Study 1, non-parametric tests were used for analyses because of these distributions (see Study 1, p. 107, & see Appendix F for full details of skew and kurtosis calculations).

Table 10.1: Descriptives for children's scores on the TedQL.4 measure

	Child TedQL.4 scores					
Age group	**n**	**Mean (SD)**	**Range**	**Skew**	**Kurtosis**	**Normality test (D)**
Year 1:						
Actual QOL	41	2.28 (0.28)	1.77-2.86	0.46**	-0.09	0.15
Discrepancy	41	0.37 (0.15)	0.05-0.64	-0.05	-0.51	0.17
Year 2:						
Actual QOL	53	1.90 (0.30)	1.18-2.50	-0.27	-0.43	0.14
Discrepancy	53	0.63 (0.26)	0.18-1.23	0.35**	-0.73**	0.17*
Year 3:						
Actual QOL	55	2.02 (0.27)	1.41-2.68	0.37**	0.06	0.16*
Discrepancy	55	0.52 (0.21)	0.00-1.05	0.35**	0.82**	0.12

*Note. Significance level of skew, kurtosis, and normality: ** p< .01; * p< .05*

Table 10.2: Descriptives for parents' scores on the TedQL.4 measure

Age group	n	Mean (SD)	Range	Skew	Kurtosis	Normality test (D)
	Parent TedQL.4 scores					
Year 1:						
Actual QOL	29	2.05 (0.19)	1.64-2.41	0.05	-0.31	0.13
Discrepancy	29	0.64 (0.19)	0.27-1.09	0.33	-0.07	0.11
Year 2:						
Actual QOL	34	2.08 (0.22)	1.64-2.45	-0.49**	-0.38	0.12
Discrepancy	34	0.66 (0.24)	0.18-1.14	0.26	-0.58	0.09
Year 3:						
Actual QOL	40	2.03 (0.26)	1.41-2.45	-0.46**	-0.33	0.12
Discrepancy	40	0.66 (0.22)	0.23-1.05	-0.42**	-1.07**	0.15

*Note. Significance level of skew, kurtosis, and normality: ** p<.01; * p< .05*

These problems with skew, kurtosis, and normality were also found for both child and parent scores on the PedsQLTM4.0 measure, i.e., scores were significantly skewed towards higher QOL reports, and high levels of kurtosis, and not normally distributed. Again this result was broadly consistent across all age groups (see Tables 10.3 and 10.4) (see Appendix F for full details of skew and kurtosis calculations and normal Q-Q plots).

Table 10.3: Descriptives for children's scores on the PedsQLTM4.0 measure

Age group	n	Mean (SD)	Range	Skew	Kurtosis	Normal -ity test (D)
	Child PedsQLTM4.0 scores					
Year 1:						
Total QOL	41	77.60 (15.38)	47.83 - 97.80	-0.17	-1.36**	0.19*
PH score	41	79.88 (14.29)	50.00 - 100.00	-0.22	-0.92**	0.14
PS score	41	76.57 (17.38)	36.67 - 100.00	-0.36	-0.96**	0.17*
Year 2:						
Total QOL	53	67.34 (14.67)	36.96 - 93.48	0.01	-0.48	0.08
PH score	53	71.81 (14.38)	37.50 - 93.75	-0.46**	-0.29	0.20*
PS score	53	64.45 (16.67)	30.00 - 96.67	0.14**	-0.43	0.16*
Year 3:						
Total QOL	55	71.58 (12.18)	45.65 - 93.48	-0.09	-0.82**	0.11
PH score	55	78.18 (12.42)	50.00 - 100.00	-0.40	-0.24	0.19*
PS score	55	67.52 (13.41)	36.67- 93.33	0.04	-0.77**	0.13

*Note. Significance level of skew, kurtosis, and normality: ** p<.01; * p<.05*

Table 10.4: Descriptives for parents' scores on the PedsQLTM4.0 measure

Age group	n	Mean (SD)	Range	Skew	Kurtosis	Normality test (D)
	Parent PedsQLTM4.0 scores					
Year 1:						
Total QOL	29	80.75 (11.16)	55.43 - 96.74	-0.75**	-0.35	0.19*
PH score	29	84.59 (12.04)	46.88 – 100.00	-1.30**	2.09**	0.26*
PS score	29	78.69 (11.97)	51.67 – 96.67	-0.75**	-0.22	0.17
Year 2:						
Total QOL	34	80.34 (11.59)	51.09 – 97.83	-0.66**	-0.17	0.15
PH score	34	87.41 (11.08)	59.38 – 100.00	-0.84**	0.40	0.17*
PS score	34	76.57 (12.74)	46.67 – 98.33	-0.54**	-0.44	0.14
Year 3:						
Total QOL	40	79.10 (12.47)	54.35 – 100.00	-0.20	-0.56	0.11
PH score	40	86.09 (11.37)	59.38 – 100.00	-0.76**	-0.47	0.17*
PS score	40	75.43 (14.05)	43.33 – 100.00	-0.19	-0.46	0.11

*Note. Significance level of skew, kurtosis, and normality: ** p<.01; * p<.05*

Potential item bias: age

Spearman's correlation coefficients (ρ) revealed that age was not correlated with children or parents' actual QOL and discrepancy TedQL.4 or total PedsQLTM4.0 scores (see Tables 10.5 and 10.6).

Table 10.5: Relationship between chronological age and mean scores on the TedQL.4 (actual and discrepancy) for child and parent reports

Age group	Correlation coefficient (ρ)					
	Year 1		Year 2		Year 3	
	n	Age	n	Age	n	Age
Child:						
Actual QOL	41	-0.19	53	0.05	55	-0.21
Discrepancy	41	0.24	53	-0.09	55	-0.02
Parent:						
Actual QOL	29	-0.03	34	0.22	40	0.01
Discrepancy	29	0.19	34	-0.32	40	0.06

Table 10.6: Relationship between chronological age and mean scores on the PedsQLTM4.0 (total QOL) for child and parent reports

Age group	Correlation coefficient (ρ)					
	Year 1		Year 2		Year 3	
	n	Age	n	Age	n	Age
Child:						
Total QOL	41	-0.26	53	0.12	55	-0.13
Parent:						
Total QOL	29	0.11	34	0.03	40	-0.01

Potential item bias: gender

A Kruskall-Wallis H test revealed that there was no effect of gender on child- or parent-rated QOL using either of the two measures (TedQL.4 or PedsQLTM4.0), with the exception of child discrepancy TedQL.4 scores (see Table 10.7).

Table 10.7: Effect of gender on scores on the TedQL.4 (actual and discrepancy) and PedsQLTM4.0 (total QOL) for child and parent reports

Gender of child	Male		Female	
Measure	n	Median (Mean)	n	Median (Mean)
Child:				
TedQL measure				
Actual QOL	82	2.05 (2.04)	67	2.00 (2.05)
Discrepancy	82	0.45 (0.48)	67	0.55 ** (0.58)
PedsQL measure				
Total QOL	82	71.74 (72.46)	67	69.57 (70.89)
Parent:				
TedQL measure				
Actual QOL	56	2.05 (2.04)	47	2.09 (2.07)
Discrepancy	56	0.64 (0.67)	47	0.64 (0.64)
PedsQL measure				
Total QOL	56	81.52 (80.90)	47	80.43 (78.91)

*Note. ** p<.01 level (Means are reported in brackets for comparison)*

10.3.2 Comparing the psychometric properties of TedQL.4 measure to established measure (PedsQLTM4.0 measure)

Internal reliability (consistency)

The internal consistency of children's and parents' responses on the TedQL.4 and on the PedsQLTM4.0 were calculated using Cronbach's alpha statistics. None of the children's or the parents' responses produced consistency values above the .70 standard for either actual QOL or discrepancy items, with the exception of Year 3 parents' actual QOL scores (see Table 10.8).

Table 10.8: Internal consistency (α) of children and parents' responses on the TedQL.4 measure

	Child TedQL.4			Parent TedQL.4		
	n	No. of items	α	n	No. of items	α
Age group						
All ages:						
Actual QOL scale	149	22	0.64	103	22	0.64
Discrepancy scale	149	22	0.52	103	22	0.59
Year 1:						
Actual QOL scale	41	22	0.54	29	22	0.51
Discrepancy scale	41	22	0.50	29	22	0.53
Year 2:						
Actual QOL scale	53	22	0.56	34	22	0.60
Discrepancy scale	53	22	0.52	34	22	0.68
Year 3:						
Actual QOL scale	55	22	0.52	40	22	0.72
Discrepancy scale	55	22	0.44	40	22	0.56

Across all the ages, the children's responses on the PedsQLTM4.0 measure produced consistency values above the .70 standard for their total QOL scores and their psychosocial summary scores (see Table 10.9). The parents' responses, across all age groups, also produced consistency values above this standard for their total QOL, psychosocial and physical health summary scores (see Table 10.9). These internal consistency values were comparable to those reported by Varni et al. (2002).

Table 10.9: Internal consistency (α) of children and parents' responses on the PedsQLTM4.0 measure

	Child PedsQLTM4.0			Parent PedsQLTM4.0		
Age group	**n**	**No. of items**	**α**	**n**	**No. of items**	**α**
All ages:						
Total QOL	149	23	0.81	103	23	0.91
PH sub-scale	149	8	0.46	103	8	0.73
PS sub-scale	149	15	0.76	103	15	0.89
Year 1:						
Total QOL	41	23	0.84	29	23	0.90
PH sub-scale	41	8	0.45	29	8	0.72
PS sub-scale	41	15	0.81	29	15	0.88
Year 2:						
Total QOL	53	23	0.80	34	23	0.91
PH sub-scale	53	8	0.41	34	8	0.71
PS sub-scale	53	15	0.75	34	15	0.88
Year 3:						
Total QOL	55	23	0.76	40	23	0.92
PH sub-scale	55	8	0.43	40	8	0.75
PS sub-scale	55	15	0.67	40	15	0.90

Item-total correlations

As in Study 4, the correlation of each item in the TedQL.4 measure with the total score was calculated to assess whether items were contributing reliability to the measure (see p. 240-1). These item-total correlations were calculated for both children's and parents' responses, and also for the PedsQLTM4.0 measure (see Appendix F for tables of all item-total correlations).

For both measures, across both child- and parent-report, 71% or more of the items were correlated to the total scores above the .20 standard (see Table 10.10, and see Chapter 8 for a discussion of this standard, p. 224). The PedsQLTM4.0 measure showed a greater amount of items with correlations above the standard compared to the TedQL.4 measure (see Table 10.10).

Table 10.10: Percentage of item-total correlations falling above the .20 standard across the TedQL.4 and PedsQLTM4.0 measures

	% of item-total correlation above .20			
	TedQL.4		PedsQLTM4.0	
Age group	Child	Parent	Child	Parent
Year 1	78	73	87	100
Year 2	71	75	92	100
Year 3	75	87	96	100

Range of measurement

Ceiling and floor effects for both measures, across both child- and parent-report, were calculated. Ceiling effects are the percentage of children and parents who endorse the highest anchor point for each item. Similarly floor effects are the number of endorsements made at the lowest anchor point. For example 1% of parents of Year 3 children reported floor levels on the discrepancy scale of the TedQL.4 (i.e., they reported no discrepancies for any of the TedQL.4 items).

On both measures, children and their parents produced a low number of ceiling effects and no floor effects (see Tables 10.11 and 10.12). The TedQL.4 measure had the least number of ceiling effects, for both child- and parent-report, compared to the PedsQLTM4.0 measure (see Tables 10.11 and 10.12).

Table 10.11: Percentage of ceiling and floor effects on the TedQL.4 measure, for children and parents

	Child TedQL.4		Parent TedQL.4	
	% ceiling	% floor	% ceiling	% floor
Age group				
Year 1:				
Actual QOL	7	0	0	0
Discrepancy	0	0	0	0
Year 2:				
Actual QOL	0	0	0	0
Discrepancy	0	0	0	0
Year 3:				
Actual QOL	0	0	0	0
Discrepancy	0	0	1	0

Table 10.12: Percentage of ceiling and floor effects on the PedsQLTM4.0 measure, for both children and parents

	Child PedsQLTM4.0		Parent PedsQLTM4.0	
	% ceiling	% floor	% ceiling	% floor
Age group				
Year 1:				
Total QOL	5	0	2	0
PH score	12	0	2	0
PS score	5	0	2	0
Year 2:				
Total QOL	0	0	2	0
PH score	0	0	15	0
PS score	2	0	2	0
Year 3:				
Total QOL	0	0	4	0
PH score	6	0	13	0
PS score	0	0	4	0

Readability statistics

Readability statistics were calculated using Flesch's readability formula to calculate the reading age (Flesch-Kincaid Grade Level) required to understand the items in both of the measures. The TedQL.4 measure had a lower grade level than the PedsQLTM4.0 measure, for both child- and parent-report versions (see Table 10.13).

Table 10.13: Reading age required for child and parent version of the TedQL.4 and PedsQLTM4.0 measures

	Flesch-Kincaid Grade Level (equivalent age range)	
Measure	**Child version**	**Parent version**
TedQL.4	0.7 (5 yrs)	1.7 (7 yrs)
PedsQLTM4.0	1.1 (6 yrs)	4.3 (9 yrs)

10.3.3 Exploring the relationship between the TedQL.4 and PedsQLTM4.0 measures

The relationship between children's and parents' scores on the TedQL.4 measure and their scores on the PedsQLTM4.0 measure was assessed using Spearman's correlation coefficients. Both children's and parents' actual QOL scores on the TedQL.4 were positively correlated with their total QOL scores on the PedsQLTM4.0, across all age groups (see Table 10.14). Both children's and parents' discrepancy scores on the TedQL.4 were negatively correlated with their total QOL scores on the PedsQLTM4.0 (see Table 10.14). This result was consistent across all age groups with the exception of parents' ratings for Year 1 children.

Table 10.14: Relationship between children and parents' ratings on the TedQL.4 (Actual QOL and Discrepancy) and PedsQLTM4.0 (Total QOL) measures

	Actual QOL to Total QOL				Discrepancy to Total QOL			
Age group	Child		Parent		Child		Parent	
	n	ρ	n	ρ	n	ρ	n	ρ
All ages	149	0.53 **	103	0.55 **	149	-0.58 **	103	-0.45 **
Year 1	41	0.46 **	29	0.45 *	41	-0.49 **	29	-0.37
Year 2	53	0.65 **	34	0.51 *	53	-0.68 **	34	-0.37 *
Year 3	55	0.42 **	40	0.61 **	55	-0.52 **	40	-0.48 *

*Note. ** p<.01; * p<.05 level*

10.3.4 Comparing agreement between child and parent ratings, across the TedQL.4 and PedsQLTM4.0 measures

The agreement between child and parent ratings of child QOL was compared across the two measures, using Wilcoxon median tests and Spearman's correlation coefficients (ρ).

Wilcoxon tests revealed significant differences between the child and parent actual QOL scores on the TedQL.4 for the Year 1 and Year 2 age group. In the Year 1 age group, children rated their QOL higher than their parents, and in the Year 2 age group parents reported higher child QOL than the children themselves (see Table 10.15). Children's and parents' discrepancy scores also differed significantly for the Year 1 and Year 3 age groups, with parents reporting higher discrepancies for their children than the children themselves (see Table 10.15).

Table 10.15: Differences between children and parents' scores on the TedQL.4 measure (Actual QOL and Discrepancy)

		Actual QOL		**Discrepancy**	
		Median (Mean)		**Median (Mean)**	
Age group	**n**	**Child**	**Parent**	**Child**	**Parent**
All ages	103	2.00 (2.04)	2.07 (2.05)	0.50 (0.52)	0.64 *** (0.65)
Year 1	29	2.30 (2.28)	2.05 ** (2.05)	0.36 (0.37)	0.64 *** (0.64)
Year 2	34	1.95 (1.90)	2.09 *** (2.07)	0.59 (0.63)	0.64 (0.66)
Year 3	40	2.00 (2.02)	2.07 (2.03)	0.50 (0.52)	0.66 ** (0.66)

*Note. *** p<0.001; ** p<0.01 level (Means are reported in brackets for comparison)*

For the PedsQLTM4.0 measure, Wilcoxon tests revealed significant differences between the child- and parent-rated total QOL across the Year 2 and Year 3 age groups (with parents' reporting higher child QOL than the children themselves, see Table 10.16).

Table 10.16:Differences between children and parents' scores on the PedsQLTM4.0 measure (Total QOL)

		Total QOL (PedsQL TM4.0)	
		Median (Mean)	
Age group	n	Child	Parent
All ages	103	71.74 (71.77)	80.43 *** (79.97)
Year 1	29	78.26 (77.60)	84.24 (80.75)
Year 2	34	67.39 (67.34)	80.43 *** (80.34)
Year 3	40	71.74 (71.58)	79.35 ** (79.10)

*Note. *** p<.001; ** p<.01; (Means are reported in brackets for comparison)*

Additionally, the correlations between child- and parent-rated child QOL were compared across the two measures (i.e., child-parent agreement). The correlations between children's and parents' child TedQL.4 scores were generally low (see Table 10.17). Children's and parents' ratings of child actual QOL were significantly correlated only within the Year 2 age group. Children's and parents' child discrepancy scores were significantly correlated only within the Year 3 age group (see Table 10.17). The agreement between children's and parents' ratings of child QOL was also low when using the PedsQLTM4.0 measure, with no significant correlations for any age group (see Table 10.17).

Table 10.17: Agreement between child- and parent-rated QOL, using the TedQL.4 and PedsQLTM4.0 measures

		Correlation coefficients (ρ)		
		TedQL.4: child to parent		PedsQLTM4.0: child to parent
Age group	n	Actual QOL	Discrepancy	Total QOL
All ages	103	0.15	0.23*	0.15
Year 1	29	-0.22	-0.11	0.01
Year 2	34	0.38*	0.31	0.17
Year 3	40	0.20	0.36*	0.27

*Note. * p<.05*

10.4 Discussion

Following the alterations and changes made to the TedQL.4 measure as a result of Study 4, we felt it was necessary to report the psychometric validation of the new child version, and also of the new parent report version. The first aim of this study was to compare the psychometric properties of the new TedQL.4 measure to an established measure (the PedsQLTM4.0). The PedsQLTM4.0 measure was developed in different ways to the TedQL.4, and differed in presentation style, item content, and response scale type.

We predicted that the TedQL.4 would be the most appropriate measure for children below eight years, and this would be reflected by this new measure producing 'better' psychometric properties in comparison to the PedsQLTM4.0. We made four predictions regarding the psychometric properties of these two measures.

Our first prediction, that the internal reliability of the children's and parents' TedQL.4 responses would be higher than their PedsQLTM4.0, was not supported. Both children's and parents' responses were found to be more reliable when using the PedsQLTM4.0 (total QOL scores: children: α= .76 - .84; parents: α= .90 - .92, see Table 10.9, p. 285), compared when using the TedQL.4 (actual QOL and discrepancy scores: children: α= .44 - .64; parents: α= .51 - .72, see Table 10.8, p. 284).

Second, we predicted that children's and parents' responses to the TedQL.4 items would show a higher number of item-total correlations above the .20 standard (compared to their responses to the PedsQLTM4.0 items). This prediction was also not supported. The percentage of item-total correlations above the .20 standard was higher for the PedsQLTM4.0 items (ranging from 87% to 100%), than for the TedQL.4 items (ranging from 71% to 87%, see Table 10.10, p. 286).

Our third prediction, that children and parents using the TedQL.4 measure would produce fewer floor and ceiling responses (compared to their responses on the PedsQLTM4.0), was partly confirmed. Although children's and parents' responses on the TedQL.4 and PedsQLTM4.0 produced no floor effects, the TedQL.4 measure had the lower number of ceiling effects, for both child and parent responses (see Tables 10.11 & 10.12, p. 287).

Fourth, we predicted that the reading age required for the TedQL.4 items would be lower than that required for the PedsQLTM4.0 items. This prediction was supported. The TedQL.4 measure had a lower grade level for both child and parent report versions (child: 0.7; parent: 1.7), compared to the PedsQLTM4.0 (child: 1.1; parent: 4.3, see Table 10.13, p. 288).

Overall contrary to predictions, we found that the PedsQLTM4.0 measure had the 'best' psychometric properties, in relation to the internal reliability of the children's and parents' responses and the item-total correlations. Although the TedQL.4 was found to produce fewer floor and ceiling effects and also to have lower reading age for comprehension of the items, the PedsQLTM4.0 measure still produced a reasonably small number of floor and ceiling effects and had an acceptable reading age level.

However, it should be noted that the TedQL.4 measure did still produce reasonable psychometric properties for both child and parent report versions, to the extent that it could be useful in child research. One possible explanation as to why the PedsQLTM4.0 measure gained higher levels of internal consistency may be that the PedsQLTM4.0 items are all closely related to each other and focused on a narrower set of issues compared with the TedQL.4 items. If items were too varied, this would serve to lower the internal reliability of a measure (which is based on the homogeneity of the items to each other). Therefore it may be that the content of the TedQL.4 items require further development and alteration, to produce a set of items which hold together better as a total scale. A tenet in measurement theory predicts that adding more items of parallel content increases the internal reliability of a given measure (Crocker & Algina, 1986). Indeed Peter and Churchill (1986) in their meta-analysis of rating scales found that the number of items in a scale were positively related to the internal reliability of that scale.

Anecdotally during the data collection all the children responded well to the presentation style of the TedQL.4 measure (i.e., the use of teddy bears as props and a graphic thermometer response scale) and frequently requested to "play the teddy bear game again". As discussed in Chapter 4 (p. 132) considering the PedsQLTM4.0 measure has been developed over more than fifteen years, it was encouraging that our relatively new measure stood up so well in comparison.

As in Studies 1 and 2 (see Chapter 4, p. 116 & p. 131), we needed to compare the psychometric properties obtained in Study 5 for the PedsQLTM4.0 measure to existing published studies using this measure. First, the internal consistency of the children's responses to the PedsQLTM4.0 items was lower than the values reported in previous published studies using this measure. The alpha values reported by Varni et al. (2001) are shown in brackets against the values we found in Study 5: total QOL: $\alpha = .81$ (.88); PH sub-scale: $\alpha = .46$ (.80); PS sub-scale: $\alpha = .76$ (.83). The internal consistency of parent's responses to the PedsQLTM4.0 items was also lower than the values reported by Varni et al. (2001): total QOL: $\alpha = .90$ (.90); PH sub-scale: $\alpha = .73$ (.88); PS sub-scale: $\alpha = .89$ (.86). (See Table 10.9, p. 285 for our alpha values). As discussed in Chapter 4 (see p. 116), we were unsure why our alpha values were lower than those reported by Varni et al. (2001). Again we were using a different sample to Varni et al. (2001, i.e., we were working in the U.K. as opposed to the U.S.).

Second, the item-total correlation values for the PedsQLTM4.0 found in Study 5 were comparable to those reported in published studies using this measure. Varni et al. (2001) reported that 19 out of 23 items (83%) of the item-total correlations were above the .20 standard for both child and proxy report. In Study 5 the percentage of item-total correlations for child report above this standard ranged from 87% to 96%, and for proxy report all of the items (100%) were above the standard (see Table 10.10, p. 286).

Third, the floor effects of both children and parent's responses on the PedsQLTM4.0 were comparable to the values reported by Varni et al. (2001). The floor percentages reported by Varni et al. (2001) are shown in brackets against the percentages found in Study 5: total QOL: 0% for all ages (0%); PH sub-scale 0% (0%); PS sub-scale 0% (0%). However the ceiling effects of both children and parent's responses were lower than those reported by Varni et al. (2001): total QOL: 0%-5% (child: 7.2%, parent 10.3%); PH sub-scale 0%-15% (child 25.8%, parent 39.6%); PS sub-scale 0%-5% (child 12.0%, parent 13.8%, see Table 10.11 & 10.12, p. 287 for our values).

The second aim of this study was to explore the relationship between the TedQL.4 and PedsQLTM4.0 measures. This aim was achieved by investigating whether children's and parents' TedQL.4 scores were correlated with their PedsQLTM4.0 scores.

Our first prediction, that children's and parents' actual TedQL.4 scores would be positively correlated to their scores on the PedsQLTM4.0, was confirmed. Both children's and parents' actual TedQL.4 scores were positively correlated with their total QOL scores on the PedsQLTM4.0 (ρ= .42 to .65), and these correlations were significant across all the age groups (see Table 10.14, p. 289).

Second, we predicted that children's and parents' discrepancy TedQL.4 scores would be negatively correlated with their scores on the PedsQLTM4.0. This prediction was also supported. Both children's and parents' discrepancy TedQL.4 scores were negatively correlated with their total PedsQLTM4.0 scores (ρ= -.37 to -.68, see Table 10.14, p. 289). Again these correlations were significant across all age groups, with the exception of parent's ratings for Year 1 children.

The results showed that our TedQL.4 measure was related to an established measure of QOL (PedsQLTM4.0), and therefore we can hypothesise that both instruments were measuring a similar construct (i.e., QOL). This result is consistent with our findings in Study 1 (see Chapter 4, p. 112) where we found children's TedQL.1 scores were positively correlated to their PedsQLTM4.0 scores.

The third aim of this study was to compare child-parent agreement across the TedQL.4 and PedsQLTM4.0 measures. These two measures differed in the way that parents were required to answer the items about their child, in that the TedQL.4 asks parents to rate items *how they think their children would answer* whereas the PedsQLTM4.0 asks to rate items *how they think their children actually are.* Based on this difference, we predicted that the level of child-parent agreement would be higher for the TedQL.4 compared to the PedsQLTM4.0.

Our prediction was partly confirmed. Agreement between children and parents' scores was low when rating child QOL on the PedsQLTM4.0 measure. Children's and parents'

PedsQLTM4.0 scores were not significantly correlated at any age group (see Table 10.17, p. 292). Although the agreement between children and parents' scores was also low on the TedQL.4 measure, child-parent ratings were significantly correlated in the Year 2 age group for actual QOL scores ($\rho = .38$, $p < .05$) and for all the children combined ($\rho = .23$, $p < .05$), and in the Year 3 age group ($\rho = .36$, $p < .05$) for discrepancy scores (see p. 292). These results could be due to the differences in the instructions given to parents in how to answer items (i.e., making them rate items *how they think their child would answer* as they did in the TedQL.4 could help raise the level of agreement between child and parent reports).

Median testing revealed that parents generally reported higher discrepancies for their children than the children themselves on the TedQL.4 (see Table 10.15, p. 290). In addition parents generally reported higher QOL for their children than the children themselves on the PedsQLTM4.0 (see Table 10.16, p. 291). Our results illustrated the point that was discussed in the introduction, that parents can both over- and under-estimate their children's QOL (Bruil, 1999, Theunissen et al., 1998, Vance, Morse, Jenney & Eiser, 2001). Parents can over-estimate the effect of their children's problems (i.e., higher discrepancy scores on TedQL.4), but on the other hand they can just as easily over-estimate their child's overall functioning (i.e., higher QOL scores on PedsQLTM4.0).

This study investigated the level of agreement between children's and parents' ratings of child QOL. Generally the amount of correlation between children's and parents' scores was low. Indeed a number of researchers have shown that the level of agreement between child- and parent-reports may not always be as high as has been originally thought (e.g. Langeveld et al., 1997, Vogels et al., 1998). The question still remains as to why children and their parents have such different views to each other. Researchers have put forward various arguments to help explain this lack of concordance. For example, Vance et al. (2001) have argued that parents and children may not agree on what is 'normal' functioning, or what is important to be good or bad at. These differing perspectives could mean that parents mis-judge the relative importance of issues for their child. Further work could be focused on investigating the variables that could influence the level of agreement between children and parents' reports. For example, the gender of children could effect how well parents can judge their child's QOL (Le Coq et al., 2000, Rutishauser, Sawyer, & Bowes, 1998). There is some evidence that parents are more likely to discuss emotional issues with daughters than sons (Fivush et al., 1991). Parent's ability to rate their children's lives accurately may also be influenced by their own mental health (e.g. Tarullo, Richardson, Radke-Yarrow, & Martinez, 1995). Future studies could be designed to assess the extent to which such external variables can influence the agreement between children's and parents' ratings of child QOL.

Chapter 11: General Discussion

11.1 Overview

The aim of this thesis was to develop a child-centred self-report QOL measure for children below eight years. In the process of developing this child measure, two questions emerged that were central to the issue of gaining information from young children. These questions were:

i. Can children below eight years self-report on their thoughts, feelings, and lives?

ii. If yes, what are the best ways to gain self-reports from young children?

In Chapters 1 and 2 we discussed the evidence that young children are capable of self-reporting on their lives (see p. 4-5 & p. 20-30), and highlighted the reasons for gaining information from children themselves (see p. 4-7). Our review of child self-report measures in Chapter 3 enabled comparison of the response scales and presentation styles that have been used in existing child measures (see p. 46-51). Chapter 3 also compared the quality of existing child measures across scale and presentation type (i.e. the item generation methods, reliability, validity, and responsiveness data reported, see p. 53-64). Based on further reviews of the literature (Chapters 6 & 7) and the results of the studies in this thesis (Chapters 4, 5, 8, 9 & 10) we produced a set of guidelines for researchers that can be applied when developing a self-report measure for children below eight years.

11.2 Implications for developing child self-report measures

We have divided these recommendations into three sections, relating to the content of measures, the way items are presented, and the response scale type used to answer items.

Content of measures

Our review of self-report measures in Chapter 3 showed that the most common method for generating items was to use information from children themselves to inform the content (see p. 53-4). As discussed in Chapter 5 (see p. 136-7), using information from children ensures that content of measures is derived from the population for which the tool is to be used. We recommend that qualitative methods (such as interviews or focus groups) be used by authors in the development of child measures, to ensure the content of items are child-generated.

Based on our review of the developmental literature and theories relating to children's cognitive development in Chapter 2, we made specific conclusions on the concepts that children below eight years could understand. These conclusions have implications for the content development of QOL measures aimed at children of this age. In relation to emotions, we concluded that children below eight years are capable of answering self-report items asking about feelings such as being happy, sad, cross, or angry (see p. 20-1). We found evidence that children's understanding of, and ability to think about, their mental states increases significantly during and after four years of age, and therefore we argued that including items about psychological and cognitive functioning in QOL measures would be viable with four year old children and above (see p. 23-4).

In relation to young children's self-concepts, we found evidence that children can hold both negative and positive self-concepts by four years of age, and that they can also understand different aspects of self (see p. 24-6). Based on this evidence, we argued that self-report measures could include items asking children as young as four years about what sort of person they were (e.g., whether they are good or bad at activities, or whether they have a lot or a few friends). In relation to health, the literature reviewed in Chapter 2 showed that by 3-4 years old children understand the physical aspects of illness and can appreciate some of the biological causes (see p. 28-30). Therefore we argued that QOL measures to be used with children below eight years could include simple items asking children about their everyday health.

Presentation style of measures

Our review of self-report measures in Chapter 3 showed that the most common way to present items to children was to read them aloud (see p. 49). The measures reviewed in Chapter 3 also highlighted the potential value of using pictures or props as visual aids when presenting items to young children (see 49-50). In Chapter 6 we considered the value of using pictures or props for QOL measures, and concluded that props can help produce more meaningful and accurate responses from children on self-report items (see p. 179-81). Props can help clarify and concretise items for children and avoid reliance on language skills. Props enable children to enact their responses, and also allow children to respond non-verbally to items if they prefer. Props can help reduce task demands on children by serving as memory aids (as they remain present while children answer items).

The literature reviewed in Chapter 6 also considered the techniques that could be used to enhance young children's abilities to respond to self-report measures. We argued that establishing rapport beforehand with children is essential when working with young ages (see p. 177-8). Researchers need to bridge the gap between themselves as an adult researcher and the child they are interviewing. Although children as young as three years understand the basic maxims of conversations, they may become confused when researchers set aside these rules during interviews. We argued that researchers should avoid repeated questioning in their measures, as this may cause children to change their second answer due to mis-understanding (see p. 182).

Readability formulas can be used to establish the reading age required for understanding wording in items (see p. 178-9). Researchers should also use clear, and unambiguous language to make the wording of items appropriate for young children (see p. 181-2). Researchers such as Chapman and Tunmer (1995) and Marsh et al. (1998) argued that authors should avoid using 'I' statements in measures, and use 'Are you?' questions instead. This relates to evidence that young children's ability to answer questions develops earlier than their ability to verify statements (Akiyama and Guillory, 1983, see p. 174-5). We argued that the use of specific recall time periods for measures (such as one to two weeks) would help children when answering self-report items (by giving them a specific time to focus on for recall of behaviour, thoughts, and feelings, see p. 182-3).

Response format for measures

Our review in Chapter 3 showed that Likert response scales were most commonly employed by researchers to represent response choices to children (see p. 46). However graphic, facial expression, and visual analogue scales have also been used successfully with young children in self-report measures (see p. 46-7). We argued that authors need justification for their response scale choice. The most common way to gain evidence for a chosen response scale was to compare the psychometric properties of a measure across different scale types (see p. 47). As we discussed in Chapter 7, we argued that such assessments involving comparison of the reliability and reproducibility of children's responses across different response scales could provide evidence as to the most appropriate scale for child self-report measures (see p. 199-201).

The literature reviewed in Chapter 7 summarised the suggestions that have been made to help children understand and use response scales in self-report measures. Children may not understand the words used as anchors in response scales, and this can impact on the accuracy of their responses to items (see p. 197-8). Researchers have suggested using specific anchors such as *very true/not true of me* rather than vague quantifiers such as *often* (see p. 197). We noted that testing children's use and understanding of anchors used in measures can be helpful to avoid problems of mis-interpretation (see p. 198).

Some researchers have suggested using graphic response scales to help children in the rating task (e.g., Champion et al., 1997, Dockrell et al., 2000, see p. 198-9). Creative ways to represent scales to young children have included using a rolled/unrolled toothpaste tube over the scale, or using a toy koala on a wooden pole to rate items. It is also relevant to consider whether children can understand, recognise, and rate frequencies. Children's understanding and use of response scales can be tested using either a calibration task (such as arranging different sized circles or blocks in the correct size order, see p. 201-2) or hypothetical questions (where researchers can reasonably assume where the correct answer should be, see p. 202). We also argued that training periods (involving practice questions) can be used to help children understand the rating task and to become familiar with the response scale used (see p. 202-3).

11.3 Evaluation of our QOL measure (the TedQL)

We followed the above criteria when developing our child QOL self-report measure (the TedQL). In the sections below we discuss the extent to which we met the aims for our instrument, and the extent to which we incorporated the recommendations from the literature in the development of the TedQL measure. As highlighted in Chapter 1 (see p. 10), we aimed to develop an instrument that would 1) use information from children themselves to inform the content of items, 2) provide an alternative presentation style to written measures, and 3) establish the most appropriate response format for young children.

Developing the content of the TedQL

The specific content of the items in the TedQL.1 measure was guided by a review of the literature (see Chapter 2, p. 20-30), from measures currently available for young children (see Chapter 3, p. 71-3), and previous experience with children. Ten items were selected for the TedQL.1 which were thought to best represent children's lives and concerns in a general way (see Chapter 4, p. 100). The initial TedQL.1 version consisted of 10 items within 5 areas (see Table 4.1, p. 101). As the children's reports on the TedQL.1 measure produced a low value of consistency in Study 1 with an alpha value below the .70 standard ($\alpha = .35$, see Table 4.5, p. 111), we expanded the number of items in our measure for Study 2 (from 10 to 23 items, see Table 4.10, p.121). These additional items were based on comments children had spontaneously volunteered in Study 1 (see p. 115). Therefore a new version of the TedQL measure was developed for use in Study 2 (the TedQL.2, see Table 4.10, p. 121). We referred back to our review of the literature concerning the concepts children below eight years can understand (Chapter 2, p. 20-30) to ensure that the content of our items did not include language, wording, or concepts beyond their understanding.

Our review in Chapter 3 showed that the most common method for developing the content of child self-report measures was to use information from children themselves to generate/justify items (see p. 53-4). Therefore we designed Study 3 to interview healthy children aged three to eight years about their thoughts and feelings in relation to their abilities, behaviour, and their school and family lives. We used the interview data to directly inform the content development of items in the TedQL.2 measure, using content analysis to identify the areas most important in the children's lives. The data from Study 3 built on the previous studies by justifying and providing evidence for the continued inclusion of any items within the TedQL.3, the removal of items that were not relevant to young children, and the addition of other items that were important in young children's lives. Based on the results of Study 3, six items were deleted, 17 were retained (with 13 of these being altered slightly), and 13 additional items were added to the TedQL.2 measure (see Chapter 5, p 162). A new version of the measure was developed which contained 30 items (see Table 5.16, p. 163). Therefore the content of the TedQL measure was based on information gained from children themselves about their lives (Study 3, Chapter 5), and also guided by our review of the developmental literature (Chapter 2). As a result, the content of the TedQL items was empirically based and justifiable.

In Chapter 8, we argued that researchers need more information on how children interpret and answer items in self-report measures (see p. 211-12). Therefore we designed Study 4 to examine the processes children engaged in when answering the TedQL.4 items (i.e., the strategies they used to rate their behaviour and functioning). We used a combination of the TAM and follow-up probes to gain information from children while they were completing a selection of ten items (see p. 222-3). We found that the majority of children used concrete examples of specific situations or instances as reasons for their response choices to items, regardless of item type (see p. 262-3). We showed that children did understand the items they were answering, and that young children tended to refer to concrete examples when attempting to rate their behaviour or functioning.

Developing the presentation style for the TedQL

The way the items were presented to children in all versions of the TedQL measure was based on our review of the literature (see Chapter 3, & Chapter 6). The review of self-report measures in Chapter 3 showed that the majority of authors read items aloud to children (see p. 49). Our review also highlighted the potential value of using pictures or props as visual aids when administering items to young children (see p. 49-50). In Chapter 6, we considered the relative value of pictures and props, and concluded that props would be most helpful for administering self-report items to children below eight years compared to pictures (see p. 179-81).

Our measure was based on two existing child self-concept/perception measures – the PSPCSA (Harter & Pike, 1984) and the BPI (Measelle et al., 1998, see p. 70). The TedQL measure was administered as a game where items were read aloud to children, with the additional use of teddy bears as props to help children understand the items (see Chapter 3, p 69-71). We chose the presentation style (verbal and props) on the basis of our review of the literature (Chapter 3 & Chapter 6). Our review in Chapter 3 revealed that no current QOL measures had used props to present items to children (see p. 51), and we argued that this method could augment children's responses to items (see p. 69-71).

In Chapter 6, we discussed various techniques that could be used to help young children respond to self-report measures (see p. 177-83). We aimed to take these recommendations into account when developing our TedQL measure. In all versions of the TedQL, we spent between one and three hours beforehand working with children to establish rapport with them before administering the measure (see p. 73). We also avoided repeated questioning in our interview task, and where we needed to give children the TedQL measure twice (Study 4, see Chapter 8, p. 210-11) we used two time points separated by a week to avoid recall bias (see p. 223).

The interview data from Study 3 was used to inform the language and wording used in the TedQL items to ensure children could understand items (see Chapter 5, p. 162-3). We also phrased the items as 'Are you' questions rather than 'I' statements to avoid mis-understanding by the children (see Chapter 6, p. 174-5). We assessed the reading level required for our measure in Study 5, and found that the reading age required for items was five years (see Table 10.13, p. 291). This could mean that we need to adjust items for the youngest children (3-4 years), or add additional prompts to ensure comprehension. We used a recall period of one week in all versions of the TedQL measure (see p. 71), to aid children in recalling and rating their behaviours and feelings for items.

Establishing an appropriate response scale for the TedQL

Our review of self-report measures for children below eight years in Chapter 3 showed that Likert scales were the scale type most commonly used in previous instruments (see p. 46), but also revealed that graphic, facial expression, and visual analogue scales had been used successfully in other child measures. In Chapter 3, we argued that researchers should provide evidence for their choice of scale type for their measure, due to the variation of scale types available for use in child self-report measures (see p. 47 & p. 65). Few researchers provided evidence for their scale choice, however for those that did the most common method used was to compare the relative value of different scales for their instrument (using psychometric properties of measures, e.g., internal reliability, see p. 47). We assessed the impact of two different scale types on children's TedQL.2 responses in Study 2 (see p. 118). We found that scale type impacted on children's responses – in that children who used a circles scale produced responses with a higher level of internal reliability and completed the measure faster than children who used a linear scale (see Tables 4.14 & 4.15, p. 128).

We designed Study 4 to investigate the impact of response scale type further, by comparing the psychometric properties of the TedQL.3 across three response scales (circles, faces, and thermometer scales, see p. 209-11). We compared the amount of 'don't know' answers, and the amount of responses to items with item-total correlations above 0.20 across scale type (see p. 234). We also compared the internal reliability and reproducibility values across scale type (see p. 235-7). We found that scale type did impact on the psychometric properties of the TedQL.3 measure. Children using the thermometer response scale produced the 'best' psychometric properties for the measure, although the results were somewhat mixed (see p. 234-7). Children using the thermometer scale showed some of the highest internal consistency ratings, and the highest reproducibility of responses over time. The children themselves also preferred this scale to the circles and faces scales. We used these results to justify our choice of the thermometer scale for use with further versions of the TedQL (see p. 265-6).

As summarised in Chapter 7, researchers have made suggestions to help overcome young children's difficulties in understanding and using conventional response scales (see p. 197-203). Children may have problems understanding the wording of anchors used in response scales (see p. 190-1). Some researchers have assessed children's understanding of the words used as anchors in child measures (e.g., Fantuzzo et al., 1996, see p. 198). However we did not provide any assessment of children's understanding and use of the anchors used in the TedQL measure in the studies reported in this thesis. Other researchers have advocated using specific anchors such as *very true/not true of me* (see p. 197). We designed the TedQL measure to use specific anchors, where the children were asked: "which bear is *most like you*?" and "are you *really like this or a little bit*?" (see p. 101 & p. 121).

The literature reviewed in Chapter 7 also showed that researchers have assessed children's rating abilities with the use of calibration tasks or hypothetical questions at the start of their measure (see p. 200-1). We did not use a calibration task in our studies, however we did employ hypothetical questions at the start of the later versions of the TedQL (see p. 122, p. 221, & p. 276). Some researchers have argued that training periods at the start of measures can help familiarise children with the rating task, and lead to more reliable responses on items (see p. 202-3). We used a training period with practice questions at the start of all versions of the TedQL (see p. 104, p. 122, p. 221, & p. 277-8).

Incorporating a theoretical model in the TedQL

We also incorporated a theoretical model in later versions of our measure, to allow children's individual preferences to be taken into account in their QOL scores. In Chapter 1, we argued that the discrepancy model could be applied for use with children below eight years (see p. 10-2). The findings from our review of the developmental literature in Chapter 2 confirmed that using this discrepancy model with young children would be a viable aim. Children need to be able to compare their own abilities and functioning to others around them to make a judgement on how good their own lives are (i.e., to form actual and ideal selves). Researchers have shown that three and four year old children can and do compare themselves to their peers, although they may focus on similarities and compare concrete aspects (as opposed to more abstract comparisons made by older children, see p. 26-8).

On the basis of the above evidence, we incorporated the discrepancy model in the TedQL using items asking about children's actual and ideal selves. In the TedQL.3 and 4 children were asked about their abilities and functioning (i.e., what they were like - their actual self), and also asked a second question for each item (i.e., what they would like to be like – their ideal self, see Chapter 8, p. 219-20). This enabled calculation of children's actual QOL scores and their discrepancy scores (i.e., a score of how their ideal self differed from their actual self, see Chapter 8, p. 223). The results of Studies 4 and 5 showed that children were able to provide reliable and consistent reports of their actual and ideal selves for the TedQL items (see Chapter 9, p. 235, & Chapter 10, p. 285). In Study 5, the children's mean discrepancy scores showed that their ideal selves were different from their actual selves (discrepancy means ranged from 0.37 to 0.52 on the 4-point thermometer scale, see Chapter 10, p. 280). The results of our studies showed that the discrepancy model can be applied successfully in a child QOL self-report measure.

Comparing the TedQL to an existing QOL measure

In addition to these recommendations, whenever possible researchers should attempt to compare children's responses on a new measure with existing similar or equivalent child measures. In Study 1, we compared the initial version of our measure (the TedQL.1) to an existing measure that had been used extensively in child research (the PedsQLTM4.0). The TedQL and the PedsQLTM4.0 were both measures of child QOL, however they differed in the way that items were presented to children and also in the content of items. The PedsQLTM4.0 used a questionnaire format to present items verbally to children (see p. 101-2), whereas the TedQL used props to present items (see p. 70). In addition, the PedsQLTM4.0 asked children about their levels of functioning (e.g., "how much of a problem have you had with walking one block?") to produce QOL scores, whereas the TedQL.1 asked children how they felt about their functioning to produce a second score of 'happiness' with their lives and abilities (see p. 100-1). In this way, the TedQL.1 allowed individual preferences to be taken in account in children's QOL scores. Despite these differences we found that children's scores were correlated across these two measures (see p. 112), therefore we hypothesised that both instruments were measuring a similar underlying construct (i.e., QOL).

In Studies 1 and 5, we also compared the psychometric properties of our measure to the PedsQLTM4.0 to assess the relative merits of both instruments for use with children. In Study 1 we found that while children preferred using the TedQL.1 measure over the PedsQLTM4.0 (see p. 112), the internal consistency of children's responses to TedQL.1 items were lower than their responses to PedsQLTM4.0 items (see p. 111). We argued that this result may have been related to the length of our TedQL measure – the TedQL.1 contained only 10 items (see p. 114-5).

For Study 5, we developed the content of our TedQL items considerably, and the version of the TedQL used in this study contained 22 items (TedQL.4, see p. 257-8). We compared the psychometric properties of our TedQL.4 with the PedsQLTM4.0 (i.e., the internal consistency of responses, the percentage of items with item-total correlations above .20, the percentage of floor and ceiling effects, and the reading age needed, see p. 281). We found that the PedsQLTM4.0 measure produced the 'best' psychometric properties compared to our TedQL.4 measure (in relation to internal consistency of responses, and item-total correlations, see p. 296-7). However our

TedQL.4 measure did produce reasonable psychometric properties, to the extent to which it could be used in child research.

Distinguishing the TedQL from similar child QOL self-report measures

Authors also need to make clear how their measure differs from existing child measures, and be convinced of what their instrument offers above these existing ones. Therefore we considered similar existing child measures in comparison to our TedQL measure. First, the Generic Children's Quality of Life measure (GCQ, Collier et al., 1997, 2000) has been developed since beginning work for this thesis, and this measure is also based around the discrepancy model of QOL. However Collier et al. (2000) did not report any development or investigation of their response scale for this measure, and the items are presented in a written format that may be less understandable to young children, than the TedQL which uses teddies as three-dimensional props.

Second, the Child Health and Illness Profile – Child Edition (CHIP-CE, Rebok et al., 2001) is a measure adapted from an adolescent health-status measure, and the authors have altered the items and response scale to be understandable to children as young as six years. The authors have collected norm data for the child version of their measure (Riley et al., 2004). This measure uses a cartoon format with a graphic response scale similar to the TedQL.4 (using graduated-sized circles). However the CHIP-CE measure may be limited by being adapted from an adolescent measure, and also because it does not allow individual preferences to be taken into account in children's health scores (i.e., it is scored as a high level of ability or functioning is equal to a high level of self-reported health).

Developing a parent report version of the TedQL

As discussed in Chapter 4 (see p. 98) and Chapter 10 (see p. 272), a necessary requirement for the validation of new child measures is moderate agreement between proxy and child reports. In Studies 1 and 2 we investigated the relationship between child and parent rated child QOL. The initial versions of the TedQL measure (the TedQL.1 & 2) did not have a parent report version, and therefore we compared child and parent report using the PedsQLTM4.0 measure. We found that children's and parents' ratings were more likely to be related to each other when they are using the same measure to rate child QOL (see p. 132). In both Studies 1 and 2 we found that when children were using the TedQL measure and parents were using the PedsQLTM4.0 to rate child QOL, their reports were not correlated to each other (see p. 113 & 129). However in Study 1 when children and parents were both using the PedsQLTM4.0 to rate child QOL, their reports were correlated with each other (see p. 113).

As a result of our findings in Studies 1 and 2, we recognised the importance of developing a parent report version of our TedQL measure. Therefore we developed a parent report version for use in Study 5, which enabled comparison of child-parent agreement across our TedQL.3 measure and the PedsQLTM4.0 measure. Due to differences in the instructions at the start of the parent questionnaires for completing these two measures, we predicted that child-parent agreement would be higher for the TedQL.4 compared to the PedsQLTM4.0. We found that children's and parents' responses were not correlated at any age group when using the PedsQLTM4.0 (see p. 295). Agreement between children's and parents' responses was also low on the TedQL.4, however significant correlations between their scores were found for Year 2 children's actual QOL scores, Year 3 children's discrepancy scores, and for all the children combined for their actual QOL scores (see p. 295). We argued that the relationship between child and parent reported child QOL was improved for the TedQL.4 measure due to the instructions given to parents when completing items (parents were asked to rate items *how they think their child would answer* on the TedQL.4, compared to rating items *how they think their children are* in the PedsQLTM4.0). As discussed in Chapter 10, the question still remains as to why parents and children have different views from each other for child QOL (see p. 297-8). Further work could investigate the variables that may impact on child-parent agreement in addition to comparing agreement across different measures.

11.4 Limitations of the TedQL measure

First, in Study 4 we shortened the number of items in the TedQL.3 to produce a version of our measure that was quicker to administer. This decision was related to the need to consider the usefulness, practicality, and feasibility of the TedQL.3 measure for paediatricians where time constraints are a constant issue. We suggested that the TedQL.3 would benefit from the removal of items that were not contributing reliability to the measure, or items that were not synonymous with the rest of the items (see

Chapter 8, p. 214-5). Therefore, we removed eight items from the measure to produce a new version (the TedQL.4, see Table 9.21, p. 256-7).

The TedQL.4 was used in Study 5, and produced responses from the children that were below the .70 standard (actual QOL and discrepancy scores: α= .44 - .64, see Table 10.8, p. 287). This may have been because the TedQL.4 items were too varied from each other (see Chapter 10, p. 294). We realise that the content of our measure does need further development, to produce a set of items that hold better together as a total scale. In retrospect, while reducing the number of items in Study 4 helped to produce a version of the measure that was quicker to administer, it may be that the number of items needs to be expanded again to help increase the reliability of children's responses on the TedQL.

Second, our TedQL measure was designed as a game where the researcher uses teddy bears to act out the items to children. While this presentation style has the advantage of making the task attractive to young children, and helping maintain their attention, it takes time and effort to administer the measure that may be burdensome in clinic settings where time is limited or in larger scale health surveys. We propose that a computerised version of the TedQL measure could be developed, using animated teddy bears and a computerised or pre-recorded voice to read items aloud to children. This would help to produce a version of our measure that would be less dependent on individual presentation, and could be administered with little supervision from the researcher (it could be used independently by children over five years, with more supervision for three to five year olds). A computerised version of the TedQL could be used easily in regular clinic check-ups with children, or even in clinical trials to enable the assessment of QOL from the child's own point of view.

Third, we developed the TedQL as a generic QOL measure, and therefore our instrument does not provide any assessment of specific symptom- or treatment-related problems in ill children. As discussed in Chapter 1 (see p. 10), QOL measure fall into two types – generic or disease-specific, and these have tended to be used for different purposes. Generic measures have been used in population-based health surveys, as they have been considered appropriate for all children regardless of their health status (Eiser & Morse, 2001). Whereas disease-specific measures have been used with specific populations of children with different conditions, and mainly used within clinical trials (Eiser & Morse, 2001). A major problem with using disease-specific measures alone is that such instruments do not allow comparisons across different health conditions (Spieth, 2001). For example, using disease-specific measures means that the impact of cancer versus asthma on children's QOL cannot be directly compared. However, researchers such as Varni et al. (1999) have shown the potential value of developing measures with a generic core and disease-specific modules to provide supplementary information on children with specific conditions. Using a generic measure with disease-specific modules is becoming a widely accepted approach (Spieth, 2001). The TedQL measure currently provides a generic measure that could be used primarily in population-based work, such as healthy surveys in schools, however the development of additional disease-specific modules would help widen the potential applications of our instrument.

11.5 Future directions

The studies in this thesis offered some evidence for the internal reliability, reproducibility, and convergent validity of the TedQL measure, for child and parent versions. However we still need to gain more psychometric data for our new measure. For example, the sensitivity of the TedQL needs to be investigated. Can the TedQL.4 scores discriminate between healthy and ill paediatric populations? Is the TedQL.4 measure sensitive to natural fluctuations in child QOL over time, or change as a result of clinical or psychological interventions with ill children? Another question relates to the criterion validity of our measure – i.e., how well do TedQL.4 scores provide an estimate of an external outcome measured at the same time, or do TedQL.4 scores predict future functioning on a given outcome? Such questions remain to be addressed in future work.

We also want to investigate further young children's use and understanding of the discrepancy model that has been applied successfully here in the TedQL measure (version 3 & 4). We could use the TAM (as used in Study 4, see p. 211-14 & p. 221-2), to investigate how children make judgements about whether "what they are like" (actual self) differs from "what they would like to be like" (ideal self). We could investigate whether children that frequently report low levels of 'normal' functioning (e.g. not being good at running), also report that this does not matter to them, as they prefer to do other things (e.g. playing inside on the computer). We could also use these data to help establish at what age children have the cognitive capacity to form a conception of their ideal self. This age cut-off point could be important for the application of discrepancy-based measures with younger children.

The role of discrepancies may also have potential for work comparing the self-reports of healthy and sick children. We have argued before that there is no simple relationship between illness and QOL (Lawford, Volvaka & Eiser, 2001), and other researchers have shown that ill children may report similar levels of QOL (or even higher) to healthy controls (e.g. Landgraf, 1998). Researchers such as Brossart, Clay, and Willson (2002) and Norman (2003) have begun to investigate the idea of response shift in QOL reports, and this concept may be relevant when considering a discrepancy model of QOL. Response shift has been defined as "the change in one's internal standard, this is, the subject's basis for determining their levels of functioning" (Sprangers & Hoogstraten, 1989, p. 265). The concept of response shift may help explain the paradox where sick populations report the same levels of QOL as their healthy peers (Norman, 2003). Children who are sick may alter their point of comparison (e.g. compare themselves to another sick child rather than a healthy one), which could mean that they alter their internal standards so that they would still report 'normal' levels of functioning, abilities and QOL. Following from this rationale, if we were able to manipulate children's choice of comparison, then we might be able to design interventions that could help improve children's perceived QOL (i.e., by changing their target comparison point and altering their internal standards).

11.6 Final comment – considering the applications of the TedQL measure

Researchers developing child self-report instruments need to ensure the content, presentation style, and response format are tailored specifically for young children to avoid inaccurate or unreliable responses from children. Over the series of studies we reported in this thesis we developed a generic child self-report measure (the TedQL) that attempted to take all these considerations into account. In the development of our instrument we met the three aims for our instrument (see Chapter 1, p. 10). First, in relation to content development, we used information from children to inform the content of the items (see Study 3, Chapter 5). Second, in relation to presentation style, we developed an interview task using teddy bears to present items to children, as an alternative to using a written presentation style (see Chapter 3). Third, we established the most appropriate response format for our instrument by comparing children's responses to the TedQL across three different scale types (see Study 4, Chapters 8 & 9). In addition, we also successfully incorporated the discrepancy model in our measure, by asking children about their actual and ideal functioning for each TedQL item (the TedQL.3 & 4, see Chapters 8 & 10).

The application and uses of the TedQL measure in medical and health research can be seen by referring to the potential value of QOL measures in general, and considering how far our instrument fulfils these goals. First, QOL measures facilitate clinical decision-making by enabling the consideration of the impact of treatments or surgeries on the wider aspects of individual's lives (Havry, 1999). Up until the 1990s outcomes in medical care and policy have been assessed primarily through the physician's report on an individual's behaviour and symptoms (Wallander, 2001). However physicians can under-estimate individual patients' views and fail to recognise the potential impact of a treatment on patients' emotional, social, and psychological lives. QOL measures provide a way for physicians to include the individual perspectives of their patient's in the clinical decision making process. Our TedQL measure addresses this need by providing health professionals with a generic tool that can be used to gain insight into children's own views of their lives, abilities, functioning, and relationships. The development of such child self-report measures is important for paediatricians, whose focus is on assessing the quantity and quality of the children's lives they are treating (Eiser & Morse, 2001). Child QOL measures help put paediatricians in better stead for making considered choices on treatment options for their child patients.

Second, QOL measures help in the evaluation of medical treatments, especially where treatment options have similar survival rates. The assessment of QOL enables medical researchers to compare the short- and long-term impacts, and the emotional, social, and psychological effects, of treatments (Eiser & Morse, 2001). Using such measures, researchers can make a broader assessment of all aspects of a patient's life, rather than a record of the absence or lessening of physical symptoms. QOL instruments can also be beneficial in evaluating the potential of new or unconventional therapies (Joyce et al. 1999), such as homeopathy or aromatherapy. In its current state as a generic child QOL measure, the TedQL would need disease-specific modules to be useful for this aspect of medical research. As we discussed previously (see p. 315), we would also need to make the administration of this measure less dependent on individual presentation.

Third, QOL instruments can be useful in evaluating interventions designed to help patients in a community setting (Eiser & Morse 2001). For example, interventions designed and run by nurses in schools or homes may increase self-esteem, confidence, social abilities, or feelings of self-efficacy (e.g., summer camp,s buddy programmes). QOL measures can provide a broad picture of a patient's functioning, thoughts, and feelings, and therefore will be sensitive to the benefits that such interventions may have for patients. The TedQL could be used to evaluate child interventions, and would allow comparisons across children with different physical conditions.

Fourth, the doctrine of informed consent has meant that medical personnel need to inform their patients of all the treatment options open to them, and the risks of all the alternatives, to enable patients to make more autonomous choices about their treatment plans (Havry, 1999). Being able to assess the potential impact on QOL of different treatments provides medical professionals with more information to help patients in their decisions. Although parents still play a large part in medical decision making for children (as discussed in Chapter 1, see p. 6-7) gaining children's views on their lives is becoming necessary from a legal standpoint. The TedQL measure has both child and parent report versions, which would enable researchers to gain information on both perspectives. Paediatricians in regular clinics or follow-up clinics could use the TedQL measure to compare children's functioning over the course of illnesses and their treatment.

Our TedQL measure provides paediatricians with a useful tool to gain information from young children themselves about their lives, abilities, functioning and relationships. The content, presentation style, and response format of our instrument has been based on the lessons learnt from our literature reviews and from the results of our empirical studies. The TedQL could be used in population-based work such as health surveys across a wide selection of children and their parents, or could be used in schools or community settings as an initial indicator measure for social, emotional, cognitive, or physical health problems.